HYBRID INTRUSION DETECTION FOR CLUSTER BASED WIRELESS SENSOR NETWORK

[1]K.RANJITH SINGH
[1]Dept. of Computer Science, Periyar University, TamilNadu, India
[2]T.HEMA
[2]Dept. of Computer Science, Periyar University, TamilNadu, India

ABSTRACT: Intrusion Detection System is an important technology in business sector as well as active area of research. It is an important tool for information security. An intrusion detection system is used to detect attacks or intrusions and report these intrusions to the user in order to take evasive action. Most of the existing commercial NIDS products are signature-based but not adaptive. Our paper proposes an Adaptive NIDS using K-Means clustering techniques of Data mining approaches. Definite behaviour of network traffic is precisely captured using Data mining approaches, and the set excavated differentiates between "normal" and "attack" traffic. Current researches comprise of single engine detection systems, whereas our proposed system is constructed by a number of Agents, which are totally different in both training and detecting processes. Using k-means clustering algorithm, respective type of packets is clustered under respective Agents formed after clustering. Each of the Agents is responsible for capturing a network behaviour type and hence the system has strength on detecting different types of attacks as well as ability of detecting new types of attacks. The experimental results show that the network traffic pattern used as reliable agents outperforms from traditional signature-based NIDS.

1. INTRODUCTION

An intrusion detection system (IDS) is a device or software application that monitors network and/or system activities for malicious activities or policy violations and produces reports to a Management Station. Some systems may attempt to stop an intrusion attempt but this is neither required nor expected of a monitoring system. They are primarily focused on identifying possible incidents, logging information about them, and reporting attempts. As networking becomes more widespread, the number of violations to normal operations is increasing. Current firewalls are not sufficient to ensure the security in computer networks, which some intrusions take advantages of vulnerabilities in computer systems or use socially engineered penetration techniques that traditional intrusion prevention techniques are not enough in protection. Network Intrusion Detection System (NIDS) will be another wall for protection. Most of the existing commercial NIDS are signature-based but not adaptive. There are many have problems such as attack stealthiest: attackers try to hide their actions from either an individual in monitoring the system or a NIDS novel intrusion: it is undetectable by signature-based NIDS; they can only be detected as anomalies by observing deviations from normal network behaviour. Whereas Anomaly detection approaches attempt to identify abnormal behaviour in patterns and can make use of supervised or unsupervised methods to detect the anomalies or attacks. Unlike the other two methods, these approaches can detect new emerging threats. The supervised anomaly detection approach trains a classifier with just 'normal' labelled patterns. Deviations from 'normal behaviour', everything that is not 'normal', are considering attacks. The disadvantage of the supervised methods for anomaly detection is that the labelling procedure of the training data is expensive and time consuming. The unsupervised anomaly detection approach overcome this problem by making use of data clustering algorithms, which makes no assumption about the labels or classes of the patterns.

The patterns are grouped together based on a similarity measure and the anomalies or attacks are the patterns in the smaller clusters. Two assumptions need to be made for this to be true: the normal patterns or connections are many more than the attacks and that the attacks are different than the normal patterns. In this paper, an adaptive NIDS using K-means technologies is developed, which accurately capture the actual behaviour of network traffic. The proposed NIDS combines the efficiency of both signature based and anomaly based constructed by different types of agent. There are six types of agent based on clustering depending on types of packet. The normal behaviour of a network can be profiled and anomaly traffic can easily be detected with the present of network portfolio. In addition, it can adopt the changes of network automatically with the adaptive learning of agents.

2. RELATED WORK

Most of the commercial NIDSs sold in the market are signature-based with a disadvantage in detection of previously known attacks only. Especially, different kinds of attack come every day. The signature-based NIDS will not be functional when new kinds of attack coming.

Therefore, many researchers have proposed and implemented different intrusion detection models based on data mining techniques to tackle this problem. In this section, a brief review on current works is given. NIDS need to be accurate, adaptive, and extensible. Developed a general and systematic method for intrusion detection and provides an overview on two general data mining algorithms that have been implemented: association rules and frequent episodes using network intrusion detection as a concrete application example, it describes how to construct models that are both accurate in describing the underlying concepts, and efficient for analyzing data in real-time. The same authors describe a framework, MADAM ID, for Mining Audit Data for Automated Models for Intrusion Detection. Proceedings of introduces a new type of clustering-based algorithm for unsupervised anomaly NIDS, which trains on unlabeled data in order to detect new intrusions. Presents a data mining based approach to support signature discovery in NIDS. Furthermore, discusses outlier detection algorithms used in data mining systems. In this paper, an adaptive NIDS based on various data mining techniques is proposed. However, unlike most of the current researches, which only one engine is used for detection of various attacks; the proposed system is constructed by a number of agents, which are totally different in both training and detection processes. In this stage, three data mining approaches: clustering, association and sequential association, are adopted and five types of agent are built. After training with normal traffic for network behaviour, when new type of attack comes, the proposed system can detect such anomaly by distinguishing it from normal traffic.

3. PROPOSED DESIGN OF AGENT BASED NIDS

Using the Data Mining Approaches, we can implement the k-means algorithm for clustering respective type of packets under respective Agents which will form after clustering. This model is supposed to improve performance.

The k-means algorithm is an evolutionary algorithm that gains its name from its method of operation. The algorithm clusters observations into k groups, where k is provided as an input parameter. It then assigns each observation to clusters based upon the observation’s proximity to the mean of the cluster. The cluster’s mean is then recomputed and the process begins again. Here’s how the algorithm works:

1. The algorithm arbitrarily selects k points as the initial cluster centers (“means”).
2. Each point in the dataset is assigned to the closed cluster, based upon the Euclidean distance between each point and each cluster center.
3. Each cluster center is recomputed as the average of the points in that cluster.
4. Steps 2 and 3 repeat until the clusters converge. Convergence may be defined differently depending upon the implementation, but it normally means that either no observations change clusters when steps 2 and 3 are repeated or that the changes do not make a material difference in the definition of the clusters.

3.1. System architecture

The proposed NIDS is composed of four modules, feature miner, Anomaly based agents, and signature based agent and agent trainers. First, a feature extractor converts the data from a monitored system into features which will be used in both training and network intrusion detection stages. Figure 1 shows the overall system architecture.

Proposed IDS

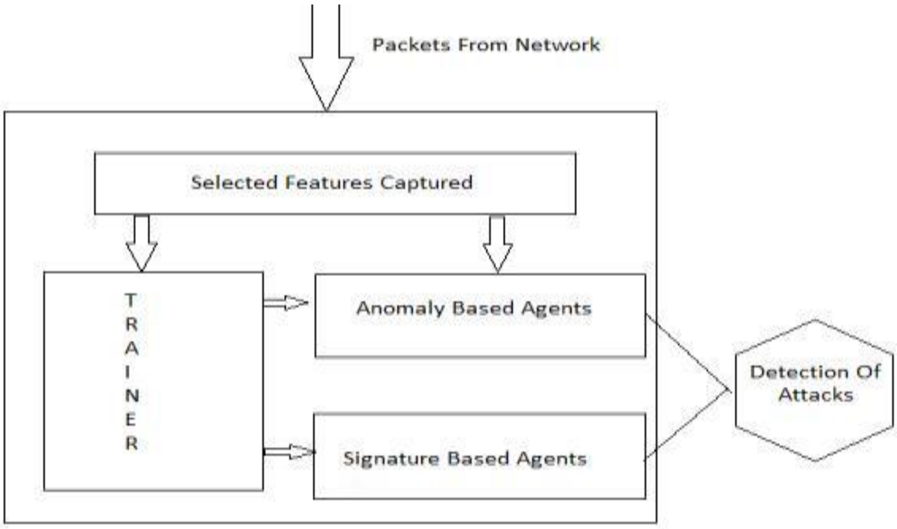

Figure 1.Architecture of agent-based NIDS

The results among all agents are gathered by the agents for concluding the final decision of the system (Figure 2).

SENSOR

It calculates using K-means to find how far (Euclidean distance) of a candidate cluster from normal. If the distance is larger than a threshold, the cluster will be regarded as an intrusion, or vice versa.

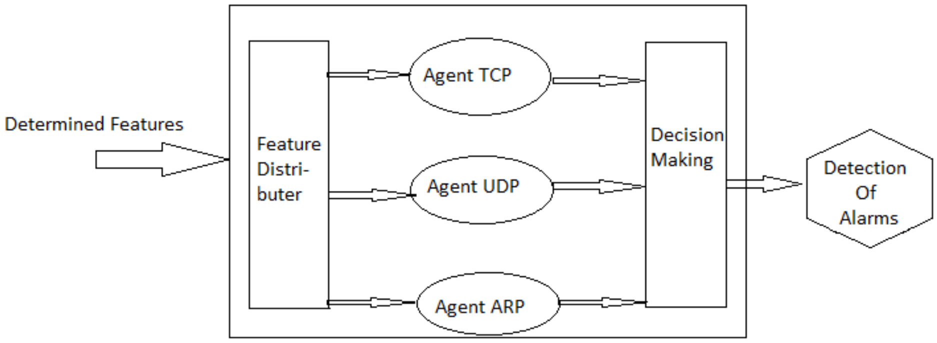

Figure 2. Architecture of the Sensor

For each agent, corresponding trainer is built for updating agent in an adaptive manner. Same as the Detection Engine, a Feature Distributor assigns necessary feature vectors to each training node. Each training node is built in a corresponding data mining approach and updated corresponding agent adaptively. Figure 3 shows the structure of a Trainer.

Agent Trainer

It reforms Corresponding Agents

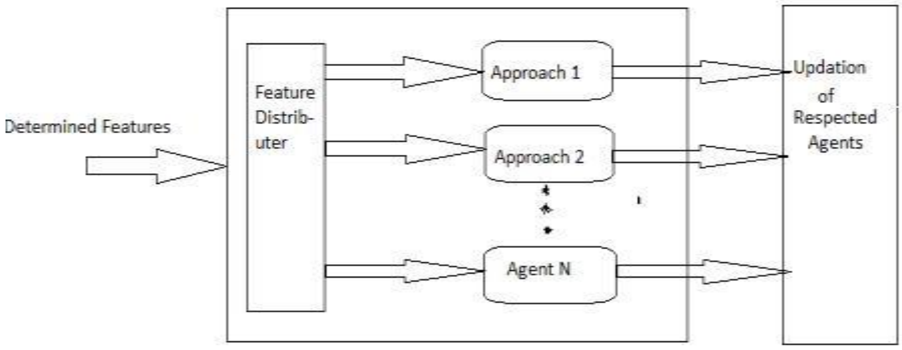

Figure3. Architecture of the agent trainer

An anomaly detection model is based on normal behaviour only and deviations from it. In other words, the normal behaviour of the network is profiled. This model is possibly high in false alarm rate as previously unseen (yet legitimate) system behaviours may be recognized as anomalies, but the adaptive ability of this model to the environment is expected in higher.

3.2. Feature miner

The Feature miner has corresponding functions for each kind of statistics, and it is flexible to use currently, the system supports the following frame feature extraction.

3.3. Clustering-based agent

It extracts behaviour pattern from traffic in terms of frames, and tries to make the normal traffic from isolated clusters in training stage. Then, each cluster will have its representative feature vectors representing certain normal property. For an unknown traffic to be clustered, its traffic property with those trained clusters is compared. If the unknown traffic vector has distance too further away from normal clusters, it is classified as attack traffic, or vice versa.

1. TTL
2. Window Size
3. Packet Length
4. Number of packets in a frame
5. Threshold value of sync bit set count
6. Number of connection attempted to open in a frame

3.3.1. Feature selection

Different feature sets for different clustering-based agents are shown in Table 1.The features selected are specifying for the quantity based attacks such as probing and denial of service.

Table1. Features for clustering-based agents

Agent Feature selected

Cluster TCP	Number of Unique ports Accessed	Mean Packet Size	Time To Live	Window size
Cluster UDP	Number of Unique ports Accessed	Mean Packet Size	Time To Live	Window size
Cluster ARP	Number of Unique ports Accessed	Mean Packet Size	Time To Live	Window size

3.3.2. Frame formation. In the proposed NIDS, consider the basic unit for feature extraction is frame containing number of packets.

3.3.3. Training phase – cluster formation. Depending upon the previous traffic pattern records, the newly arrived packets are trained accordingly and divided into clusters and lead to recalculation of cluster's mean.

3.3.4. Sensor-It is based on how far (Euclidean distance) of a candidate cluster from normal. If the distance is larger than a threshold, the cluster will be regarded as an intrusion, or vice versa.

4. EXPERIMENTS

Investigations on the performance of proposed NIDS is studied, and also, different types of attack are tested to evaluate the strength and limitation of each agent.

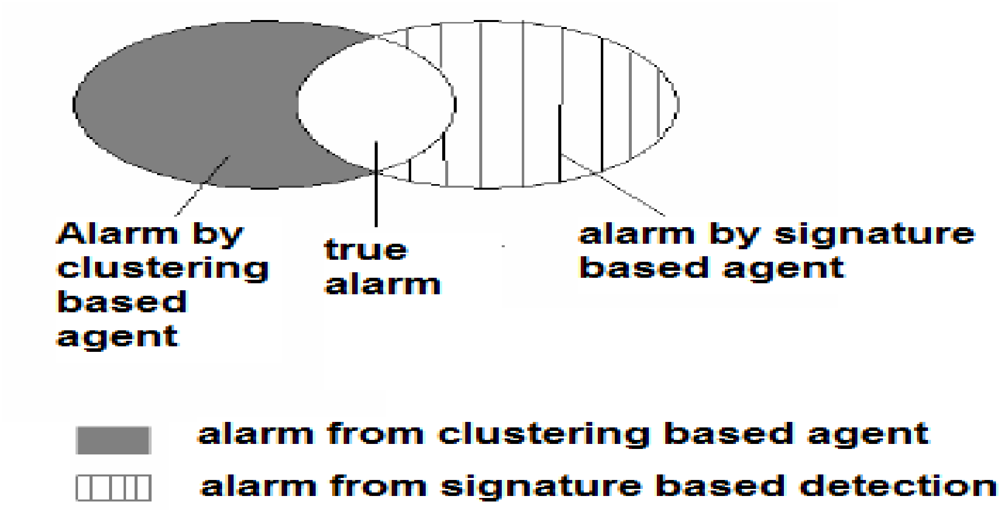
Alarm by clustering based agent
true alarm
alarm by signature based agent
alarm from clustering based agent
alarm from signature based detection

Figure4. ALARM GENERATION RULE

4.1. Experiment Parameters

The packets are captured from the incoming network traffic. In cluster-based agent, the *k*-means clustering approach was adopted and the statistic seed is set with *k* = 256. Squared-error threshold is set to 0(did not set) and maximum loop count is 500. Detail features selected for each clustering-based agent were specified in section 3.3.1. In signature based detection agent, minimum support is set to 100% and depreciation percentage, *depr*= 96. For the tidy representation of data, the different combinations of agents are represented in Table 8.

Table 2 : SAMPLE POLICY FOR DIFFERENT SET OF AGENTS

TCP AGENT 1	Cluster of TCP packets using k-means algorithm
TCP AGENT 2	Cluster of TCP packets using signature based detection
UDP AGENT 1	Cluster of UDP packets using k-means algorithm
UDP AGENT 2	Cluster of UDP packets using signature based detection
ARP AGENT 1	Cluster of ARP packets using k-means algorithm
ARP AGENT 2	Cluster of ARP packets using signature based detection
RULE 1	TCP AGENT 1 (AND) TCP AGENT 2
RULE 2	UDP AGENT 1 (AND) UDP AGENT 2
RULE 3	ARP AGENT 1 (AND) ARP AGENT 2

4.2. Results

Each agent inspects specific kind of traffic, there is lower detection rate in each agent and higher false alarm rate from certain agent. After the alarm decision, a higher detection rate can be achieved, while policy applied limits the false alarm rate. The signature based detection agents yield better detection rate than clustering-based agents on the same attack type, due to its tolerance on noisy background and the adapting ability on attack speed. The benchmarking traditional signature-based NIDS shows that for the attack type without signature, the detection rate is very low, while the proposed NIDS which only based on normal traffic shows its strength on capturing “unseen” attack.

4.3 INTRUSION DETECTION PERFORMANCE

ATTACKS	**AGENT**						**RULES**		
	TCP1	TCP2	UDP1	UDP2	ARP1	ARP2	RULE1	RULE2	RULE3
TCP LAND	98.4	98.1	0.0	0.0	0.0	0.0	98.25	0.0	0.0
DOS	99.1	98.9	0.0	0.0	0.0	0.0	99	0.0	0.0
UDP FLOOD		0.0	98.6	98.0	0.0	0.0	0.0	98.3	0.0
TCP SYN	99.6	91.4	0.0	0.0	0.0	0.0	95.5	0.0	0.0

The performance is analysed using accuracy and false alarm rate as the parameters. Hence, using the average of both the agents designated on the basis of clustering based detection agents and signature based detection agents, the rate of accurately finding the attacks increases since the attacks are first tested with the signature based detection and then the remaining unsuspected packets to the clustering based agents.

5. CONCLUSIONS

Most of the existing commercial NIDS products are signature-based but not adaptive. In our paper, an **adaptive NIDS** using data mining technology is developed. Data mining approaches are used to accurately capture the actual behaviour of network traffic, and portfolio mined is useful for differentiating “normal” and “attack” traffics. On the other hand, most of the current researches are using only one engine for detection of various attacks; the proposed system is constructed by a number of **agents**, which are totally different in both **training and detecting processes**. Using the Data Mining Approaches, we can implement the **k-means algorithm** for clustering respective type of packets under respective Agents which will form after **clustering**. Each of the agents has its own strength on capturing a kind of network behaviour and hence the system has strength on detecting different types of attack. In addition, its ability on detecting new types of attacks. The experimental results show that the frequent patterns mined from the audit data could be used as reliable agents, which **outperformed** from traditional signature-based NIDS. For future development, the following directions are proposed: (i) To develop more agents which are strength on other aspects

(ii) To set the thresholds by the system with minimum human interrupt

(iii) To introduce incremental updating mechanism for the detection agents.

REFERENCES

[1] IBMs Aglet Mobile Agent Implementation.

[2] G.W. Lee and S. J. Stolfo, "Data Mining Approaches for Intrusion Detection," In Proceedings of the Seventh USENIX Security Symposium (SECURITY '98), San Antonio, TX, January 1998.

[3] W. Lee, S. Stolfo and K. Mok, "Mining Audit Data to Build Intrusion Detection Models," In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD '98), New York, NY, August 1998.

[4] R. Agrawal, T. Imielinski and A. Swami, "Mining Associations between Sets of Items in Massive Databases," In Proceedings of the ACM-SIGMOD Int'l Conference on Management of Data, Washington D.C., pp. 207-216, May 1993.

[5] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," In Proceedngs of the 20th Int'l Conference on Very Large Databases (VLDB94), Santiago, Chile, September. 1994.

[6] H. Mannila, H. Toivonen and A. I. Verkamo, "Discovery of Frequent Episodes in Event Sequences," Data Mining and Knowledge Discovery

1(3), pp. 259-289, 1997.

[7] W. Lee, S. Stolfo and K. Mok, "Mining in a Data-flow Environment: Experience in Network Intrusion Detection," In Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '99), San Diego, CA, August,1999.

[8] W. Lee and S. J. Stolfo, "A framework for constructing features and models for network intrusion detection systems," ACM Transactions on Information and System Security (TISSEC), Vol. 3, Issue 4, pp. 227-261, November 2000.

[9] L. Portnoy, E. Eskin and S. J. Stolfo, "Intrusion detection with unlabeled data using clustering," In Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001), Philadelphia, PA, November 2001.

[10] Han, X. L. Lu, J. Lu, C. Bo, and R. L. Yong, "Data mining aided signature discovery in network-based network intrusion detection system," ACM SIGOPS Operating Systems Review, Vol. 36, No. 4, pp. 7-13, October 2002.

[11] M. I. Petrovskiy , "Outlier Detection Algorithms in Data Mining Systems," Source Programming and Computing Software, Vol. 29 , Issue 4, pp. 228 237, July-August 2003.

[12] R. Agrawal, T. Imielinski, A. Swami and R. Srikant, "Mining Association Rules between Sets of Items in Large Databases," In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207- 216, May 1993

[13] R. Agrawal and R. Srikant, "Mining Sequential Patterns," In Proceedings of the Int'l Conference on Data Engineering (ICDE), Taipei, Taiwan, March 1995.

[14] http://www.snort.org

[15] http://www.ll.mit.edu/SST/ideval/data/data_index.html

[16] Malgalhaes, Ricky. "Host based IDS vs. Network based IDS." [Online] Available http://www.windowsecurity.com/articles/Hids_vs_Nids_Part1.html, July 10, 2003.

[17] VarunChandola, Arindam Banerjee, and Vipin Kumar. Anomaly Detection: A survey.2007

[18] Leonid Portnoy, EleazarEskin, Sal Stolfo. Intrusion detection with unlabeled data using clustering. 2001

[19] Stefano Zanero, Sergio M. Savaresi. Unsupervised learning techniques for a intrusion detection system. 2004

[20] Alexandar Lazarevic, LeventErtoz, Vipin Kumar, AyselOzgur, Jaideep Srivastava. A Comparative study of anomaly detection schemes in network intrusion detection. 2003

[21] Anita K. Jones, Robert S. Sielken. Computer system intrusion detection: A survey. 2000