ISSN: 2321-7782 (Online)
Volume 3, Issue 11, November 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

A Survey of Intrusion Detection Using Genetic K-Means Algorithm in Wireless Sensor Networks

Manali Mandanna[1]
Dept. of CSE
BMSCE
Bangalore, India

Kiran L[2]
Dept. of CSE
BMSCE
Bangalore, India

Madhavi R P[3]
Dept. of CSE
BMSCE
Bangalore, India

Abstract: ***A high level of security is required in the area of wireless sensor networks. Security in communication has become a major concern. The field of network security faces many challenges i.e. the ability to identify and prevent attacks on the network. Wireless sensor networks (WSN) consist of sensor nodes deployed in a manner to collect information about the surrounding environment. Their distributed nature, multi hop data forwarding and open wireless medium are the factors that make wireless sensor networks highly vulnerable to security attacks at various levels. An effective intrusion detection system can play an important role in identifying and preventing attacks which is needed to ensure the network against security breaches.***

Keywords: Network Security; Wireless Sensor Network; Intrusion Detection System; security attacks; genetic algorithm;

I. INTRODUCTION

A wireless sensor network is spatially distributed with autonomous sensors to monitor physical or environmental conditions such as temperature, sound, pressure, etc. and pass their data to a main location. The development of wireless sensor networks was motivated by military applications such as battlefield surveillance but today these networks are found in several industrial and consumer applications such as industrial process monitoring and control, machine health monitoring, etc.[1]. The wireless sensor networks are built from a few to several hundreds or even thousands of nodes where each node is connected to one or several sensors. Sensor nodes have the ability of self-healing and self-organizing.

Wireless sensor networks are deployed in harsh physical conditions and hostile environments where the nodes are exposed to damages as well as physical security risks. Characteristics of the nodes such as self-organizing nature, low battery power, limited bandwidth supply, with distributed operations using open wireless medium, multi hop traffic forwarding and dependency on other nodes are such characteristics of sensor networks that makes it vulnerable to security attacks to all levels of OSI layers [2,3].

a) Features of Wireless Sensor Networks

The features of wireless sensor networks [4] can be depicted below:

» Variable size of network – The network can be formed from a single node to thousands of nodes.

» Low cost – Since the number of nodes can be very high the cost of each node must be inexpensive.

» Long lifetime – The network must be designed with protocols so that the network is able to run efficiently from a long period of time.

» Self-Organization - Wireless sensor networks are formed through the connection of a large number of nodes which have the ability to organize themselves without external configuration.

» Data centric – The network is focused on collecting data efficiently where the data collected from one node does not overlap with the data that was collected from another node.

» Application specific – Each network is created for a specific application that all the nodes in the network are aware of.

» Cooperation amongst nodes – All the nodes must be able to cooperate with each other so that the data collected can be aggregated in a meaningful manner which contributes to the overall efficiency of the network.

» Multi hop routing protocols – Protocols have been implemented in the wireless sensor network so that the packets are routed efficiently with minimization in loss of energy and also maximize network life.

b) Security Issues in Wireless Sensor Networks

The several areas of security issues and goals are:

» Data confidentiality – This means that a sensor must be able to keep its data private and not leak it to any external or neighbouring network.

» Data integrity – The information that is collected from the sensors must remain unaltered and not be tampered by any external agent.

» Data authenticity – The source of the collected data must be from an authentic verified sensor and should not be injected into the network from an outside source.

» Robustness – The network must be robust and able to return to normal functioning with minimal damage from a security attack.

c) Security Attacks in Wireless Sensor Networks

According to the survey conducted by [5], wireless sensor networks are vulnerable to security attacks and these attacks can be classified under:

» Spoofed, altered or replayed routing information – This attack targets the routing information that is exchanged by the nodes. It involves creating unnecessary loops, generating false error messages, changing the routes used in the network, etc.

» Selective forwarding – Here the malicious nodes will decide not to send certain packets to its neighbours and simply drops the packets which prevents them from being propagated further. The simplest case is when the node decides to not forward any of the packets that it has received to its neighbours i.e. it behaves as a black hole.

» Sinkhole attacks – In this kind of attack the traffic in the network is lured through a compromised node to a particular part or area of the network which can be related to as a sinkhole. This kind of attack can enable other attacks on the network for example the selective forwarding attack.

» Sybil attacks – In the Sybil attack a malicious node tries to portray itself as multiple nodes so that it degrades the usage as well as the efficiency of the distributed algorithms that are used in the network.

» Wormholes – This attack involves the transfer of messages from part of the network to another part through a link that usually does not exist and convinces two nodes that they are neighbours even when they are not. This is similar to a wormhole and hence the name. This attack can be used along with the selective forwarding attack and also with eavesdropping.

» HELLO flood attacks – Many protocols use HELLO messages in order to establish a connection with its neighbours. This attack takes advantage of such a situation and a malicious node can flood a network with many HELLO messages which causes the nodes to think that its neighbours are compromised. This situation leads to a state of confusion in the network.

Several solutions have been proposed to solve these security issues like authentication, key exchange and secure routing or security mechanisms for specific attacks. These can prevent some of the attacks but it is not enough to protect the network from all of the security attacks. An Intrusion Detection System (IDS) is referred to as the second line of defense where it is capable of only identifying the threats but not prevent or respond to the attacks that it detects.

Intrusion detection systems use one of the two detection techniques available:

» Statistical anomaly-based IDS-An intrusion detection system that uses this technique will monitor network traffic and compare it with the established baseline. This baseline is used to define what is normal for this network and also the bandwidth and protocols used. The ports are devices are connected to each other and alert the system when anomalous traffic is detected or traffic which is different to that of the baseline defined. The issue this system can encounter is that it can give a false positive indication when there is legitimate use of bandwidth if the system is not intelligently configured.

» Signature-based IDS-A signature based intrusion detection system is also called as a rule-based system where it monitors the packets on the network and compares them to a database of signatures or attributes from known malicious attacks. The issue with this system is that there is a lag between a new threat being discovered and the signature being available in the database for comparison. During this lag the system is vulnerable to these new attacks.

Several schemes are available for the implementation of intrusion detection schemes [6] such as:

» ***Neural networks***

- Pros – It has the ability to generalize from noisy, limited and incomplete data. Expert knowledge is not required to be able to find out unknown or novel attacks.
- Cons – The training process is slow and is not practical for real time situations.

» ***Bayesian networks***

- Pros – Able to recognize probabilistic relationships among the variables under consideration and also has the ability to include prior known data.
- Cons – It is difficult to handle continuous features and also if the prior knowledge is wrong it may not contain good classifiers to handle the data.

» ***Support vector machines***

- Pros – It has a better learning ability for smaller samples, high training rate, and decision rate and is insensitive to the size of input data.
- Cons – A longer training period is required and it uses a binary classifier so it cannot provide additional information about the detected attack.

» ***Genetic algorithm***

- Pros – It can derive the best classification rules and selects the optimal parameters. It is biologically inspired and uses the evolutionary algorithm.
- Cons – It cannot insure constant optimization response times.

» ***Fuzzy logic***

- Pros – The reasoning employed in this technique is approximate rather than precise and it is effective against port scans and probes.
- Cons – It consumes a large number of resources and dynamic rule updating at runtime is difficult.

The above section gives an overview about the security issues that are faced by wireless sensor networks as well as the various techniques and schemes that can be employed to develop an intrusion detection system to help prevent attacks on the network. The next section discusses the various systems that were developed by others and the techniques that they have employed in their intrusion detection systems.

II. EXISTING WORK

In this section, we will concentrate on the literatures of several related research areas to Network Intrusion Detection Systems. An Intrusion Detection System (IDS) is a security system which implements the process of intrusion detection and reports the intrusion accurately to the appropriate authority. There are two general categories of intrusion detection systems (IDSs): misuse detection and anomaly detection [23]. The IDS monitors the packets from various network connections in order to detect an intrusive activity. If an intrusion is detected, the IDS simply logs in a message into system audit file to be analyzed later or to stop such connections to end an intruder's attack or perform some other action as defined by the organization's rules. False alarms could occur sometimes. [20, 12]

Many methods were used to implement an IDS namely, a GA-based method to detect anomalous network behaviors [8, 10, 11, 13, 15] or that uses GP to directly derive a set of classification rules from historical network data as done by W. Lu and I. Traore [7] or SNORT and GA were combined which would detect the network attacks by scanning each of the data packets [9] or Fuzzy logic was combined with GA so as to efficiently detect various types of network intrusions [12] and so on. A support-confidence framework was used as a fitness function in GP and GA based approaches. [7, 16] In a GA based IDS, a simple GA was employed to represent and derive rules from network audit data. The generated rules are used to classify the incoming network connections. Appropriate GA parameters were chosen based on a large number of experiments. [8, 11, 16, 23] In many of the methods, the standard dataset of KDD Cup 1999 was used to evaluate the performance of the method. [9, 10, 13, 15, 19, 21]

In the approach proposed by Lu *et al.* [7], the use of GP makes the implementation more difficult and more data or time was required to train the system. In the method proposed by Li [8], both quantitative and categorical features of network data are included when deriving classification rules using GA. The inclusion of quantitative features may lead to increased detection rates. However, no experimental results are available yet. Dave *et al.* [9] present an approach by combining traditional SNORT and Genetic Algorithm to reduce the detection time, CPU Utilization and memory utilization. To evaluate the performance of SNORTGA, the standard dataset of KDD Cup 1999 was used.

Anup Goyal and Chetan Kumar [10] used the GA algorithm that takes into consideration different features in network connections to generate a classification rule set. Each rule in rule set identifies a particular attack type. In the approach proposed by Xia, Hariri and Yousif [11], some network features can be identified with network attacks based on mutual information between network features and type of intrusions and then using these features, a linear structure rule and also a GA is derived. The approach seems very effective because of the reduced complexity and higher detection rate. The only problem is that it considered only the discrete features.

Mostaque Md. Morshedur Hassan [12] devised a method of applying genetic algorithms with fuzzy. The fitness of a chromosome is measured using a fuzzy confusion matrix. The proposed system can upload and update new rules to the system as new intrusions become known and hence it is cost effective and adaptive. Srinivasa K G *et al.* [14] presents IGIDS, where the

genetic algorithm is used for pruning best individuals in the rule set database. The search space of the resulting rule set is much compact when compared to the original rule and hence the process of decision making is faster. This makes IDS faster and intelligent. It exhibits high detection rate with low false positive rate.

B. Uppalaiah *et al.* [13] presents the Genetic Algorithm for the Intrusion detection system for detecting DoS, R2L, U2R, Probe from DD99CUP data set. Sunil *et al.* [15] proposed an approach that aims at gaining maximum detections of the Denial of Service attacks with minimum false positive rate. By using GA on KDD Dataset, the rule set was extended which would then be integrated with the network sniffer to detect denial of service attacks. This approach is very useful for the attack discovery in today's varying attack methodologies. However, the rules are not updated dynamically with the firewall's log data in this approach.

A A Ojugo *et al.* [16] proposed a genetic algorithm based approach, which is a software, driver or device. The software implementation is aimed at improving system security in networked settings allowing for confidentiality, integrity and availability of system resources. As a future work, other learning algorithms suitable for optimization can be implemented to achieve a secured environment for distributed computing. Vivek *et al.* [18] mainly concentrated on misuse detection system. It can find only those attacks whose matching rules are already stored in rule set. It collects data for audit which contains normal and abnormal data. After collecting the data, network sniffer will analyze the data and will send it to the genetic algorithm. After applying fitness function, rules are added to rule set which are stored in rule base.

K Marx *et al.* [17] designed and built an Intrusion Detection System (IDS) in JAVA that implements pre-defined algorithms for identifying the attacks over a network. The packets in the network are captured online. However, this system has a few limitations. This software does not completely shield the network from intruders. Although it is platform independent it was tested only on Windows XP. However, it can be employed and tested on various other machines which run different operating systems. It employs a log doesn't store the information about past sessions. This system just displays the log information but doesn't employ any techniques to analyze that information. Mohammad Sazzadul Hoque *et al.* [19] present a GA based approach. To measure the fitness of a chromosome the standard deviation equation with distance was used. If a better equation or heuristic is used in this detection process then the detection rate and process will improve, especially false positive rate will surely be much lower. The future work would be to improve this intrusion detection system with the help of more statistical analysis and with better and may be more complex equations.

V. Moraveji Hashmei, Z. Muda and W. Yassin [21] present a genetic algorithm based intrusion detection system. Software implementation of the proposed system is presented. The system is flexible enough to be used in different application environments, if proper attack taxonomy and proper training dataset exist. High detection rate and low false positive rates are the highlights of the proposed system. Sandhya and Anitha [22] proposed an approach for intrusion detection that employs genetic k-means algorithm. This algorithm is applied to differentiate normal and abnormal intrusion behavior and the rule base of intrusion detection is updated. A real-time intrusion detection rule base is set. It automatically detects the groups of similar objects in data training. High detection rate and low false positive rate are achieved.

III. Conclusion

Security plays an important role in the protection of the nodes that are connected in the network and also in safeguarding the data that is acquired by these nodes. This paper discusses the features of wireless sensor networks, the various security issues and the attacks that these networks are susceptible to. The techniques that can be employed to design intrusion detection systems have also been showcased along with the pros and cons of each technique. A survey on the various existing intrusion detection systems that have developed by others previously has been described. These systems try to solve one problem but encounter a new problem. As the security domain is ever increasing the research in this field evolves every single day to provide better and more efficient techniques to curb attacks on networks.

ACKNOWLEDGMENT

The work discussed in this paper is supported by the college through the Technical Education Quality Improvement Programme [TEQIP-II] of the MHRD, Government of India.

References

1. F. Akyildiz and I.H. Kasimoglu, "Wireless Sensor and Actor Networks: Research Challenges,"; Ad Hoc Networks, vol. 2, no. 4, pp. 351-367, Oct. 2004.
2. Scarfone, Karen; Mell, Peter (February 2007). "Guide to Intrusion Detection and Prevention Systems (IDPS)" (PDF). Computer Security Resource Center (National Institute of Standards and Technology) (800–94). Retrieved 1 January 2010.
3. I.F. Akyildiz, W Su, Y. Sankarasubramaniam and E Cayirci, "Wireless Sensor Networks, A Survey," Communication Magazine, IEEE, August 2002, Vol. 40, Issue 8, pp. 102-114.
4. Rajashree.V.Biradar , Dr. S. R. Sawant , Dr. R. R. Mudholkar , Dr. V.C .Patil, "Multihop Routing In Self-Organizing Wireless Sensor Networks" IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011.
5. Sushma , Deepak Nandal , Vikas Nandal, "Security Threats in Wireless Sensor Networks" IJCSMS International Journal of Computer Science & Management Studies, Vol. 11, Issue 01, May 2011
6. Jayveer Singh , Manisha J. Nene, "A Survey on Machine Learning Techniques for Intrusion Detection Systems" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2013
7. W. Lu and I. Traore, "Detecting New Forms of Network Intrusion Using Genetic Programming", Computational Intelligence, vol. 20, pp. 3, Blackwell Publishing, Malden, pp. 475-494, 2004.
8. W. Li, "A Genetic Algorithm Approach to Network Intrusion Detection", SANS Institute, USA, 2004.
9. Mit H. Dave, Dr. Samidha Dwivedi Sharma, "Improved Algorithm for Intrusion Detection Using Genetic Algorithm and SNORT", 2014.
10. Anup Goyal, Chetan Kumar, "GA-NIDS: A Genetic Algorithm based Network Intrusion Detection System", 2008.
11. T. Xia, G. Qu, S. Hariri, M. Yousif, "An Efficient Network Intrusion Detection Method Based on Information Theory and Genetic Algorithm", Proceedings of the 24th IEEE International Performance Computing and Communications Conference (IPCCC '05), Phoenix, AZ, USA, 2005.
12. Mostaque Md. Morshedur Hassan, "Network Intrusion Detection System Using Genetic Algorithm and Fuzzy Logic", 2013.
13. B.Upalhaiah, K. Anand, B. Narsimha, S. Swaraj, T. Bharat, "Genetic Algorithm Approach to Intrusion Detection System", 2012.
14. Srinivasa K G, S Chandra, S Kajaria, S Mukherjee, "IGIDS: Intelligent intrusion detection system using Genetic Algorithm", 2011.
15. Sunil Kumar, Surjeet Dalal, "Optimizing Intrusion Detection System using Genetic Algorithm", 2014.
16. A A Ojugo, A O Eboka, O E Okonta, R E Yoro, F O Aghware, "Genetic Algorithm Rule-Based Intrusion Detection System (GAIDS)", 2012.
17. Allam Appa Rao, P Srinivas, B Chakravarthy, K Marx, P Kiran, "A JAVA Based Network Intrusion Detection System (IDS)", 2006.\
18. Vivek K Kshirsagar, Sonali M Tidke, Swati Vishnu, "Intrusion Detection System using Genetic Algorithm and Data Mining: An Overview", 2012.
19. Mohammad Sazzadul Hoque, Md. Abdul Mukit, Md. Abu Naser Bikas, "An Implementation of Intrusion Detection System Using Genetic Algorithm", 2012.
20. Shaveta, Er. Abhinav Bhandari, Dr. Krishan Kumar Saluja, "Applying Genetic Algorithm in Intrusion Detection System: A Comprehensive Review", 2014.
21. V. Moraveji Hashmei, Z. Muda and W. Yassin, "Improving Intrusion Detection using Genetic Algorithm", 2013.
22. Sandhya G, Anitha Julian, "Intrusion Detection in Wireless Sensor Network Using Genetic K-Means Algorithm", 2014.
23. Ren Hui Gong, Mohammad Zulkernine, Purang Abolmaesumi, "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection", 2005.

