ISSN: 2321-7782 (Online)
Volume 4, Issue 3, March 2016
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Implementation 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: ***Security in communication has become a major concern. A high level of security is required in the area of wireless sensor networks. 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. Due to 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. Intrusion detection systems include pattern analysis techniques to discover useful patterns of system features. The derived patterns comprise inputs of classification systems, which are based on statistical and machine learning techniques. Clustering methods are used to detect unknown attacks. Elimination of insignificant features is essential for simplified, faster and more accurate detection of attacks. We present a conceptual framework for identifying attacks for intrusion detection by applying genetic k-means algorithm.***

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.

Security mechanisms can be divided into attack prevention, detection and recovery [2]. In order to remove the vulnerabilities of attack prevention methods, intrusion detection can be used as a second wall of defense. Generally, intrusion detection systems (IDS) can be categorized into anomaly detection and misuse detection based on their detection methods. One disadvantage of misuse detection is that it can only detect previously known attacks, based on their signatures; whereas in anomaly detection, attacks are detected through deviation from normal behavior [3].

Networks suffer from many attacks that can be classified into four main categories:

- DOS: Denial-of-service, e.g. syn flood;
- R2L: Unauthorized access from a remote machine, e.g. guessing password;

- U2R: Unauthorized access to local superuser (root) privileges, e.g., various ``buffer overflow'' attacks;
- Probing: Surveillance and other probing, e.g., port scanning.

The IDS should be able to deliver reliable detection outcomes. The detection methods have to be effective in identifying intrusions since poor detection performance will ruin the trustworthiness of the IDS. The IDS should survive in hostile environments. However, maintenance of high detection accuracy is challenging. IDS that employ attack signatures to detect intrusions cannot discover novel attacks. Protecting computers and applications becomes difficult as the number of new intrusions increases. Therefore, an effective detection methodology that is able to discover novel attacks is necessary for building reliable IDS.

II. EXISTING SYSTEM

Many methods were used to implement an IDS namely, a GA-based method to detect anomalous network behaviors [5, 7, 8, 10, 12] or that uses GP to directly derive a set of classification rules from historical network data as done by W. Lu and I. Traore [4] or SNORT and GA were combined which would detect the network attacks by scanning each of the data packets [6] or Fuzzy logic was combined with GA so as to efficiently detect various types of network intrusions [9] and so on. A support-confidence framework was used as a fitness function in GP and GA based approaches. [4, 13] 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. [5, 8, 13, 16] In many of the methods, the standard dataset of KDD Cup 1999 was used to evaluate the performance of the method. [6, 7, 10, 12, 14, 15]

In the approach proposed by Lu *et al.* [4], 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 [5], 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.* [6] 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.

Mostaque Md. Morshedur Hassan [9] 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.* [11] 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.

III. PROBLEM IDENTIFICATION

The characteristics of Wireless sensor network such as unreliable channel, dynamic topology, dependence on node routing mechanism, lack of monitoring and management center, etc., makes it vulnerable to various attacks. Intrusion detection in network security plays the important role of resistance against attacks. There are some prominent drawbacks of existing intrusion detection models, such as lack of adaptive ability, inability to detect new or unknown attacks, low detection rate and false positives rate. Several types of intrusions have to be detected from a large set of features. To address this need, methods for better classification of anomalies is needed. Feature selection analysis emphasizes the relation that exists between the type of feature and the kind of attacks detected. Hence feature-based diagnosis methodology is indispensable. Low detection rate, mainly due to the high false positive rate degrades the performance of intrusion detection. Low throughput attained due to the high data rates is also another drawback. Decision on a feature that is not related to the content will lead to unfair detection of an

attack. One of the important drawbacks of Intrusion Detection Systems is that if an intrusion is not identified or if it is falsely detected, it continues to make the same mistake every time as there is no possibility to change its behavior.

The objective of this paper is to introduce the design of an intrusion detection system that will be able to continuously learn about new attacks and also from the errors of the previous runs so that a better and more reliable system capable of adapting to any scenario based on the what the network has to offer can be applied to the network.

IV. PROPOSED METHODOLOGY

Our proposed method has the following modules: Genetic K-Means Algorithm for intrusion detection, Selection, Mutation, K-Means operator, Implementation of clustering and Identification of attacks.

A. *Genetic K-Means Algorithm for intrusion detection*

Genetic K-Means Algorithm (GKA) maintains a population of coded solutions. The population is randomly initialized and is evolved over generations. Population in the next generation is attained by applying genetic operators to the current population. The evolution continues till a terminating condition is achieved. The steps involved can be seen below.

- Selection: The selection operator involves random selection of a chromosome from the previous population according to the distribution given by

$$P(s_i) = \frac{F(s_i)}{\sum_{j=1}^{N} F(s_j)}$$

$F(s_i)$ represents fitness value of the string s_i in the population. Solutions in the current population are estimated based on their merit to survive in the next population. This requires that each solution in a population be associated with a fitness value. The fitness value of a solution string depends on the total within cluster variation. A solution string that has relatively small square error must have relatively high fitness value.

- Mutation: Mutation changes the value of solutions based on the distance between the cluster centroids and the corresponding data points. Each solution corresponds to a data point whose value represents the cluster to which that data point belongs. The probability of changing value of a cluster number is more if the corresponding cluster center is closer to data point.

- K-Means Operator: The k-means operator consists of two steps: (1) Firstly the cluster centers are calculated and (2) Each data point to the cluster is reassigned with the nearest cluster center.

- Implementation of clustering: The population is divided into clusters which are assigned sequence numbers 0,1, ..., (n-1). A cluster with sequence number m is denoted as (m). The N*N matrix is denoted by D = [d(i,j)]. The level of kth clustering is given as L(k). The proximity between clusters (p) and (q) is denoted as d [(r),(s)]. We begin with the disjoint clustering having level L(0) = 0 and sequence number m = 0. Then determine the least dissimilar pair of clusters in the current clustering, say for pair [(p),(q)], according to d[(p),(q)]=min d[(i),G)], where the minimum is the overall pair of clusters in the current clustering. Increment the sequence number according to m = m + 1 and hence merge the clusters (p) and (q) into a single cluster so as to form the next clustering m. The level of this clustering is set to L(m) = d[(p),(q)]. The proximity matrix D is updated by deleting the rows and columns corresponding to clusters (p) and (q). A row and column corresponding to the newly formed cluster is added. The proximity between the new cluster, denoted by (p,q) and the old cluster (k) is defined as below:

$$d[(k), (p,q)] = \min d[(k),(P)], d[(k),(q)]$$

V. EXPERIMENTAL RESULTS AND ANALYSIS

Java in NetBeans IDE 7.2.1 was used for implementing the genetic k-means algorithm to detect intrusions in wireless sensor network. After applying selection, mutation and K-Means operator, the clusters are chosen for detecting the intrusion as shown in Fig. 1.

The effectiveness of our proposed intrusion detection system using Genetic K-Means algorithm is evaluated using the detection rate. Fig. 3. shows that the Genetic K-Means algorithm has the higher detection rate when compared to K-Means algorithm. [17]

Fig. 2. Shows the K-Means and Genetic K-Means metrics evaluation tables. Our system also showed low false positive rate as the false alarms detected remain minimum and it also shows a low false negative rate which are shown in Fig. 3.

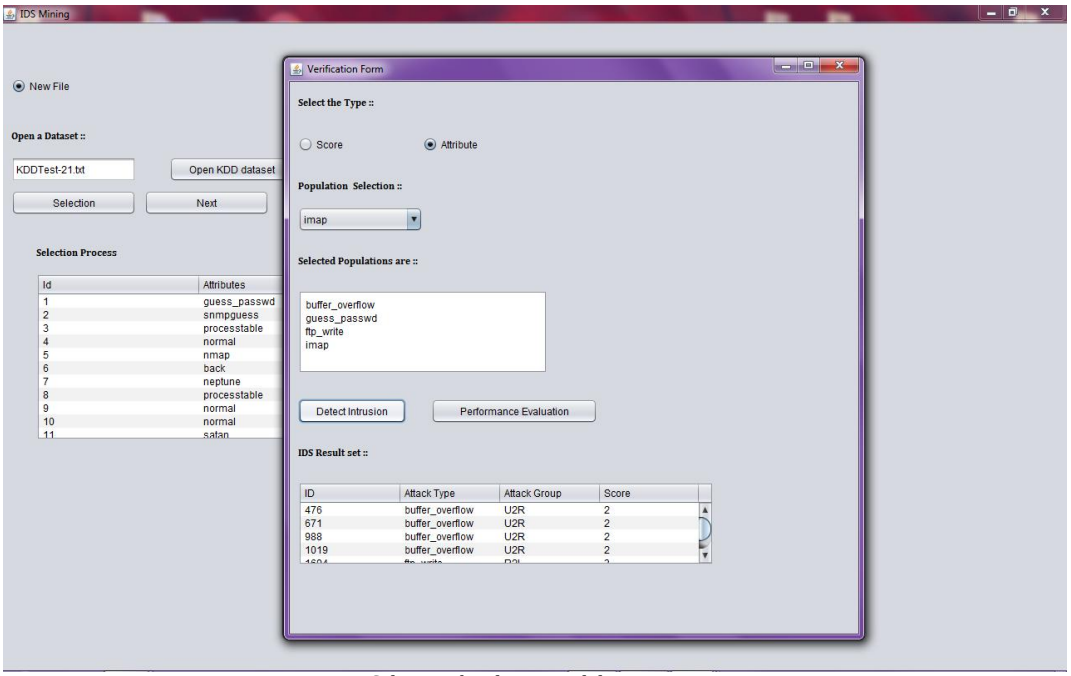

Fig. 1 Selecting the clusters and detecting intrusion.

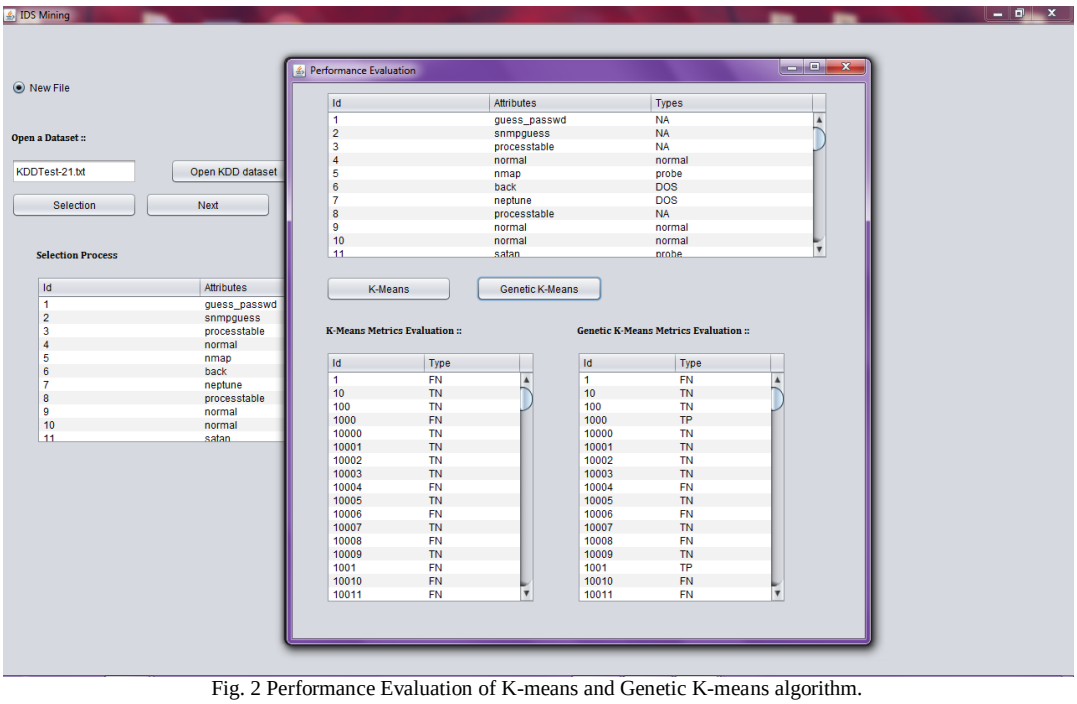

Fig. 2 Performance Evaluation of K-means and Genetic K-means algorithm.

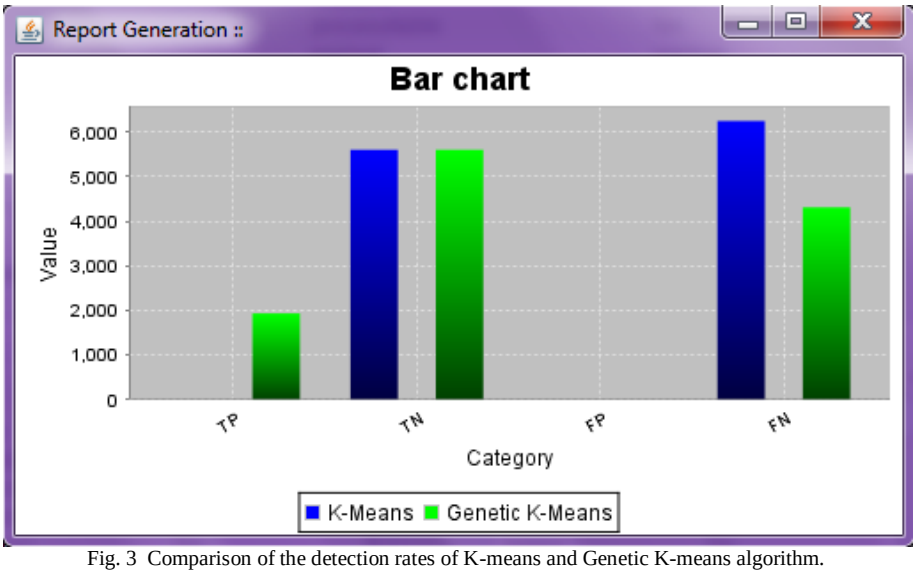

Fig. 3 Comparison of the detection rates of K-means and Genetic K-means algorithm.

VI. 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 problems faced with respect to the security of a network, the various security issues and the attacks that these networks are susceptible to. We have discussed a technique that can be employed to design an effective intrusion detection system. It solves many issues faced in some other existing systems like the high rate of false positives and also has a higher rate of detection. Our proposed methodology is more suitable for dynamic networks. It can also detect new attacks without depending on the signatures already known to the system based on genetic k-means algorithm.

Acknowledgement

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. Stallings, W., "Cryptography and network security: principles and practice", Prentice Hall, 2009.
3. Timmis, J.; Lemos, R.; Ayara, M.; Duncan, R., "Towards Immune Inspired Fault Tolerance in Embedded Systems", 9th International Conference on Neural Information Processing, pg. 1459-1463, IEEE, 2002.
4. 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.
5. W. Li, "A Genetic Algorithm Approach to Network Intrusion Detection", SANS Institute, USA, 2004.
6. Mit H. Dave, Dr. Samidha Dwivedi Sharma, "Improved Algorithm for Intrusion Detection Using Genetic Algorithm and SNORT", 2014.
7. Anup Goyal, Chetan Kumar, "GA-NIDS: A Genetic Algorithm based Network Intrusion Detection System", 2008.
8. 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.
9. Mostaque Md. Morshedur Hassan, "Network Intrusion Detection System Using Genetic Algorithm and Fuzzy Logic", 2013.
10. B.Upalhaiah, K. Anand, B. Narsimha, S. Swaraj, T. Bharat, "Genetic Algorithm Approach to Intrusion Detection System", 2012.
11. Srinivasa K G, S Chandra, S Kajaria, S Mukherjee, "IGIDS: Intelligent intrusion detection system using Genetic Algorithm", 2011.
12. Sunil Kumar, Surjeet Dalal, "Optimizing Intrusion Detection System using Genetic Algorithm", 2014.
13. A A Ojugo, A O Eboka, O E Okonta, R E Yoro, F O Aghware, "Genetic Algorithm Rule-Based Intrusion Detection System (GAIDS)", 2012.
14. Mohammad Sazzadul Hoque, Md. Abdul Mukit, Md. Abu Naser Bikas, "An Implementation of Intrusion Detection System Using Genetic Algorithm", 2012.
15. V. Moraveji Hashmei, Z. Muda and W. Yassin, "Improving Intrusion Detection using Genetic Algorithm", 2013.
16. Ren Hui Gong, Mohammad Zulkernine, Purang Abolmaesumi, "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection", 2005.
17. Shi Zhong, Taghi and Naeem Seliya, "Clustering-Based Network Intrusion Detection," International Journal of Reliability, Quality and Safety Engineering, Vol. 14, No. 2, pp. 169-187,2007.

