International Journal of Innovative Research in Advanced Engineering (IJIRAE)
Volume 1 Issue 1 (March2014)

A Survey on Wireless Intrusion Detection using Data Mining Techniques

R.Venkatesan
dr_san_phd12@outlook.com

Abstract - An Intrusion Detection System (IDS) is a system for detecting intrusions and reporting to the authority or to the network administration. In recent years, since the computer network keeps expanding drastically, the incidents of data theft and hacking are also increased. In order to control and monitor such activity that violates the security policy of system, it is necessary to have intrusion detection system (IDS) in a network. Intrusions are commonly high in wireless network when compared to wired networks, irrespective of wireless networks like Wireless sensor network (WSN)/Wireless ad-hoc network (WAHN) the vulnerability is very high because there is no perimeter security as defined or stated in wired networks. Data mining techniques have been successfully applied in many fields like Network Management, Education, Biology, Management studies, Production engineering, Process control, and Fraud Detection. The objective of this survey paper is to discuss the various methods which are being used for wireless intrusion Detection in WSN & WAHN based on Data mining concepts.

Keywords: Data Mining, Intrusion Detection System (IDS), Wireless IDS, Network Security, Misuse Detection, Anomaly Detection

1. INTRODUCTION

In recent years, there are many research articles are published for IDS using Data mining techniques [1] – [6]. Many researchers are focusing to use Data Mining concepts for Intrusion Detection, because the range of data they need to analyze is very huge and the necessity to do such analysis has become vital with respect to network security policies and standards framed by the concerned organization. Data mining is a process to extract the implicit information and knowledge from a large database. The intrusion detection in an information system or a process used to identify intrusions or any violations in the network security. In this survey paper, we are going to study the Data Mining approaches which are being followed to detect intrusion in a wireless network.

1.1 DATA MINING

The Knowledge Discovery in Databases (KDD) Process is used to denote the process of extracting useful knowledge from large data sets. Data mining is the process of discovering interesting patterns (or knowledge) from large amounts of data. The data sources can include databases, data warehouses, the Web, any other information repositories, or data that are streamed into the system (dynamically). Data can be associated with *classes* or *concepts* that can be described in summarized, concise, and yet precise, terms, such descriptions of a concept or class are called *class/concept descriptions.* These descriptions can be derived via Data Characterization and Data Discrimination. Data describes the actual state of the world and the Knowledge describes the structure of the world and consists of principles, directives and rules. The KDD process involves a number of steps and is often interactive, iterative and user-driven decision making rules [7]. Data mining is the most vital step in the KDD process, and it applies *data mining techniques* to extract patterns from the data.

Know the application domain: to understand the back ground of the knowledge and to specify the goal.
Data Collection: includes creating a target dataset which is relevant to the analysis
Data Mining: applying an appropriate algorithm to extract useful information using techniques.
Data Interpretation: to understand the discovered patterns and to confirm the goal is achieved.
Knowledge Representation: the final stage of representing the discovered knowledge.

Data mining functionalities are used to specify the kind of patterns to be found in data mining tasks and it can be classified into two categories:

- *Descriptive*: to characterize the general properties of data in the database
- *Predictive*: to perform inference on data and to make predictions

1.2 INTRUSION DETECTION

The Intrusion Detection concept was introduced by Anderson J. in 1980 [8]. Intrusion detection systems (IDSs) are usually deployed along with other preventive security mechanisms, such as access control and authentication, as a second line of defense that protects information systems. The IDSs may be classified into (i) Host-base IDSs (ii) Distributed IDSs and (iii) Application based IDSs.

Denning D.E in 1986 [9] presented the first intrusion detection model, which has six main components: *subjects*, *objects*, *audit records*, *profiles*, *anomaly records*, and *activity rules*.

- *Subjects* refer to the initiators of activity in an information system; they are usually normal users.
- *Objects* are the resources managed by the information system, such as files, commands and devices.

- *Audit records* are those generated by the information system in response to actions performed or attempted by subjects on objects. Examples include user login, command execution, etc.
- *Profiles* are structures that characterize the behavior of subjects with respect to objectives in terms of statistical metrics and models of observed activity.
- *Anomaly records* are indications of abnormal behaviors when they are detected.
- *Activity rules* specify actions to take when some conditions are satisfied, which updates the profiles, detects abnormal behaviors, relate anomalies to suspected intrusions, and produce reports.

1.3 WIRELESS INTRUSION DETECTION

The concept of Wireless intrusion detection should be redefined when compared with wired network IDS. The privacy settings like WEP/WPA will characterize the differentiation between the network types (wired/wireless network).The characteristics of wireless networks like, it can be used to cover an area, but not limited like wired networks. The chances are very high for the intruder to stay in a covered area and unseen in the network. There are many differences between wired or network IDS and wireless IDS, because the insider and outsider attacks need to defined wired networks and the same should be redefined for wireless environment. Moreover, there is no clarification about the limits/border for internal and external network. Some of specific threats in wireless environment for our analysis:

1. *Unauthorized access (rogue) points*
2. *Unauthorized use of service*
3. *DOS attack*
4. *MAC spoofing*
5. *Eavesdropping*
6. *War Driving*
7. *Null Association*
8. *Flooding and etc,*

In general, attacks are designed to steal the login credentials of the wireless network or to break the association between nodes. Usually intrusion are detected either real time or after the attack. Data Mining for IDS is the technique which can be used to identify unknown attacks and to raise alarms when security violations are detected. If the privacy settings WEP/WPA are weak (password/pass code settings) in a wireless network (IEEE802.11), then it will be an easy task to the intruders to access the network and leads to information theft or misusing the service. The scenario will be entirely different when it comes to WSN/WAHN because the possibilities of vulnerability are extremely high. Though the WSN/WAHN has been constructed on basis of Topology control (both centralized and distributed environment) [10] we don't sophisticated IDS to stop the intrusion. The vulnerability and different types of attacks in WSN is discussed in the literature [11].

1.4 RELATED WORK IN WIRELESS INTRUSION DETECTION USING DATA MINING TECHNIQUES

Many researches applied/implemented data mining techniques to design/model IDS. The detailed reports of such developments can be studied in the literatures [12] - [15].

2. INTRUSION DETECTION TECHNIQUES

The intrusion detection techniques based upon data mining [16] - [19] are generally falling into one of two categories: *anomaly detection* and *misuse detection*. The signatures of some attacks are known, whereas other attacks only reflect some deviation from normal patterns. A study was done on IDS frameworks (wired network) for data mining in the literature [20]. The article [21] states the basics of IDS and its correlation methods.

2.1 ANOMALY DETECTION

Anomaly detection attempts to determine whether deviation from an established normal behavior profile can be flagged as an intrusion [1].Anomaly detection consists of first establishing the normal behavior profiles for users, programs, or other resources of interest in a system, and observing the actual activities as reported in the audit data to ultimately detect any significant deviations from these profiles. Most anomaly detection approaches are statistical in nature. Anomaly detection may be divided into static and dynamic anomaly detection. A static anomaly detector is based on the assumption that there is a portion of the system being monitored that does not change. The static portion of a system is the code for the system and the constant portion of data depends upon the correct functioning of the system. Dynamic anomaly detection typically operates on audit records or on monitored networked traffic data. An audit record of operating systems does not record all events; they only record events of interest. Strength of anomaly detection is its ability to detect previously unknown attacks.

2.2 MISUSE DETECTION

Misuse detection works by searching for the traces or patterns of well-known attacks. Lee et al. [18] designed a signature-based database intrusion detection system (DIDS) which detects intrusions by matching new SQL statements against a known set of transaction fingerprints. Misuse detection is considered complementary to anomaly detection.

The rationale is that known attack patterns can be detected more effectively and efficiently by using explicit knowledge of them. This system usually searches for patterns or user behavior that matches known intrusion or scenarios, which are stored as signatures. If a pattern match is found, it signals an event then an alarm is raised. But it is unable to detect new or previously unknown intrusion. Pattern, Data mining, and state transition analysis are some of the approaches of misuse detection. To perform this detection method, each scenario need to described or modeled. Misuse detection is based on extensive knowledge of patterns associated with known attacks provided by human experts.

3. DATA MINING APPROACHES

Data mining generally refers to the process of (automatically) extracting models from large stores of data [22]. The recent rapid development in data mining has made available a wide variety of algorithms, drawn from the fields of statistics, pattern recognition, machine learning, and database. There are several types of algorithms [22] which are particularly related to intrusion detection.

- **Classification**: classifies a data item into one of several pre-defined categories. These algorithms normally output "classifiers". An ideal application in intrusion detection would be to gather sufficient "normal" and "abnormal" audit data for a user or a program, then apply a Classification algorithm to learn a classifier that can label or predict new unseen audit data as belonging to the normal class or the abnormal class;
- **Link analysis**: determines relations between fields in the data base records. Correlations of system features in audit data, for example, the correlation between command and argument in the shell command history data of a user, can serve as the basis for constructing normal usage profiles.
- **Sequence analysis**: models sequential patterns. These algorithms can discover what time-based sequences of audit events are frequently occurring together. These frequent event patterns provide guidelines for incorporating temporal and statistical measures into intrusion detection models.

3.1 ASSOCIATION RULE OR DEPENDENCY MINING

Association analysis is the discovery of association rules showing attribute – value conditions that occur frequently together in a given set of data. Association analysis widely used in transaction data analysis. This approach work on data dependency, in which one item is modify another item refer with this also modify. The concept of Association mining is to find all co-occurrences relationship called associations. Association Mining has been used in various domains and many efficient algorithms, extensions and applications have reported. In general, Association analysis has been considered as an unsupervised technique, so it can be applied for KDD task. *Apriori Association Rules Algorithm* is widely used for mining frequent item sets for Boolean association rules [23]. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent item set properties.

3.2 CLASSIFICATION

Classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of training set of data containing observations (or instances) whose category membership is known. An algorithm that implements classification, especially in a concrete implementation, is known as a *classifier* [24]. The term "classifier" sometimes also refers to the mathematical function, implemented by a classification algorithm that maps input data to a category. Classification can be thought of as two separate problems – Binary and Multiclass classification In binary classification, a better understood task, only two classes are involved, whereas multiclass classification involves assigning an object to one of several classes [25]. Since many classification methods have been developed specifically for binary classification, multiclass classification often requires the combined use of multiple binary classifiers. This is supervised learning. The class will be predetermined in training phase.

3.3 CLUSTERING

The clustering technique is used to map data items into groups according to similarity or distance between them. There are many clustering methods available, and each of them may give a different grouping of a dataset. The choice of a particular method will depend on the type of output desired. In general, clustering methods may be divided into two categories based on the cluster structure which they produce.

3.3.1 NON – HIERARCHICAL

The non-hierarchical methods divide a dataset of N objects into M clusters, with or without overlap. These methods are sometime divided into partitioning methods, in which the classes are mutually exclusive, and the less common clumping method, in which overlap is allowed. Each object is a member of the cluster with which it is most similar however the threshold of similarity has to be defined.

3.3.2 HIERARCHICAL – CONNECTION ORIENTED

The hierarchical methods produce a set of nested clusters in which each pair of objects or clusters is progressively nested in a larger cluster until only one cluster remains. The hierarchical methods can be further divided into agglomerative or divisive methods. In agglomerative methods, the hierarchy is build up in a series of N-1 agglomerations, or Fusion, of pairs of objects, beginning with the un-clustered dataset. The less common divisive methods begin with all objects in a single cluster and at each of N-1 steps divide some clusters into two smaller clusters, until each object resides in its own cluster. Some of the important Data Clustering Methods are described below.

4. CONCLUSION

As the wireless sensor network/ wireless ad – hoc networks keeps on increasing every day, we need to have efficient and effective IDS to monitor the networks and to identify, if there is any security violation breached. Data mining techniques for IDS are capable of extracting patterns automatically and adaptively from a large dataset. Various methods related to intrusion detection system are studied briefly. This survey paper states the methods and techniques of data mining to aid the process of Intrusion Detection in wireless environment. The concept intercepting these two different fields gives more scope for the research community to work in this area and to built new IDS and new threads are increasing along with the network expansion.

REFERENCE

[1] W. Lee and S. J. Stolfo. Data mining approaches for intrusion detection. In Proceedings of the 7th USENIX Security Symposium, San Antonio, TX, January 1998.
[2] W. Lee, S.J. Stolfo, K.W. Mok, Algorithms for Mining System Audit Data, in Proc. KDD, 1999.
[3] Sekeh.M.A,Bin Maarof.M.A, "Fuzzy Intrusion Detection System Via Data Mining with Sequence of System Calls", in the Proceedings of International Conference on Information Assurance & security (IAS)2009,IEEE Communication Magazine, pp-154-158,ISBN:978-0-7695-3744-3,DOI:10.1109/IAS.2009.32.
[4] Norouzian.M.R, Merati.S, "Classifying Attacks in a Network Intrusion Detection System Based on Artificial Neural Networks", in the Proceedings of 13th International Conference on Advanced Communication Technology(ICACT), 2011,ISBN:978-1-4244-8830-8,pp-868-873.
[5] Prabhjeet Kaur, Amit Kumar Sharma, Sudesh Kumar Prajapat, Madam id for intrusion detection using data mining, International Journal of Research in IT & Management Volume 2, Issue 2 (February 2012) (ISSN 2231-4334) (pg 256 – 263)
[6] R.Venkatesan Dr. R. Ganesan Dr. A. Arul Lawrence Selvakumar " A Novel Approach Design in Interference Finding System for Data Mining Using MADAM ID" International Journal of Research in Engineering & Applied Sciences (IJREAS), Volume 3, Issue 3, March 2013
[7] Fayyad, U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. The KDD process of extracting useful knowledge from volumes of data. Communications of the ACM, 39(11):27–34, November 1996
[8] Anderson J., "Computer Security Threat Monitoring and Surveillance," February 26, 1980- revised April 15, 1980.
[9] Dorothy E. Denning. "An Intrusion-Detection Model" 1986 IEEE Computer Society Symposium on Research in Security and Privacy, pp 118-31
[10] R.Venkatesan, Dr. A. Arul Lawrence Selvakumar "An Analysis of Topological Control Protocols & its Issues in Wireless Sensor Network", CiiT International Journal of Networking and Communication Engineering, September 2012 DOI: NCE092012004
[11] Zinaida BENENSON, Peter M. CHOLEWINSKI , Felix C. FREILING "Vulnerabilities and Attacks in Wireless Sensor Networks" http://pi1.informatik.uni-mannheim.de/filepool/publications/zina/attacker-models-bookchapterIOS_Press.pdf
[12] Karapistoli and Economides ADLU: "a novel anomaly detection and location-attribution algorithm for UWB wireless sensor networks", EURASIP Journal on Information Security 2014, 2014:3
[13] Murad A. Rassam, Anazida Zainala, Mohd Aizaini Maarof " An Efficient Distributed Anomaly Detection Model for Wireless Sensor Networks", 2013 AASRI Conference on Parallel and Distributed Computing and Systems, ScienceDirect, AASRI Procedia 5 (2013) 9 – 14
[14] Sutharshan Rajasegarar , Christopher Leckie , Marimuthu Palaniswami " Hyperspherical cluster based distributed anomaly detection in wireless sensor networks", J. Parallel Distrib. Comput. 74 (2014) 1833–1847
[15] Mishra A, Nadkarni, K, Patcha, A," Intrusion detection in wireless ad hoc networks", Wireless Communications, IEEE (Volume:11 , Issue: 1)Feb 2004 pages 48-60
[16] Daniel Barbara, Ningning Wu and Sushil Jajodia "Detecting novel network intrusion using bayes estimators". In Proceedings of First SIAM Conference on data mining Chicago, 2001.
[17] Eric Bloedorn, Alan D. Christiansen, William Hill, Clement Skorupka, Lisa M. Talbot, and Jonathan Tivel. Data mining for network intrusion detection: How to get started.
[18] S.Y. Lee, W. L. Low and P. Y. Wong, "Learning Fingerprints for a Database Intrusion Detection System", In Proceedings of the 7th European Symposium on Research in Computer Security, Pages 264-280, 2002.
[19] R.Venkatesan Dr. R. Ganesan Dr. A. Arul Lawrence Selvakumar"A Survey on Intrusion Detection using Data Mining Techniques "International Journal of Computers and Distributed Systems Vol. No.2, Issue 1, December 2012
[20] R.Venkatesan , Dr. R.Ganesan, Dr. A.Arul Lawrence Selvakumar "A Comprehensive Study in Data Mining Frameworks for Intrusion Detection" International Journal of Advanced Computer Research (IJACR), Volume-2 Number-4 Issue-7 December-2012 pg 26 – 31
[21] R.Venkatesan , Dr. A.Arul Lawrence Selvakumar "Basics Of Intrusion Detection System And Correlation Methods" Asian Academic Research Journal of Multidisciplinary: Volume 1 Issue 5 January 2013
[22] W. Lee, S.J. Stolfo, K.W. Mok, Algorithms for Mining System Audit Data, in Proc. KDD, 1999.
[23] Agrawal R. and Srikant R., "Fast algorithms for mining association rules," in Proceeding 20th VLDB Conference, Santiago, Chile, pp. 487–499, 1994.
[24] J.R.Quinlan. C4.5: Programs for machine learning.Morgan Kaufman Publishers, 1993.
[25] Har-Peled, S., Roth, D., Zimak, D. (2003) "Constraint Classification for Multiclass Classification and Ranking." In: Becker, B., Thrun, S., Obermayer, K. (Eds) Advances in Neural Information Processing Systems 15: Proceedings of the 2002 Conference, MIT Press. ISBN 0-262-02550-7