

A Prototype for Intrusion Detection in Wireless Sensor Networks Using Data Mining Methods

Asha R N¹, Dr Venkatesan S²

¹Asst. professor, Computer Science & Engineering, Global Academy of Technology, Bangalore

²professor & Dean, Computer Science & Engineering, Dayanana d Sagar College of Engineering, Bangalore

¹harshaaru@gmail.com, ²selvamvenkatesan@gmail.com

Abstract- The Wireless Sensor Networks (WSNs) are highly distributed networks of tiny, light-weight wireless nodes, placed in large numbers to monitor the environment or system. Monitoring the system includes the measurement of physical parameters such as pressure, temperature, relative humidity and passing their data to the main node (sink). WSN faces various security attacks which can affect the overall performance and security of the system. So, it is necessary to detect and prevent the attacks on WSN. Intrusion Detection is one of the major and efficient method against attacks. Intrusion Detection Systems can act as a second line of defence and it provides security primitives to prevent attacks against computer networks. This paper focuses on a hybrid approach for intrusion detection system (IDS) based on data mining techniques. The approach is clustering analysis with the aim to improve the detection rate and decrease the false alarm rate.

Keywords—Intrusion detection system; data mining; clustering; k-means; ensemble; detection rate; false alarm rate.

I. INTRODUCTION

Security of wireless sensor networks is becoming an important issue, as network attacks have increased in number over the past few years. It is essential to find an effective way to protect it as more sensitive and confidential information is being stored and manipulated online. The network based attacks can also be referred as some kind of intrusion. An intrusion can be defined as “any type of attack or action of misuse that attempt to compromise the confidentiality, integrity, availability, or availability of any resource”. To detect and control intrusions, intrusion detection systems are introduced.

An Intrusion Detection System [2] is a defense system that plays an important role to secure or protect a network system and its main purpose is to monitor network activities automatically and to detect malicious attacks and intruders. Intrusion detection system (IDS) is increasingly becoming a major and important component to secure the network in today’s world of Internet and wireless sensor networks.

Intrusion Detection Systems are divided into two types [1,3,4] according to the detection approaches: Misuse Detection and Anomaly Detection.

A. Anomaly detection

Anomaly detection defines the expected behavior of the network or profile in advance. Any significant behavior from such defined expected behavior are reported as possible attacks. But not all such deviations are attacks. The main advantage of this approach is that it can examine unknown and more complicated intrusions. The

shortcoming of this approach is its high false alarm rate and low detection rate.

B. Misuse detection

Misuse detection works on pattern which is predefined for malicious behavior and then identify intrusion based on this known pattern i.e. it finds intrusions by looking for activity corresponding to known techniques for intrusions. The shortcoming of this approach is that it can only detect intrusions that follow predefined patterns. The main advantage of misuse detection is its higher detection accuracy to all known attack.

Between these two methods[6,7,9] only anomaly detection has the ability to detect unknown attacks, since misuse detection can only detect intrusions which contain known predefined patterns of attack.

Clustering techniques in data mining [8] can be useful for detecting intrusions from network data, since clustering methods can discovers intrusions over a different time period. It is the process of assigning the data into groups of similar objects and each group is called as cluster. Each group consists of members or nodes from the same cluster that are similar and members from the different clusters are different from each other. Clustering is an unsupervised machine learning mechanism for discovering patterns and deals with unlabeled data with many dimensions.

Anomaly based IDS have the ability to detect new attacks, as any attack will differ from the normal activities. In order to detect attacks, a number of clustering based detection methods has been proposed.

Yu Guan and Ali A. Ghorbani, Nabil Belacel proposed a Y-means clustering algorithm [6] which overcomes two shortcomings of k-means: that is number of clusters dependency and degeneracy. It partitioned a data set into an appropriate number of clusters automatically.

A clustering algorithm that uses K-Means [11] for intrusion detection was proposed in which when the SOM finish its training process, K-means clustering is adopted to refine the weights obtained by training, and when SOM finish its cluster formation, K-means is applied to refine the final result of clustering.

K-means [10] is one of the simple portioning algorithms that solve the clustering problem. The f K-means algorithm follows a very simple and easy way to classify a given data set through a certain number of k clusters that are fixed a prior.

A hybrid learning approach [13] by using a combination of K-means and naive bayes classification, cluster all data based on the parameters into the corresponding group before applying a classifier for classification purpose.

A hybrid anomaly detection system [14] was proposed which combine k-means, and two classifiers: k-nearest neighbor and naive bayes. First, the feature selection process has to be covered from the intrusion data using an entropy based feature selection algorithm which selects the major attributes and removes the redundant attributes. The next step is cluster formation using k-Means and then it further classifies them by using a hybrid classifier.

To improve the performance of IDS and to achieve high accuracy and detection rate as well as low false alarm rate an Intrusion Detection System with clustering ensemble is proposed.

This paper presents a clustering algorithm for unsupervised anomaly detection. The proposed method is based on K-means clustering, which is a typical clustering algorithm. It overcomes the drawbacks of K-means thereby dealing with noise and outliers and generates the strategy for calculating the number of the cluster centroid and choosing the appropriate initial cluster centroid automatically

The rest of the paper is organized as follows. Section II presents proposed system architecture. A brief discussion on drawbacks of previous methods and advantages of proposed method is given in Section III. Finally, Section IV presents conclusion and future work.

II. PROPOSED METHOD

This section describes the system architecture for intrusion detection system (IDS) based on hybrid data mining techniques.

Clustering is a process of grouping and labeling data and assigning that data into groups of similar objects. Each group is called as cluster. It consists of members from the same cluster that are similar and members from the different clusters that are different from each other.

The proposed method is based on K-means clustering, which is a typical clustering algorithm. It overcomes the drawbacks of K-means thereby employing a hybrid approach.

K-means is one of the simplest unsupervised learning clustering algorithms. Its procedure follows a simple way to classify a given data set through a certain number of k clusters that are fixed a priori. First locations k center (c_1, \dots, c_k) are initialized. Then each data point x_i is assigned to its nearest cluster centre c_i . The positions of the k centers are recalculated until the centers no longer move. Each cluster centre c_i is updated as the mean of all data point x_i that has been assigned closest to it.

Fig. 1 depicts the proposed system architecture for intrusion detection. It consists of components such as data set for intrusion and clustering with k-means algorithm, behavioral analysis for intrusion and detection rate.

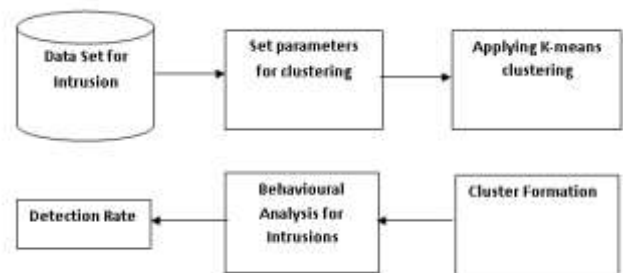


Fig. 1. System Architecture Proposed for IDS

Selecting parameters is important if the data set consist of large number of attributes. It consists of selecting parameters using an information gain feature selection method which selects the important attributes from the data set. It calculates the sum of the distance of each point from every other point and also calculates the sum of the average distance. For any point, if the distance is greater than the average distance then the center selected point it is considered as an outlier and it is removed from the data set.

After the clusters is formed using K-means algorithm.

The clusters that are formed by running the K-means algorithm are divided and merged again. By dividing and merging the clusters the number of k cluster centroids is calculated. The density of each point is calculated in filtered dataset to choose the appropriate initial centroids. These points are sorted as their density in descending order.

Then the k points with the larger density are selected as the initial centroids. Since the single clustering algorithm is difficult to get the great effective detection, the clustering ensemble is introduced by varying the value of k for the effective identification of attacks to achieve high accuracy and detection rate as well as low false alarm rate.

III. DISCUSSION

. This section discusses the limitations of previous existing methods and advantages of proposed method over them.

A Y-means clustering algorithm [8] has low false alarm rate and better detection rate. But it cannot solve real time anomaly detection, since it cannot update the data set dynamically during the process.

The major drawback of k-means m are its sensitivity to initial conditions such as the number of partitions and the initial centroids, and it is also sensitive to outliers and noise.

The major advantages of K-means [11] are that it is a lightweight, fast iterative algorithm which is easy to understand and implement.

A parallel clustering ensemble algorithm [12] forms the clusters more efficiently to mass data. It also achieves high detection rate but its false alarm rate is low.

Hybrid learning approach [13] by using K-means clustering and naive bayes classification overcomes the drawback of high false alarm rate and moderate detection rate.

With the aim to improve intrusion detection rate and decrease false alarm rate, this paper presents a hybrid learning approach for intrusion detection system. Feature selection helps in selecting important and relevant features from the data set and reduces the time required to process the data set.

Since the single clustering algorithm is difficult to get the great effective detection, clustering ensemble is employed for the effective identification of both known and unknown patterns of attacks to achieve high accuracy and detection rate as well as low false alarm rate.

IV. CONCLUSION

A hybrid data mining approach for intrusion detection system is proposed in this paper. The main research method is clustering analysis with the aim to achieve high detection rate for intrusion detection and very low or no false alarm rate.

Feature selection selects the important attributes from the data set. By the more better and accurate method of finding initial k clustering centers or nodes, the intrusion detection model with clustering ensemble is presented to achieve high accuracy and detection rate as well as very low false alarm rate.

The future work is the implementation of this approach and its comparison with existing methods.

REFERENCES

- [1] Snehlata S. Dongre and Kapil K. Wankhade, "Intrusion Detection System Using New Ensemble Boosting Approach", In International Journal of Modeling and Optimization, Vol. 2, No. 4, August 2012, pp 488-492.
- [2] Z. Muda, W. Yassin, M.N. Sulaiman, N.I. Udzir, "Intrusion Detection based on K-Means Clustering and OneR Classification", In Proceedings of 7th International Conference on Information Assurance and Security (IAS), IEEE, 2011, pp.192-197.
- [3] Kapil Wankhade, Mrudula Gudadhe, Prakash Prasad, "A New Data Mining Based Network Intrusion Detection Model", In Proceedings of International Conference on Computer and Communication Technology (ICCT 2010), IEEE, 2010, pp.731-735.
- [4] Z. Muda, W. Yassin, M.N. Sulaiman, N.I. Udzir, "Intrusion Detection based on K-Means Clustering and Naïve Bayes Classification", In Proceedings of 7th International Conference on IT in Asia (CITA), IEEE, 2011.
- [5] Kapil Wankhade, Sadia Patka, Ravindra Thool, "An Overview of Intrusion Detection Based on Data Mining Techniques", In Proceedings of 2013 International Conference on Communication Systems and Network Technologies, IEEE, 2013, pp.626-629.
- [6] Yu Guan and Ali A. Ghorbani, Nabil Belacel, "Y-Means: A Clustering Method For Intrusion Detection", In Proceedings of Canadian Conference on Electrical and Computer Engineering, Montreal, Quebec, Canada, May 4-7, 2003, IEEE, 2003, pp.1083-1086.
- [7] Yang Zhong, Hirohumi Yamaki, Hiroki Takakura, "A Grid-Based Clustering for Low-Overhead Anomaly Intrusion Detection", IEEE, 2011, pp.17-24.
- [8] WANG Huai-bin, YANG Hong-liang, XU Zhi-jian, YUAN Zheng, "A clustering algorithm use SOM and K-Means in Intrusion Detection" In Proceedings of 2010 International Conference on E-Business and E-Government, IEEE, 2010, pp.1281-1284.
- [9] Deepthy K Denatiou, Anita John, "Survey on Data

-
- Mining Techniques to Enhance Intrusion Detection”, In Proceedings of International Conference on Computer Communication and Informatics (ICCCI - 2012), Jan. 10 – 12, 2012, Coimbatore, INDIA, IEEE, 2012.
- [10] Hongwei Gao, Dingju Zhu, Xiaomin Wang, “A Parallel Clustering Ensemble Algorithm for Intrusion Detection System” In Proceedings of 2010 Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science, IEEE, 2010, pp.450-453.
- [13] Z. Muda, W. Yassin, M.N. Sulaiman, N.I. Udzir, “Intrusion Detection based on K-Means Clustering and Naïve Bayes Classification”, In Proceedings of 7th International Conference on IT in Asia (CITA), IEEE, 2011.
- [14] Hari Om, Aritra Kundu, “A Hybrid System for Reducing the False Alarm Rate of Anomaly Intrusion Detection System”, In Proceedings of 1st Int’l Conf. on Recent Advances in Information Technology (RAIT-2012),IEEE, 2012.