A Survey of Various Techniques to Avoid Intrusion Detection in Wireless Sensor Networks

Chethan K S
Assistant Professor, Department of CSE
GSSSIETW, Mysuru

Dr. S. Meenakshi Sundaram
Professor & Head, Department of CSE
GSSSIETW, Mysuru

***Abstract* - Intrusion Detection System is a security management system that monitors the system and detects malicious activity.Machine learning techniques are used to evaluate the performance of intrusion detection system and some of the techniques discussed in this paper are decision tree, ripper rule, back-propagation neural network, bayesian network, naive bayesian classifier, radial basis function neural network. Swarm Intelligence approaches are used to solve complicated problems by multiple simple agents without centralized control. The swarm intelligence algorithms inspired by animal behaviour in nature such as ants finding shortest path in finding food; a flock of birds flies or a school of fish swims in unison and approaches such as ant colony optimization and practical swarm optimization are discussed in this paper.IDBGC algorithm automatically establish clusters and detect intruders by labeling normal and abnormal groups.In overall this paper gives a survey on different techniques of intrusion detection system.**

Keywords: Intrusion Detection, Swarm Intelligence, Clustering

I .INTRODUCTION

Internet services have become essential to business commerceas well as to individuals. With the increasing reliance on network services, the availability, confidentiality, and integrity of critical information have become increasingly compromised by remote intrusions[7]. Enterprises are forced to fortify their networks against malicious activities and network threats. Therefore, a network system must use one or more security tools such as a firewall, antivirussoftware or an intrusion detection system to protect important data/services from hackers or intruders.

Network security has become an indispensable factor of computer technology with the development of internet. The security of a computer system or network is compromised when an intrusion takes place. An intrusion can be defined as any set of actions that makes an attempt to compromise the integrity, confidentiality or availability of a resource. Intrusion prevention techniques such as firewalls, access control or encryption have failed to fully protect networks and systems from increasing attacks and malwares. As a result, Intrusion Detection System (IDS) have become an essential component of security infrastructure to detect these threats, identify and track the intruders. As IDS must have a high attack Detection Rate (DR), with a low False Alarm Rate (FAR) at the same time, construction of IDS is a challenging task. In the recent past, biology inspired approaches have made their appearance in a variety of research fields, ranging from engineering, computer science, economics, medicine and social sciences. Likewise, many biology inspired techniques have been proposed for intrusion detection to improve their efficiency and performance. Swarm intelligence is one of them. Techniques and algorithms of this research field draw their inspiration from the behaviour of insects, birds and fishes, and their unique ability to solve complex tasks in the form of swarms.

II. INTRUSION DETECTION

Intrusion Detection System has received great attention from researchers all over the world because of their ability to keep track of the network behaviour, so that abnormal behaviour can be detected quickly[6]. Figure 1 shows the intrusion detection model.

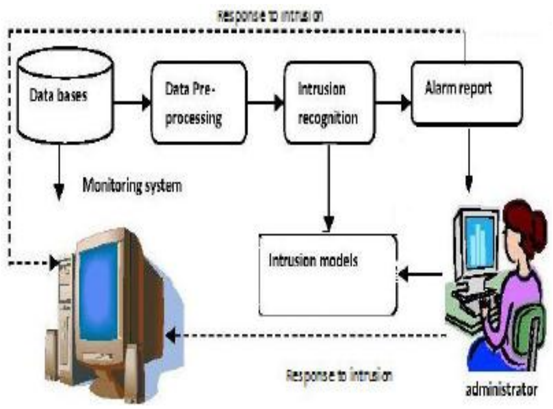

Figure 1: Intrusion Detection model

An IDS is generally categorized as misuse detection and anomaly detection. The misuse detection can detect intrusions with low false alarm rate, but it fails to detect new attacks. IDS analyze the information it gathers and matches with the large databases of intrusive behaviour or attack signatures. It is also known as signature-based detection. Anomaly detection has the capability of detecting new types of attacks and is classified as static and dynamic. It determines whether deviation from the established normal usage patterns and is stated as intrusions.

The two most popular performance evaluation metrics in IDS are: Detection Rate (DR), which is defined as the ratio of the number of correctly detected attacks to the total number of attacks, and False Alarm Rate (FAR), or False

Positive Rate (FPR), which is the ratio of the number of normal connections that are misclassified as attacks to the total number of normal connections.

Table 1: Comparison of Performance evaluation of IDS in Computer Network

Computer Network	Behaviour Analysis	Attack Type	Attack Categories	Methods	Evaluation Metrics
Sensor network	Limited range radio communications nodes.	Jamming, Tampering, Selective forwarding, sinkhole, Sybil attack [9]	Probe, DOS, U2R,R2L	K-means, SVM, Navie-bayes	Network delay, Throughput, Success rate, Latency, Energy Consumption.[13]
Cloud Network	Network traffic, detect known attacks, log files for auditing	Back door channel attack, Insider attack, Flooding attack.[10]	DOS, U2R, Port scanning.	Anomaly based detection, Signature based detection, SVM	Fairness, Isolation
Adhoc network	On demand routing protocol, RREQ, RREP packets	Black-hole attack, warmhole attack, Replay attack, Message tampering.[11]	DOS,RTO	Network load, Network Size, Channel error, node mobility	Normalized routing overhead, average end to end delay, throughput, Collisions[12]

A. Sensor Network

Wireless sensor networks are composed of several sensors deployed in areas where the aim is to collect data and forward it for the analysis. It has become an increasingly interesting field of research in solving such challenging real-world problem, as environmental monitoring, military applications, geographical sensing, traffic control, and home automation. The properties of WSN show that that sensor node is completely restricted by resources, including memory, energy, computing, communication and bandwidth. Therefore, the deployment of these kinds of networks with their resource restrictions makes their security issue essential, and vulnerable to various security threats. Key management and authentication have been used to protect WSNs from different attacks, encryption and authentication are the first security measures as the first line of defense for protecting WSN. But cryptography based on secret key management are not enough to protect the WSN, because even in the presence of this first line of defense, several attacks may extract sensitive information, and use them for malicious reason.

B.Cloud Network

Cloud computing is an emerging technology adopted by organizations of all scale due to its low-cost and pay-as-you-go structure. The distributed nature of cloud environment makes it most vulnerable and attractive environment for the intruders to perform attacks. Intrusion detection systems can be used to enhance the security of such systems by systematically examining the logs, network traffic as well as configurations. However, conventional intrusion detection systems (IDSs)—which can be classified into host-based intrusion detection systems (HIDS) and network-based intrusion detection systems (NIDS)—are not appropriate for cloud environment as these are unable to locate the hidden attack trail, e.g., the network-based IDS is unable to detect any event in case of encrypted node communication and it is possible for the attacker to gain control over the installed virtual machines if the hypervisor is compromised. Attackers can use the compromised hypervisor to gain control over the host. Owing to the fact that the IDS techniques were not designed with the specific context of virtualization under consideration, they do not offer the same protection in such environments. There are certain trade-offs that need to be faced when deploying IDS in the virtual environment, mostly because of their inability to inspect the internal working of the operating systems. Despite the huge benefits that are offered by virtualization, there are a number of security risks that are associated with it. It introduces a number of new problems that did not exist in a traditional computing environment.

C. Adhoc Network

As network-based computer systems play increasingly vital roles in modern society, they have become the targets of our enemies and criminals. When an intrusion takes place, intrusion prevention techniques, such as encryption and authentication (e.g., using passwords or biometrics), are usually the first line of defense. However, intrusion prevention alone is not sufficient because as systems become ever more complex, while security is still often the after-thought, there are always exploitable weaknesses in the systems due to design and programming errors. The primary assumptions of intrusion detection are: user and program activities are observable, for example via system auditing mechanisms; and more importantly, normal and intrusion activities have distinct behavior. Intrusion detection therefore involves capturing audit data and reasoning about the evidence in the data to determine whether the system is under attack. Based on the type of audit data used, intrusion detection systems (IDSs) can be categorized as network-based or host-based. A network-based IDS normally runs at the gateway of a network and captures" and examines network packets that go through the network hardware interface. A host-based IDS relies on operating system audit data to monitor and analyze the events generated by programs or users on the host. Intrusion detection techniques can be categorized into misuse detection and anomaly detection.

III. MACHINE LEARNING TECHNIQUES

There are many machine learning techniques which can be used for data classification. We present the general concepts and a brief description of some popular supervised learning techniques that generally give high detection accuracy for our IDS approach [1]below.

A. Decision Tree

Decision Tree is a well-known classification algorithm which can efficiently classify data [3]. A Decision Tree consists of non-terminal nodes (a root and internal nodes) and terminal nodes (leaves). The root node is the first attribute with test conditions to split each input data record toward each internal node depending on characteristics of the data record. First, the Decision Tree istrained with known data before it can classify new or untrained data. The training process builds the Decision Tree by identifying attributes and values that would be used to test the input data at each internal node. After training, the tree can predict or classify new data by starting from a root node and traversing internal nodes based on attributes of test conditions until it arrives at a leaf (terminal) node consisting of an answer class. The C4.5 version is an efficient and popular algorithm for training and building a DecisionTree [2].

B. Ripper Rule

Ripper (Repeated Incremental Pruning to Produce Error Reduction) Rule is an efficient rule-based learning algorithm that can process various noisy datasets [2]. The Ripper Rule algorithm consists of two stages. The first stage is to initialize the rule conditions. The next stage uses a rule optimization technique. This step repruneseach rule of the rule set to minimize the errors. A training dataset consists of a growing set and a pruning set. When the Ripper Rules have already been trained by the training dataset, this algorithm checks the condition in each rule to classify the testing dataset. The rules are in if-then-else form. Each class is considered one by one until the condition in a rule is met, or the data is classified into a class.

C. Back-Propagation Neural Network

Back-Propagation Neural Network (BPNN) model is a well knownsupervised learning model that can effectively classify many types of data. BPNN is a feed-forward multi-layer network. Its input vectors and the corresponding target vectors are used in training the network until it can approximate a function, and associate its input vectors with specific output vectors. BPNN has many possible training algorithms such as gradient descent, momentum and resilient algorithm.

D. Bayesian Network

A Bayesian Network involves both a graphical model and probabilistic model representing random variables and conditional independence through a directed acyclic graph. Nodes in the graph represent random variables, while edges represent conditional dependencies. Thus nodes that are not connected represent variables that are conditionally independent from each other. The Bayesian Network learns the casual relations between attributes and class labels from the training dataset before it can classify unknown data.

E. Naive Bayesian classifier

Naive Bayes is a simple technique for classification using a simple probabilistic model from Bayes's theorem with the assumptions of independent attributes. Naive Bayes is a type of supervised learning algorithm that uses a maximum likelihood method for parameter estimation. It requires a set of training data to estimate means and variances of the attributes for classification. This technique works well with many complex real-world applications [4].

F. Radial Basis Function Neural Network

Radial Basis Function Neural Network (RBF-NN) has a feed-forward structure similar to the Back-Propagation Neural Network (BPNN). Both models have an input layer, a hidden layer and an output layer. RBF-NN uses radial basis functions as activation functions. They can be used for function approximation and time-series prediction. The RBF-NN is also similar to a K-Nearest Neighbor (k-NN) model, where data is classified by measuring Euclidean distances between input and the hidden layer centers. In dealing with large dataset, the RBF-NN usually uses much less computation time than a BPNN does [5].

G. Datasets

Different datasets were used as training dataset and testing dataset. The KDD99 and the RLD09 training datasets each had 20,000 records consisting of 10,000 attack records and 10,000 normal records. The KDD and the RLD09 testing datasets each had 10,000 records consisting of 5000 attack records and 5000 normal records. The attack records in each dataset contain various types of DoS and Probe attacks.

IV. SWARM INTELLIGENCE

The term Swarm Intelligence (SI) was first introduced by Beni in the context of cellular robotics system. A swarm can be considered as a group of cooperating agents to achieve some purposeful behaviour and task. The simple scheme of a swarm is shown in Figure 1. It links to artificial life, in general, there are several collective behaviour like birds flocking, ant colonies, social insects and swarm theory.

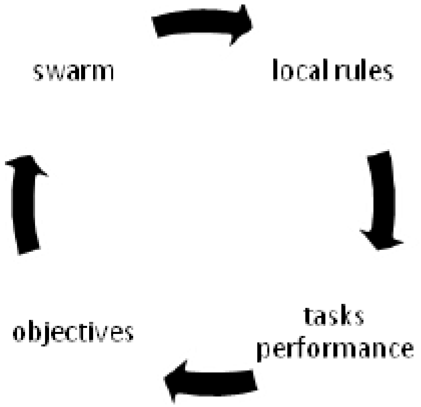

Figure2: Scheme of swarm

Swarm Intelligence approaches in Intrusion detection are Ant Colony Optimization (ACO), Particle swarm Optimization (PSO) and Bee Colony Optimization (BCO). The foraging behaviour of ants and their ability to find the shortest path from their nests to food source as shown in Figure 2, has inspired the creation of the algorithmic model which is known as Ant Colony Optimization.

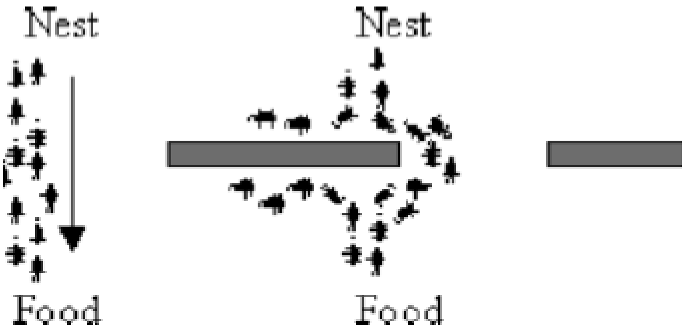

Figure3.Behaviour of Ants

A. Ant Colony Optimization

Dorigo et al. (1999, 2004) presented an algorithmic implementation of the ant behaviour for solving minimum cost path problems on graphs known as simple Ant Colony Optimization[6]. ACO is set apart from the other approaches, as it is primarily applied to combinatorial optimisation. There are recent ACO algorithms proposed for continuous numerical optimisation, (Dréo and Siarry (2002) and Socha and Dorigo (2006)), however, their application to this domain is limited. ACO represents the problem as a graph and treats it as combinatorial optimization.

B. Particle swarm Optimization

Kennedy and Eberhart (1995) introduced the term Particle Swarm Optimization and their work was the main influence of the basic PSO model. According to this model a fitness function exists which measures the quality of the current solution[6]. A number of particles (solutions) are placed randomly inside the hyperspace, each having a random velocity. The particles move in the hyperspace and at each step evaluates their position according to the fitness function. Each particle in the swarm represents a possible solution. Two key features of this model are: (a) the speed (and therefore the next position) of each particle is calculated according to the findings of both that particle and the findings of the rest of the swarm and (b) the global best solution is communicated among all particles of the swarm. Dozier et al., (2004) presented PSO technique that can be used as a part of IDS to identify possible attacks. The basic variants of PSO have been developed to improve speed of convergence and quality of solution found by the PSO.

V. IDBGC ALGORITHM

Clustering in D-dimensional Euclidean space RD is a process of partitioning a given set of n instances into K groups based on some similarity or dissimilarity metrics[8]. Let the set of n points {x1; x2; : : : ; xn} be representedby the set C and K clusters be representedby c1; c2; : : : ; cK . Then

ci_= _; i = 1; : : : ; K;
ci∩ cj = _; i; j = 1; : : : ; K and i_= j and
_K
i=1
ci = C:

The IDBGC algorithm consists of two stages. They arestatedas follows:

1. The stage of nearest neighbor clustering: To establish the set of original clusters using the nearest neighbor methodby grouping very similar instances into a cluster and filter noisy data objects based on some similarity or dissimilarity metrics.

2. The stage of genetic optimization: To combine original clusters by genetic algorithms and obtain the near optimal result, then label the cluster including most activities asthe normal according to above assumptions.

VI. CONCLUSION

This paper discuss about survey on techniques of intrusion detection system. Machine learning techniques evaluate the performance of intrusion detection system. These techniques uses two types of datasets KDD99 and RLD09. By evaluating various machine learning algorithms, decision tree gave higher total detection rates than other algorithms. Swarm Intelligence can be derived from the behaviour of the animals.The low complexity of ACO algorithm establishes it as a major candidate for the creation of fast, robust and adaptive IDS. The combination of ACO with some other machine learning technique is expected to lead to highly adaptive IDS. On the other hand, PSO based IDS have been extensively studied in combination with other ML techniques constantly providing solid DR rates. IDBGC provides automatically establish clusters and detect intruders by labeling normal and abnormal groups. The first stage takes the nearest neighbor method to group network data, which avoids the limitation of the clustering center in some clustering problems. The second one is a genetic optimization process to obtain the near optimal detection result. The main purpose of the first stage is to reduce the size of data objects to a moderate one so as to be suitable for genetic algorithms in the second stage.

REFERENCES

[1] P. Sangkatsanee, N. Wattanapongsakorn, C. Charnsripinyo, Network intrusion detection with artificial neural network, decision tree and rule basedapproaches, in: The International Joint Conference on Computer Science andSoftware Engineering, Thailand, 2009.

[2] Z. Pan, S. Chen, G. Hu, D. Zhang, Hybrid neural network and C4.5 for misusedetection, in: The 2nd International Conference on Machine Learning andCybernetics, China, 2003, pp.2463–2467.

[3] P.N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Pearson Addison Wesley, 2005.

[4] M. Panda, M.R. Patra, Semi-Naı¨ve Bayesian method for network intrusion detection system, Neural information processing, Lecture Notes in ComputerScience (Springer Link) 5863 (2009) 614–621.

[5] S.X. Wu, W. Banzhaf, The use of computational intelligence in intrusiondetection system: a review, Applied Soft Computing 10 (2010) 1–35.

[6] P.Amudha,H.AbdulRauf,A Study on Swarm Intelligence Techniques in Intrusion Detection in:IJCA Special Issue on “Computational Intelligence & Information Security" CIIS 2012

[7] PhurivitSangkatsanee, NaruemonWattanapongsakorn ,ChalermpolCharnsripinyo, Practical real-time intrusion detection using machine learning approaches:www.elsevier.com/locate/com,Computer Communications.

[8] YongguoLiua, Kefei Chen, Xiaofeng Liao, Wei Zhang, A genetic clustering method for intrusion detection:www.elsevier.com/locate/com

[9] Yousef El Mourabit, Anouar Bouirden, Ahmed Toumanari, Intrusion Detection Techniques in Wireless Sensor Network using Data Mining Algorithms: Comparative Evaluation Based on Attacks Detection.

[10] U. Oktay and O.K. Sahingoz, :Attack Types and Intrusion Detection Systems in Cloud Computing

[11] P. Narendra Reddy , CH. Vishnuvardhan , V. Ramesh : Routing Attacks In Mobile Ad Hoc Networks

[12] A.A.A. Radwan , T.M. Mahmoud, E.H. Houssein : Evaluation comparison of some ad hoc networks routing protocols

[13] L. Alazzawi and A. Elkateeb: Research article: Performance Evaluation of the WSN Routing Protocols Scalability

[14] Dr.Andhe Dharani, Mr.Manjuprasad B, Dr.Shantharam Nayak, Dr.Vijayalakshmi M.N., "Performance Analysis of Cluster Based Protocols in Sensor Networks and their Vulnerabilities", International Journal of Innovative Research in Computer and Communication Engineering(IJIRCCE), Vol. 3, Issue 4, ISSN (Online): 2320-9801, April 2015,pp.3045-3051. DOI: [10.15680/ijircce.2015.0304107] Impact Factor (2013) - 4.447.