ISSN:0975-9646

A Modified Classification Based Technique for More Accurate Classification & Prediction of the Intrusion

Neha Agrawal, Priyanka Vijayvargiya

Department Of Computer Science -RGPV University
Indore Sanwer Road, Gram Baroli, Alwasa, Indore, Madhya Pradesh INDIA

***Abstract*— IDS play an important role in network security. Most of the current intrusion detection systems are signature based systems. Signature based IDS also known as misuse detection looks for a specific signature to match, signaling an intrusion. Provided with the signatures or patterns they can detect many or all known attack patterns but they are of little use for as yet unknown attacks. Rate of false positives is close to nil but these types of systems are poor at detecting new attacks or variation of known attacks or attacks that can be masked as normal behavior.**

This paper presents a modified classification technique for the efficient intrusion detection system. The method is based on the concept of decision tree. The experimental results have proved that the accuracy of the proposed technique is better then the existing techniques.

***Keywords*— IDS , SIDS.**

I. INTRODUCTION

The field of intrusion detection has received increasing attention in present years. First reason is the explosive growth of the internet and the large number of networked systems that exist in all types of organizations. The intrusion detection techniques using data mining have attracted more and more interests in recent years. As an important application of data mining these techniques aim to meliorate the great burden of analyzing huge volumes of audit data and realizing performance optimization of detection rules. The objective of proposed work is to try out the intrusion detection on large dataset by classification algorithms and improved its learning time and detection rate in the field of Network based IDS.

RELATED WORK:

In [6] **Jake Ryan et al** applied neural networks to detect intrusions. Neural network can be used to learn a print (user behavior) & identify each user. If it does not match then the system administrator can be alerted. A back propagation neural network called NNID was trained for this process. **Denning D.E et al** [7] has developed a model for monitoring audit record for abnormal activities in the system. Sequential rules are used to capture a user's behavior [8] over time. A rule base is used to store patterns of user's activities deviates significantly from those specified in the rules. High quality sequential patterns are automatically generated using inductive generalization & lower quality patterns are eliminated. An automated strategy for generation of fuzzy rules obtained from definite rules using frequent items. The developed system [9] achieved higher precision in identifying whether the records are normal or attack one. **Dewan M et al** [10] presents an alert classification to reduce false positives in IDS using improved self adaptive Bayesian algorithm (ISABA). It is applied to the security domain of anomaly based network intrusion detection.

S.Sathyabama et al [11] used clustering techniques to group user's behavior together depending on their similarity & to detect different behaviors and specified as outliers. **Amir Azimi Alasti et al** [12] formalized SOM to classify IDS alerts to reduce false positive alerts. Alert filtering & cluster merging algorithms are used to improve the accuracy of the system.SOM is used to find correlations between alerts. **Alan Bivens et al** [13] has developed NIDS using classifying self organizing maps for data clustering. MLP neural network is an efficient way of creating uniform input for detection when a dynamic number of inputs are present.

An ensemble approach [14] helps to indirectly combine the synergistic & complementary features of the different learning paradigms without any complex hybridization. The ensemble approach outperforms both SVMs MARs & ANNs. SVMs outperform MARs & ANN in respect of Scalability, training time, running time & prediction accuracy. This paper [3] focuses on the dimensionality reduction using feature selection. The Rough set support vector machine (RSSVM) approach deploy Johnson's & genetic algorithm of rough set theory to find the reduct sets & sent to SVM to identify any type of new behavior either normal or attack one. **Aly Ei-Senary et al** [1] has used data miner to integrate Apriori & Kuok's algorithms to produce fuzzy logic rules that captures features of interest in network traffic.

Oswais.S et al [2] proposed genetic algorithm to tune the membership function which has been used by IDS. A survey was performed using approaches based on IDS and on implementing of Gas on IDS. **Norouzian M.R et al** [4] **defined** Multi- Layer Perceptron (MLP) for implementing & designing the system to detect the attacks & classifying them in six groups with two hidden layers of neurons in the neural networks. Host based intrusion detection is used to trace system calls. It does not exactly need to know the program codes of each process. The normal & intrusive behavior are collected through system call & analysis is done through data mining & fuzzy technique. The clustering and genetic optimizing steps [5] were used to detect the intrude action with high detection rate & low false alarm rate.

II ARCHITECTURAL DIAGRAM OF PROPOSED METHOD

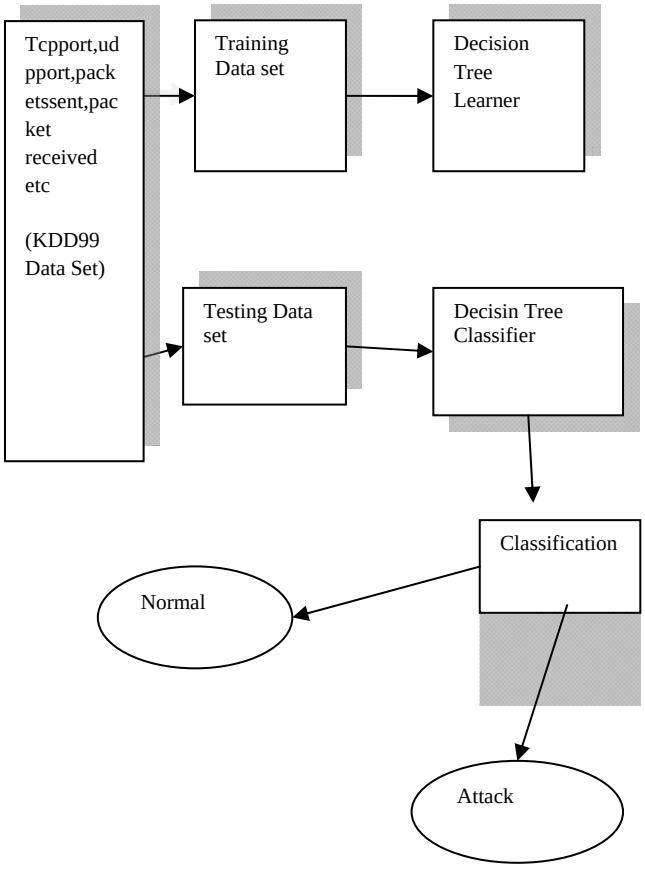

Fig.1 Proposed Architecture

A. Algorithm:

INPUT:

1. A TRAINING DATA SET
2. A TESTING DATA SET

OUTPUT:

1. A DECISION TREE
2. CLASSIFICATION
3. PREDICTION

B. Procedure:

STEP 1: STEP 1:

IF THE TRAINING DATA SET IS EMPTY THEN RETURN A SINGLE NODE WITH VALUE "UNSUCCESS".

IF ALL THE RECORDS OF THE TRAINING DATA SET CONTAINS SAME VALUE FOR THE OUTPUT ATTRIBUTE THEN CREATE A NODE WITH THE SAME VALUE AND RETURN.

ELSE IF EVERY ROW OF THE DATA BASE HAS DOMAIN VALUE NO THEN CREATE A NO NODE AND STOP

ELSE SELECT AN ATTRIBUTE USING THE PROPOSED ATTRIBUTE SELECTION APPROACH AND THEN CONSTRUCT A DECISION NODE

STEP 2: PARTITION THE TRAINING DATA SET T INTO T1,T2,……, TN ACCORDING TO STEP 1 CRITERIA

STEP 3: APPLY THE PROPOSED DECISION TREE CLASSIFIER ALGORITHM RECURSIVELY ON EACH DATA SET T1,T2,……,TN.

C. Attribute Selection:

Our proposed ATTRIBUTE SELECTION methodology is based on a modified Gain. The proposed modified Gain is based on the concept of the greedy approach. It selects the best attribute to create the decision tree then this node is used to perform the partition the data set in to smaller partitions. Then the same process is repeated again and again.

The Entropy will be calculated as follows:

Entropy (D) = D [-p(J) log2 p(J)]

Where:

p (J) is portion of data set of D belonging to class J.

Log2 is log base 2.

D is entire data set

If the value of Entropy(D) is zero then all the members of the data set D are in the same class. Also if the value of the Entropy (D) is one then the members of the data set D are completely different from each other.

Gain(D, A) is information gain of example set D on attribute A is defined as

Modified Gain(D, A) = (Entropy(D) - S ((|Da| / |D|) *Entropy(Da))) * AUV

Where:

AUV is the attribute utility value . It ranges between $0 < AUV < N$.

Da = subset of D for which attribute A has value a.

| Da| = number of elements in Da

|D| = number of elements in D.

III. EXPECTED OUTCOME:

Expected outcomes are as follows:

1. Classification Accuracy is higher
2. Success rate is higher
3. Good for novelty attack detection.
4. Error rate is less

IV. CONCLUSION

Intrusion is harmful for the data flowing on the internet. Intrusion reduces the authenticity. To make sure that the information or data is secure from the unauthorized access, the IDS is must. It identifies the system, which is performing malfunctioning. In this paper, we have proposed a modified method for the intrusion detection system. It is based on the decision tree. It is a decision tree based classifier. It classify data in more accurate manner. It predicts the intrusion on the basis of the learning the existing test data.

REFERENCES

[1] Aly Ei-Semary, Janica Edmonds, Jesus Gonzalez-Pino, Mauricio Papa, "Applying Data Mining of Fuzzy Association Rules to Network Intrusion Detection", in the Proceedings of Workshop on Information Assurance United States Military Academy 2006, IEEE Communication Magazine, West Point, NY,DOI:10.1109/IAW.2006/652083.

[2] Sadiq Ali Khan, "Rule-Based Network Intrusion Detection Using Genetic Algorithm", International Journal of Computer Applications, No: 8, Article: 6, 2011, DOI: 10.5120/2303-2914S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.

[3] Shilendra Kumar, Shrivastava ,Preeti Jain, "Effective Anomaly Based Intrusion Detection Using Rough Set Theory & Support Vector Machine(0975-8887), Vol:18,No:3, March 2011,DOI: 10.5120/2261-2906.

[4] Jin-Ling Zhao, Jiu-fen Zhao ,Jian-Jun Li, "Intrusion Detection Based on Clustering Genetic Algorithm", in Proceedings of International Conference on Machine Learning & Cybernetics (ICML),2005, IEEE Communication Magazine,ISBN:0-7803-9091-1,DOI: 10.1109/ICML.2005.1527621.

[5] Anderson.J.P, "Computer Security Threat Monitoring & Surveilance", Technical Report, James P Anderson co., Fort Washington, Pennsylvania, 1980.

[6] Jake Ryan, Meng - Jang Lin, Risto Miikkulainen, "Intrusion Detection With Neural Networks", Advances in Neural Information Processing System 10, Cambridge, MA:MIT Press,1998,DOI:10.1.1.31.3570.

[7] Denning .D.E, "An Intrusion Detection Model", Transactions on Software Engineering, IEEE Communication Magazine, 1987,SE-13, PP-222-232,DOI:10.1109/TSE.1987.232894.

[8] Teng.H.S, Chen.K and Lu.S.C, "Adaptive Real-Time Anomaly Detection using Inductively Generated Sequential Patterns, in the Proceedings of Symposium on research in Computer Security & Privacy, IEEE Communication Magazine,1990, pp-278-284.

[9] Sekeh.M.A,Bin Maarof.M.A, "Fuzzy Intrusion Detection System Via Data Mining with Sequence of System Calls", in the Proceedings of International Conference on Information Assurance & security (IAS)2009,IEEE Communication Magazine, pp- 154-158,ISBN:978-0-7695-3744-3,DOI:10.1109/IAS.2009.32.

[10] Dewan Md, Farid, Mohammed Zahidur Rahman, "Anomaly Network Intrusion Detection Based on Improved Self Adaptive Bayesian Algorithm", Journal of Computers, Vol 5, pp-23-31, Jan 2010, DOI:10.4.304/jcp 5.1.

[11] Sathyabama.S, Irfan Ahmed.M.S, Saravanan.A,"Network Intrusion Detection Using Clustering: A Data Mining Approach", International Journal of Computer Application (0975-8887), Sep-2011, Vol: 30, No: 4, ISBN: 978-93-80864-87-5, DOI: 10.5120/3670-5071.

[12] Amir Azimi, Alasti, Ahrabi, Ahmad Habibizad Navin, Hadi Bahrbegi, "A New System for Clustering & Classification of Intrusion Detection System Alerts Using SOM", International Journal of Computer Science & Security, Vol: 4, Issue: 6, pp-589-597, 2011

[13] Alan Bivens, Chandrika Palagiri, Rasheda Smith, Boleslaw Szymanski, "Network-Based Intrusion Detection Using Neural Networks", in Proceedings of the Intelligent Engineering Systems Through Artificial Neural Networks, St.Louis, ANNIE-2002, and Vol: 12, pp- 579-584, ASME Press, New York.

[14] Srinivas Mukkamala, Andrew H. Sung, Ajith Abraham, "Intrusion Detection Using an Ensemble of Intelligent Paradigms",Journal of Network & Computer Applications ,pp-1-15, 2004