International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI November 2017- Available at www.ijraset.com

A Machine Learning Approach for Secure Intrusion Detection in Wireless Sensor Networks

Arun Kunar Silicery[1], Suvarna S[2]
[1]*Dept of CSE, Raja Mhendra College Of Engineering Lbrabmpthnam, Hyderabad*
[2]*Assistant professor, Dept of CSE, Raja Mhendra College Of Engineering Lbrabmpthnam, Hyderabad*

Abstract: ***The wireless sensor network connects millions of nodes over the world running on various platforms for providing secure communication correspondence and business services. In any case, this interconnectivity among nodes likewise enables malwares to corrupt resources and mount Internet threats. Persistently designing internet threats represents a serious challenge to create a flexible, adaptive security oriented techniques. Presently, a large portion of commercially available intrusion detection frameworks are signature based. These frameworks perform well in identifying known attacks whose signature lives in the database. Such frameworks require frequent rule base updates and updated signatures, and are not equipped for recognizing unknown attacks. For discovery of known and zero day unknown attacks, inconsistency based system intrusion detection frameworks are best methodologies. anomaly based strategies are reasonably exceptionally appealing be that as it may, numerous issues stays to be solved some time recently being adopted generally. Issues including false alarm rate, reorganization accuracy and inability to scale high speed etc. and so forth still should be understood. In this paper, we can assemble Intrusion detection framework model to identify attacks on wireless sensor networks and can enhance the framework utilizing captured information. By applying feature selection approach in machine learning in, the NSLKDD data collection gained can be decreased and furthermore can enhance the Intrusion detection utilizing the captured information. By machine learning procedures, we can build number of new unknown attacks the network of Intrusion detection can be implemented.***
Keywords: Network security, Feature Selection, intrusion detection system, Machine Learning

I. INTRODUCTION

Data access through Internet gives intruders different methods of corrupting a node system. To an ever increasing more organizations have turned out to be vulnerable because of intrusion of digital attacks which compromise the security targets of their network system. There are five security goals: first security objective is the Availability—it secures the framework against intentional or inadvertent attempts to deny trusted clients access to data or frameworks. Second security objective is the Integrity of Data or Systems—it guarantees the un-modification of information in any unauthorized way. Third goal is Consistency of Data or Systems—it guarantees the security of data against unauthorized access or utilize. Fourth objective is Accountability—it guarantees the essential control to find activities to their source. Accountability directly accepts non-repudiation, discouragement, intrusion prevention, security verification, recovery, and legitimate suitability of records. Last security objective is the Assurance—it guarantees the confidence that specialized and operational security measures work in as planned. Assurance features the thought that security frameworks give the planned functionality while resolving undesired actions. An intrusion or node attack can be characterized as "any collection of activities that endeavor to compromise the security objectives". There are different sorts of attack that can compromise the security goals. DARPA has partitioned these attacks into four distinctive classes: top of the line is Probe: here an attacker checks a system to capture data with a specific end goal to discover known vulnerabilities. This kind of attack influences the consistency and accountability security goals. Second kind of attack is Denial of Service (DoS) attack: here an attacker makes a few registering or memory resources excessively occupied or too full, making it impossible to deal with legitimate demands, denying authorized client's access to a machine. This attack straightforwardly influences the availability, accountability and confirmation security objectives. Third kind of attack is User to Root (U2R) attack: here an attacker begins with access to an ordinary client account on the framework by picking up root get to access. This attack influences privately and integrity targets. Last and fourth kind of attack is Remote to Local (R2L) attack: here an attacker sends packets information to a machine over a system that adventures the machine's vulnerability to pick up neighborhood access as a client illegally. This attack may prompts influence of confidentiality and trustworthiness targets. These security targets are between inner dependent. Any kind of nodeattack influencing one of security objective additionally influences the others in some way. With the weakness of nowadays software and protocols joined with the expanding sophistication of attacks, it will come as no one that system based attacks is on the rise. The

Computer Emergency Response Group/Coordination Center (CERT/CC) has detailed that the number of nodeattacks has maximized exponentially in the recent years from 1990 to 2003. To minimize the economical losses, different security components are connected to network. Around 86 percent of the organizations utilized firewalls but firewalls independent from anyone else are definitely not adequate to give adequate protection. These uncertain security components has presented regularly evolving field of intrusion detection and prevention.

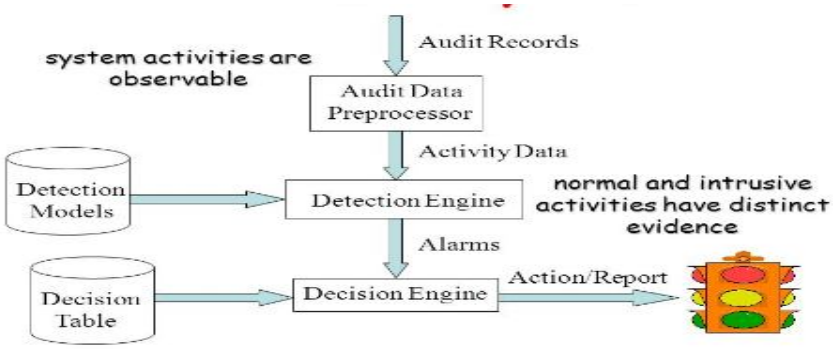

Figure 1: System Architecture

An intrusion detection framework accumulates and examines data from various zones of nodes or system to distinguish possible attempts to compromise the security targets of a framework/network. The principle of this content is to show a short overview of different types of intrusion detection framework; Artificial Intelligence (AI) based procedures utilized for intrusion detection and to adopt the multi classifier approach for identifying intrusions efficiently.

II. INTRUSION DETECTION SYSTEM

An Intrusion Detection framework (IDS) characterized as “aneffective security framework, which can identify, prevent and possibly respond to the node attack” is one of the standard segments in security frameworks. It observes target sources of actions, for example, audit and wireless network traffic information in node or system frameworks and deploys different methods to give security services. The primary target of IDS is to distinguish all interruptions in an effective way. The development of IDS permits network administrators to distinguish security target violations. These security objective violations extend from outside attackers trying to pick up unauthorized access to network security environment or making resources inaccessible to insiders abusing their entrance of the framework resources.

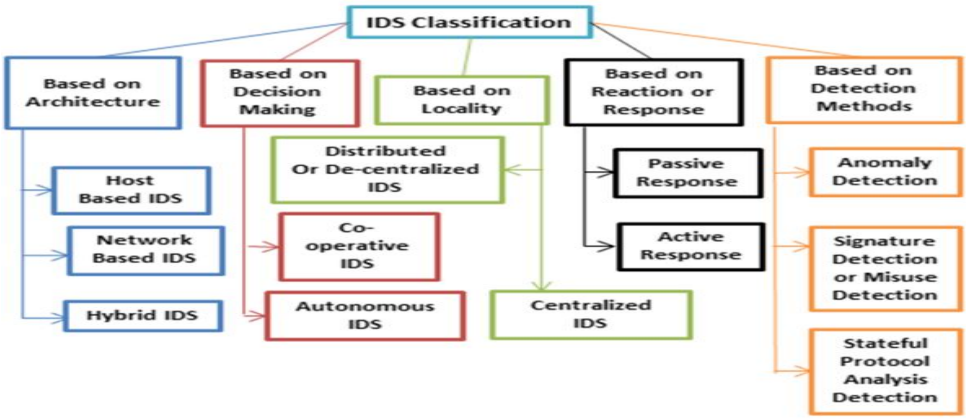

Figure 2: Classification of Intrusion Detection Systems

Based upon the various modules, IDS can be classified into different classes. In view of dataset collection and memory storage unit, IDS can be separated into two classes:

A. Host based IDS

Host based IDS gathers the information from a host to be secured. They gathers the information for the most part from network calls, operating system log documents, NT occasions log record, CPU usage logs, application log documents etc. the purpose of

Host based IDS is that they are platform dependent and are exceptionally effective to identify attacks like buffer overflow. These frameworks become noticeably wasteful if there should be an occurrence of encoded information and switched system. The availability host based IDS are IDES, MIDAS and Haystack.

B. Network based IDS

Network based IDS gathers the information from network specifically in type of packets. These IDS are platform independent and simple to deploy to different frameworks. The available system based IDS are NSM, Bro. Based on criteria adopted for information analysis and preprocessing unit, IDS can be isolated into two classes:

C. Misuse or signature based IDS

Misuse or signature based IDS keep up a database of known attack signatures. The discovery of attack includes comparison of information from information collection unit and information put away in database. On the off chance that the similarity happens them attack signal get produced. The difficult task is to keep the database of signatures up to date. Signature based IDS perform well for attack whose signatures are in database however they are wasteful to distinguish zero day attacks. They additionally have low false alarm rate.

D. Anomaly based ID

Anomaly based IDS responds to anomaly behavior as characterized by some history of checked frameworks previous behavior or some already characterized profile of that framework. The framework matches the current profile with past profile, if there is any significant deviation, at that point that action is informed as an attack. These frameworks are fit for distinguishing zero day attack. The available anomaly based IDS are IDES, W and S, Comp Watch.

E. Based on criteria adopted for creating the response, IDS can be separated into two classes:

1) *Passive IDS:* Passive IDS reacts to attacks by notified the best possible authority. They don't themselves attempt to moderate the malware done, or effectively try to harmor hamper the attacker. The available IDS in this class are IDES, GrIDS and NIDES
2) *Active IDS*: Active IDS reacts to attacks by starting certain activity. The activity can be against two elements, which additionally classifies Active IDS into sub classes. The available active IDS are EMERLARD and Janus. Those elements can be:
3) *Attacking system*: In this class, the IDS attempt to control the attacking system. They attempt to attack the attack ersystemto eliminate his stage of operation.
4) *Attacked system*: in this class, the IDS attempt to control the attacked system. They modified the state of attacked system to mitigate the attack. They can eliminate the network interconnections, maximize the security logging or kill the concerned processes etc.

III. INTRUSIONDETECTION TECHNIQUES

Previously, various types of procedures from different disciplines have been applied to recognize the intrusions. The strategies are a) Statistical Techniques b) Knowledge-based Techniques and c) Artificial Intelligence (AI) based Techniques. In the statistical based IDS, the nature of the framework is described from an arbitrary perspective. Then again, Knowledge based IDS procedures attempt to catch the claimed behavior from available framework information (protocol determinations, network traffic activities, and so on.). At long last, AI based IDS techniques depend on the foundation of an explicit or implicit model that permits the patterns examined to be categorized.

IV. ARTIFICIAL INTELLIGENCE BASED TECHNIQUES

Mario Castro Ponce has recorded many points of interest of AI based new patterns); Fast computing (quicker than people, really) also, Learning capacities. Many researchers have classified AI based detection approaches into various classes. The real classes include followings:

A. Decision Tree Based Approach

Choice trees are more powerful and well known software's for grouping and prediction. A decision tree is a tree that has three primary segments: nodes, curves and leaves. Every node is marked with a feature property which is most informative among the properties not yet considered in the way from the root, each arc segment out of anode is marked with a feature value for the node's feature and each leaf is marked with a category or class. A decision tree would then be able to be utilized to order an information

point by beginning at the base of the tree and traveling through it until a leaf node is come to. The leaf node would then give the classification of the information point. Levin made a collection of locally ideal decision trees from which ideal subset of trees is chosen for predicting new cases. 10% of KDD Cups database is utilized for preparing and testing. Information is randomly examined from the whole preparing informational collection. Multi class detection approach is utilized to recognize various attack classifications in the KDD informational collection. Much the same as Agarwal and Joshi [23], Levin tries to characterize the information into five unique classes: Normal, Probing, DOS, U2R, and R2L. The last trees give high detection rates for all classes incorporating the R2L in the whole preparing information set. The outcome of trial is appeared in table.

	Normal	Probe	DoS	U2R	R2L
Detection accuracy	99.42 %	84.52 %	97.5%	11.8%	7.32%
False positive rate	-	21.6%	73.1%	36.4%	1.7%

B. Parzen Window Based Approach

Yeung and Chow implements a novelty detection approach utilizing non-parametric density estimation depend on Parzen window estimators with Gaussian parts. KDD informational collection was utilized to prepare and test framework. The aggregate traffic was grouped into four classes. The outcome is appeared in table.

	Probe	DoS	U2R	R2L
Detection accuracy	99.2%	96.7%	93.6%	31.2%
False positive rate	No data available			

C. Rule Based Approac

Rule based approach commonly include the application of a set of association rules and random episode patterns to characterize the audit information. In this specific circumstance, if a rule states that "if event X happens, at that point event Y is probably going to happen", at that point events X and Y can be described as sets of (variable, value) sets where the goal is to discover the sets X and Y with the end goal that X "implies" Y. In the domain of classification, we settle Y and attempt to discover sets of X which are great indicators for the correct characterization. The benefit of utilizing rules is that they have a tendency to be straightforward and natural, unstructured and less rigid. As the disadvantages they are troublesome to keep up, and sometimes, are inadequate to describes too many sorts of data. Various inductive rule generation techniques have been proposed in literature review. A few of them initially build a decision tree and after that concentrate a collection of classification rules from the decision tree. Different calculations directly consists rules from the information by utilizing a dividend conquer approach.

D. Machine Learning Approach

1) Neural Network Approach: In the neural system approach to deal with intrusion detection, the neural system figures out how to predict the nature of the different clients and daemons in the framework. On the off chance that legitimately designed and executed, neural systems can possibly address huge numbers of the issues encountered by rule based methodologies. The fundamental advantage of neural systems is their resistance to imprecise information and unverifiable data and their ability to gather solutions from information without having earlier knowledge of the regularities in the information. This in mix with their ability to generalize from learned information has indicated made them a proper way to deal with intrusion detection. Keeping in mind the end goal to apply this way to deal with Intrusion Detection, we would need to present information representing to attacks and non-attacks to the Neural Network to change automatically coefficients of this System during the preparation stage. Neural systems can be utilized as a part of following ways:

2) Supervised Learning: The managed technique figures out how to map inputs to outputs utilizing the right values characterized by the manager. A feed forward neural network (FFNN) and recurrent neural network (RNN) are two essential techniques that utilization supervised learning. A multi-layered feed forward (MLFF) NN and radial basis function (RBF) are two cases of FFNNs. The estimation of the separation amongst inputs and the centers of hidden neurons is the premise of RBF classification. In examination with the MLFF back propagation (BP), the RBF is better for extensive information in light of the fact that it is faster.

3) Unsupervised Learning: There is no administrator in unsupervised learning, and it is prepared using unlabeled information only. Unsupervised learning is like a statistical clustering, in which they recognize different groups of sources of inputs utilizing their comparability. The self organizing maps (SOM) and the adaptive reverberation Theory (ART) are two cases of unsupervised learning. The SOM is an essential neural system technique utilized for the abnormality and misuse detection. Be that as it may, the performance analysis of ART and SOM based intrusion detection are looked at in , which demonstrates the higher detection analysis of ART on both online and offline data.

E. Clustering Based Approach: Clustering methods work by gathering the identified information into clusters, as per a given similarity or distance measure. The applicants strategies are distance measurement are Euclidean distance and Mahalanobis distance. Similarity can be measured by utilizing cosine calculation, double weighted cosine formula proposed by Sanjay Rawat and some more. The technique most regularly utilized for this consists in choosing a representative point for each cluster. At that point, each new information point is delegated having a place with a given cluster as indicated by the proximity to the relating reprehensive point. There exist no less than two ways to deal with clustering based anomaly detection. In the main approach, the anomaly recognition method is prepared utilizing unlabelled information that consists of both normal and also attack traffic. In the second approach, the model is prepared utilizing just normal information and a profile of ordinary action is generated. The thought behind the primary approach is that abnormal or attack information shapes a little level of the add up to information. In the event that this suspicion holds, anomalies and attacks can be recognized in view of cluster sizes—vast clusters relate to normal information, and whatever remains of the information points, which are anomalies, relate to attacks.

F. Multi Classifier Approach: Multi classifier approach has been supported by numerous specialists in literature review. They demonstrated that no single classifier is sufficiently skilled to recognize all classes of attacks to supportable false alarm rate and identification accuracy. Srinivas Mukkanmalla et al. examined the SVM, ANN, LGP and MARS for classification of KDD information into five classes. They utilized detection accuracy, attack seriousness, preparing and testing time (versatility) as evaluation measurements and utilized DARPA intrusion detection dataset. They demonstrated that none of single classifier performs well in all attack classes. LGPs outperform MARS, SVMs and ANNs as far as detection accuracy to the detriment of time. MARS is better than SVMs in regard classifying the most basic classes (U2Su and R2L) as far as the attack severity. SVMs beat ANNs are important parts of scalability (SVMs can prepare with a bigger number of examples, while ANNs set take a long opportunity to prepare or, then again neglect to join at all when the number of patterns gets huge); preparing time and running time; and forecast precision. The outcomes are outlined in table.

	Normal	Probe	DoS	U2R	R2L
LGP	99.64	99.86	99.90	64	99.47
SVM	98.42	98.57	99.11	64	97.33
MARS	99.71	56.57	99.60	28	98.93
NN (RBP)	99.57	92.71	97.47	48	95.02

G. Hybrid Systems: Numerous scientists have proposed that the checking capability of current intrusion detection frameworks can be enhanced by adopting a hybrid strategy that comprises of both anomaly and in addition signature detection systems. The anomaly detection strategy helps in the recognition of new or unknown zero day attacks while the signature identification procedure recognizes known attacks. Tsong Song Hwang et al has proposed a 3 level hybrid way to deal with detect intrusions. First level of framework is signature based approach to deal with filter the known attacks utilizing black list idea. Second level of framework is anomaly detector that uses the white list idea to recognize the ordinary and attacks traffic activity that has by passed first level. Third level part of framework uses the SVM to describe the unknown traffic activity into five classes i.e. normal, probing, DoS, U2R and R2L. KDD dataset was utilized to prepare and test the framework. The outcomes are outlined in table.

Class/detection accuracy	Old attacks	New attack	Total detection
Probing	99.92%	98.16%	99.16%
DoS	99.99%	18.03%	97.65%
U2R	20.57%	87.83%	76.32%
R2L	79.84%	26.94%	46.53%

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI November 2017- Available at www.ijraset.com

V. RESULTS AND DISCUSSION

It will describes normal artificial intelligence based IDSs as well as various high speed intelligent IDSs. Modern IDSs were evaluated utilizing IPv4 traffic activity. Migration to IPv6 gives new security challenges. In spite of the fact that, IPv6 is, in general, more secure than IPv4 due to the commanded IPsec, the transaction progress presents various security issues. These security dangers were considered in this part. Intrusion detection in different systems is as vital as in wired network systems. Artificial intelligence based IDSs are acceptable to numerous different environments as well. In this part, the utilization of Artificial intelligence in social grid computing, distributed computing, and remote systems was likewise considered.

VI. CONCLUSION

In spite of the fact that quickly, we represented by and large the security objectives also, possible attack abusing the security targets of any organization and classification of IDS in view of different processing segments. Subtle elements of different AI based strategies connected in intrusion detection are introduced. At last, a multi classifier approach is talked about that outcome into detection of known and unknown attacks with high accuracy and low false alarm rate. Every one of these methods examined above either prepare single finder to recognize different of classes of attacks or minimize the number of features to make the approach computationally sufficient. This compromization identification accuracy and false alarm rate. There is intense need to distinguish different features to recognize distinctive classes of attacks and productive AI based procedures that can ensure the frameworks. More research on productive and exact AI based classification procedures stays to be finished with the goal that adaptable and versatile security framework can be designed to adapt to consistently developing number of severe Internet attacks.

REFERENCES

[1] Pathan, A.-S. K., Lee, H.-W., and Hong, C. S. Security in wireless sensor networks: Issues and challenges. In Proc. Of IEEE ICACT ’06, Vol. II, Phoenix Park, Korea, (February 20- 22, 2006)
[2] Cam, H., Ozdemir, S., Muthuavinashiappan, D., and Nair, P. Energy efficient security protocol for wireless sensor networks. In IEEE 58th VTC 2003 Fall, 2003, 5 (October 6– 9, 2003)
[3] Cam, H., Ozdemir, S., Nair, P., Muthuavinashiappan, D., and Sanli, H. O. Energy-efficient secure pattern based data aggregation for wireless sensor networks. Com. Commun., 29, I.4, (2006)
[4] Yin, C., Huang, S., Su, P., and Gao, C. Secure routing for large-scale wireless sensor networks. In Proceedings of IEEE ICCT 2003, 2 (April 9–11, 2003)
[5] Hass, Z. J. Design methodologies for adaptive and multimedia networks. IEEE Communications Magazine, 39(11), (November 2001)
[6] Heinzelman, W. B., Chandrakasan, A. P., and Balakrishnan, H. An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. Wire. Commun., 1(4) (2002)
[7] M. Almgren and E. Jonsson. Tuning an ids - learning the security officer’s preferences. In 11th Nordic Workshop on Secure IT Systems - Nordsec 06, 2006.