CONTENTS

Intrusion Detection for 802.11*i* Wireless LAN

(Synopsis)

1. Introduction

Computer Network is a collection of interrelated self-governing nodes. It interacts with authorized nodes using well defined and mutually agreed set of rules and conventions known as protocols, also allows resource sharing. The devices can be linked in various ways different equipments available to connect them.

Over the past ten years, the world has become increasingly mobile. As a result, traditional ways of networking have proven derisory to meet the challenges posed by our new collective lifestyle. The movement is dramatically reduced for users connected to a network by physical cables. Wireless connectivity, however, poses no such restriction and allows a great deal more free movement on the part of the network user. As a result, wireless technologies have rapid and huge acceptance among user over the traditional "fixed" or "wired" networks.

The wireless network is the fastest emerging segment of the communications industry. It has captured the attention of the media and the imagination of the public. The exponential growth in cellular systems over the last decade and there are currently around eight billion users worldwide. Indeed, cellular phones have become a critical business tool and part of everyday life in most developed countries, and are rapidly supplanting antiquated wire line systems in many developing countries. In addition, wireless local area networks currently supplement or replace wired networks in many homes, businesses, and campuses. Like all networks, wireless networks transmit data over a network medium. The medium is a form of electromagnetic radiation. [1] Many new applications emerging from research ideas to concrete systems are wireless sensor networks, automated highways and factories, smart homes and appliances, and remote telemedicine etc. The tremendous development of wireless technology coupled with the proliferation of laptop and palmtop computers indicate a bright future for wireless networks, both as stand-alone systems and as part of the larger networking infrastructure. Still, many technical solutions remain in designing robust wireless networks that deliver the performance necessary to support emerging applications.

Wireless Network requires the necessary consideration like wireless PC Card, wireless Access points, DSL Router or wireless cabling and many more. The challenging work for wireless networks is frequency allocation, Interference, Reliability, Security, Power Consumption, and throughput. The IEEE 802.11 family supports medium-range, higher data rate applications, such as Ethernet networks, and to address mobile and portable stations. 802.11 is the original WLAN standard, designed for 1 Mbps to 2 Mbps wireless transmissions. [1]

Wireless technologies are the most important ones in current trends. This dramatically changes computer networks, considering not only mobility, availability and quality of service but also security. It is much more difficult to ensure security in this type of network than in ordinary (wired) networks. Different mechanisms are developed to provide the secure environment. Some protocols are developed to provide effective encryption and decryption policies. Recently WEP, WAP and WPA are used but they have some limitations. In the wireless network intrusion detection and prevention systems are also developed to identify the intrusive activities performed on the network. The intrusion detection is classified based on intruder type, detection behavior and approaches used. This research work focuses on the one of the intrusion detection approaches known as Anomaly detection in WiFi.

As per the current scenario, Data mining is used to identify the anomalous activities. Data Mining is the non trivial process of identifying valid, potentially and understandable patterns in the form of knowledge discovery from the databases or artificial intelligence. The main aim of this process is to discover patterns and associations among preprocessed and transformed data. Anomaly detection means any significant deviations from the expected behavior are reported as possible attacks. Data mining provides various techniques to find out the knowledge from the data. Anomalies are some type of activities that would be performed by intruders. Anomaly detection is the process of finding the objects that are not related to other normal objects. Data mining provides unsupervised learning techniques to perform behavioral analysis. Unsupervised learning is the method of grouping the data as per behavior of data. It is also known as descriptive method. Clustering is one of the unsupervised learning techniques. Clustering works on the data directly no any predefined label are required. Clustering also executes or gives the different groups as per the user wants to generate. Clustering techniques generate the groups as per the distance criteria among the data.

This research work is aimed to find out anomalous activities in wireless network observing the large quantity network data. The methodologies used are based on unsupervised learning techniques from data mining. The proposed model would be helpful in research and other business areas such as finance, medicine, tourism etc. to find out the groups according to some pattern.

2. Chapter Summary

2.1 Introduction of Wireless Network

Wireless networks function as the transport mechanism between devices and among devices and the traditional wired networks. Wireless Network and Wired Network have many differences. As per the current scenario Wireless Network usage become very incredible over wired networks because of the facilities provided by the wireless network. Wireless network offers primary benefits like Use Mobility, Rapid Installation, Flexibility, and Scalability. A variety of standards and varying levels of security features are available in Wireless technologies.

Wireless LAN technology is evolving with full of pace. Within just a few years the industry has experienced the highest data rates provided by products based on the 802.11 standards journey starts from 2 Mbps (802.11) to 11 Mbps (802.11b) and now to 54 Mbps (802.11a/g). [17] The Components of wireless LAN is Access Point, PC Card, PCI Adapter and Router. Wireless LANs can be generally categorized into two categories: ad hoc wireless LANs and wireless LANs with infrastructure. [16] The users of wireless networks want the similar services and capabilities that they have commonly faces challenges and constraints like Frequency Allocation, Security, Mobility, and Power Consumption etc… [4] This topic also gives the details about the key characteristics of WLAN and also shows the comparisons between different 802.11 varieties. [1] [5] The detail architecture of the WLAN is also discussed. 802.11 are sometimes referred to as "wireless Ethernet." The core elements present in Ethernet are present in 802.11. 802.11 also provides different types of network services like to distribute, integration, association, Re-association, Disassociation, Authentication, De-authentication, privacy and MSDU delivery. Among these services only three of the services are used for moving data; the remaining six are managing operations that allow the network to keep track of the mobile nodes and deliver frames accordingly. [1]

802.11 also support the mobility. Mobility is the major motivation for deploying an 802.11 network. Stations can move while connected to the network and transmit frames while in motion. 802.11 have been widely and rapidly adopted, security issues have continued to grab headlines. Network managers will undoubtedly be asked to comment on security issues, especially in any wireless LAN proposals. WEP and WPA have been fully broken and the IEEE is forging a successor to it based on 802.1x. [1] The key to the 802.11 specification is the MAC. It rides on every physical layer and controls the transmission of user data into the air. It provides the core framing operations and the interaction with a wired network backbone. Different physical layers may provide different transmission speeds, all of which are supposed to interoperate. [4] [1] The challenges for the MAC are RF link quality and Hidden Node Problem. Access to the wireless medium is controlled by coordination functions. Ethernet-like CSMA/CA access is provided by the distributed coordination function (DCF). If contention-free service is required, it can be provided by the point coordination function (PCF), which is built on top of the DCF. Contention-free services are provided only in infrastructure networks. [9] Carrier sensing is used to determine if the medium is available. Two types of carrier-sensing functions in 802.11 manage this process: the physical carrier-sensing and virtual carrier-sensing functions. [6]

This topic focuses on the basics of wireless local area network. It gives detailed information about WLAN components and architecture. It also describes the family of 802.11 Architecture and working of 802.11with MAC.

2.2 Study of Security in Wi-Fi at present (WEP/WPA/WAP)

Wireless technologies are the most important ones in current trends. This dramatically changes computer networks, considering not only mobility, availability and quality of service but also security. Since most of the communications in the future will use both wired and wireless networks on the communication paths, new hardware and software technologies will be needed to ensure security. Among these, the intrusion detection and prevention (IDS/IPS) technologies will have to be exposed to the most dramatic changes [2]. As wireless communication and the Internet become truly interoperable, users will want this communication channel to be secure and available when needed. For a message sent using this communication channel, the user expects assurance of authentication, confidentiality and integrity. [7] An Introduction to Computer Security generically classifies security threats in nine categories of NIST ranging from errors and omissions to threats to personal privacy. All of these represent potential threats in wireless networks as well. Ensuring confidentiality, integrity, authenticity, and availability are the prime objectives of all government security policies and practices. [5]

As with wired networks, agency officials need to be aware of liability issues for the loss of sensitive information or for any attacks launched from a compromised network. Maintaining a secure wireless network and associated devices requires significant effort, resources, and vigilance. [5] A wireless access point (AP) may be accessed from off the premises if the signal is detectable. Hence wireless networks require secure access to the AP in a different manner from wired LANs. In particular it is necessary to isolate the AP from the internal network until authentication is verified. The device attempting to connect the AP must be authenticated. The 802.11 standard provides the means to satisfy these security requirements - validation of the access device, user authentication and a secure channel. [7]

Basic three methods for 802.11 securities are used to secure access to an AP and provide a secure channel. These are SSID (Service Set Identifier), MAC address filtering and WEP security. [7] The most common WLAN security exploits are Insertion attacks, Interception and Unauthorized Monitoring, Denial of Service (DoS), Client to Client attacks, Brute Force Attacks against AP Passwords, Encryption Attacks and Misconfigurations.

As per Matthew Gast the author of “802.11 Wireless Networks: The Definitive Guide”, seven major security problems related to the 802.11 wireless network standard have been identified and some solutions have been proposed from the corporate point of view is discussed in this topic. These are Easy Access to WLAN, Rogue Access Points, Unauthorized use of service, Service and Performance Restriction, MAC Spoofing and session Hijacking, Traffic Analysis and Eavesdropping and DoS Vulnerability and other higher level attacks. [2]

The IEEE 802.11 specification identified several services to provide a secure operating environment. The security services are provided largely by the Wired Equivalent Privacy (WEP) protocol to protect link-level data during wireless transmission between clients and access points. WEP does not provide end-to-end security, but only for the wireless portion of the connection. The security services like authentication techniques, privacy policy and integrity offered by 802.11 families. The IEEE 802.11 specification does not, unfortunately, identify any means for key management (life cycle handling of cryptographic keys and related material). Therefore, generating, distributing, storing, loading, escrowing, archiving, auditing, and destroying the material is left to those deploying WLANs. [5] Key management for 802.11 is left largely as an exercise for the users of the 802.11 network. As a result, much vulnerability could be introduced into the WLAN environment. These vulnerabilities include WEP keys that are non-unique, never changing, factory-defaults, or weak keys. For that various security protocols are developed to ensure the security like WAP Protocol Family, WEP and WPA with different authentication, encryption and key size configuration policies. [15] [7]

This topic describes the security conventions implemented by wireless local area network. It also explores the mechanisms of different types of protocols with the detailed mechanisms used to provide better security with them limitations.

2.3 *Study of Intrusion detection approaches in Wi-Fi*

Computer intrusion as a set of actions aimed at compromising the security goals (confidentiality, integrity, availability of a computing/ networking resource). Intrusion detection can be defined as the process of identifying and responding to intrusion activities. An Intrusion prevention system (IPS) is software that has all the capabilities of an intrusion detection system and can also attempt to stop possible incidents. [23] An intrusion prevention system (IPS) combines IDS with a firewall, a virus detection algorithm, a vulnerability assessment algorithm, etc. The ambition of such a system is to manage both preventive and responsive actions against attacks on a computer network. [2] An IDPS might be able to block reconnaissance and notify security administrators, who can take action if needed to alter other security controls to prevent related incidents. Because reconnaissance activity is so frequent on the Internet, reconnaissance detection is often performed primarily on protecting internal networks. [23] The key function of IDPS technologies are Recording information related to observed events, Notifying security administrators of important observed events and Producing Reports. Some IDPSs are also able to change their security profile when a new threat is detected. IPS technologies are differentiated from IDS technologies by one characteristic: IPS technologies can respond to a detected threat by attempting to prevent it from succeeding. They use several response techniques like IPS stops attacks itself, the IPS change security environment, or it may change the attacks contents. [2] The following figure shows the classification of Intrusion Detection System.

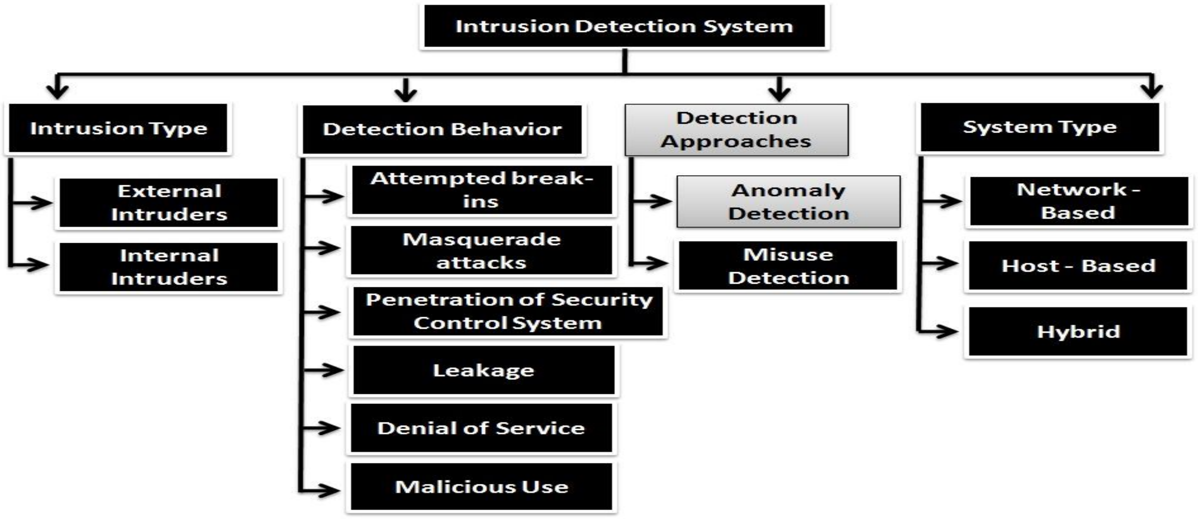


Figure 1: IDS Classification

The components of IDPS are Sensor or Agent, Management Server, Database Server and Console. The wireless network traffic is monitored by the wireless IDPS monitor and analyzes its wireless networking protocols to identify suspicious activity involving the protocols themselves. This section provides a detailed discussion of wireless IDPS technologies. [23] Wireless intrusion detection systems can be divided into misuse detection and anomaly detection. [2] The main problem with misuse detection system is the distribution of the elements of the IDS for that three possible approaches are used like Wireless IDS sensors and processors are integrated into the access points, Overlay IDS with centralized processing and • Overlay IDS with decentralized processing. [2]

A multilevel wireless IDS/IPS is uses agents on the hosts, sensors, an IDS server and a reporting tool in order to combine host based and network based detection in a wireless network environment. [2] This section focuses on the component, architecture and objective of wireless intrusion detection network. The commonly intrusive activities detected by the wireless intrusion detection sensors. It’s also focused on the key difference between wired IDS and wireless IDS. The wireless IDS techniques are also used for the MAC spoofing, Rogue access points and for many others. [30] MAC spoofing allows an adversary to assume the MAC address of another WLAN node and launch attacks on the WLAN using the identity of the legitimate node. A number of different techniques have been suggested to detect MAC spoofing activity in a WLAN some of the techniques are Sequence Number Monitoring, Fingerprinting, Location Determination and Signal Strength Fourier Analysis [30]

This topic focuses on Intrusion detection and prevention system. It reviews the usage of intrusion detection and prevention mechanisms and introduces recently used mechanisms to detect or protect the wireless local area network intrusive activities with its confines.

2.4 *Anomaly detection approach for intrusion detection in Wi-Fi*

Anomaly-based detection is the process of comparing definitions of what activity is considered normal against observed events to identify significant deviations. The activities of intrusion detections are then noticed as anticipated or normal behaviors of the monitored system, with intrusions defined as a variation of monitored behavior from the baseline that system is known as anomaly based intrusion detection system. This system does not require explicit signatures of security events. They use probable or non-malicious behavior and elevate any variations from this behavior as security events. [30] The anomaly based intrusion detection system can be classified in two ways such as statistical models and specification based models.

The statistical modeling comes up to identify the events which are significance to anomaly based system. In this approach variables and its characteristics are deliberated in excess of certain time scales by the intrusion detection system and statistically outlined to extend a baseline of normal or estimated behavior of the observed computer system or network. So this modeling requires the training phase for intrusion detection system to understand estimated or normal behavior of the observed system. The Specification based models does not require any training phase instead the expected correct behavior is overtly provided in a declarative manner. [30]

The potentiality of the anomaly-based IDSs are that they are competent of sensing both existing and novel attacks without having to be reconfigured or updating. Anomaly based intrusion detection systems ensign observed activities that depart significantly from the conventional normal usage profiles can be defined as anomalies or probable intrusions. [24] The main advantage of anomaly detection is that it will able to find new intrusions without requiring any prior understanding of the intrusion and disadvantage is that it may not be able to illustrate what attack is.

Anomaly detection technique presumes that all intrusive activities are essentially anomalous. The following figure shows the anomaly based intrusion detection system.

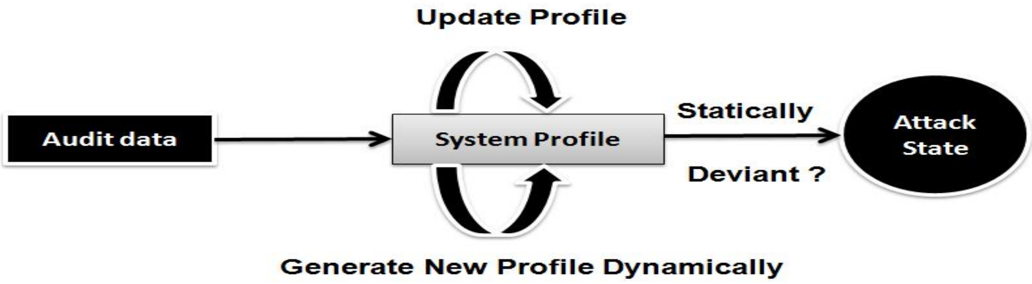


Figure 2: Anomaly based IDS

Anomaly based intrusion detection system is used to differentiate intrusive activities from the normal activities for that different anomaly detection techniques are used which are classified into Statistical analysis techniques, Data mining techniques, and Rate limiting techniques. The following figure shows the classification of anomaly detection techniques.

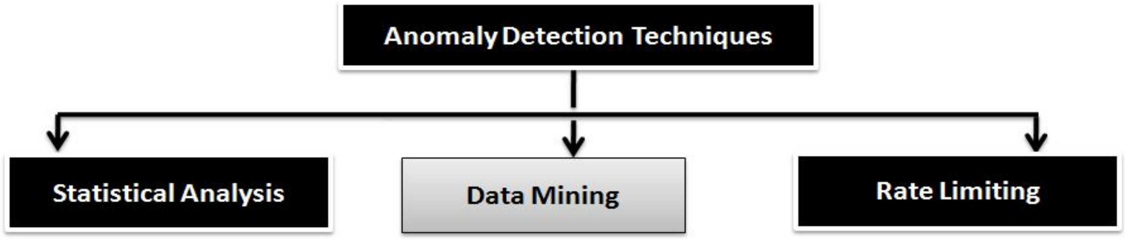


Figure 3: Anomaly detection techniques classifications

In this research work data mining techniques are used to determine patterns of system features that describes events and user behavior and compute a classifier that can recognize anomalies and intrusions. For this wireless network log is used. Data mining provides various machine learning techniques to find out unknown patterns from the data. To find out anomaly behavior activities from the wireless network log first perform preprocessing of the data. Pre processing of data means to find out the missing values from the log, which are the most related attributes, perform dimension reduction etc… After this Statistical analysis is used to prove the applied log is statistically appropriate to perform further analysis.

This topic focuses on the anomaly detection approaches used by a wireless local area network to identify intrusive activities. It discusses the recently used approaches and how they work. This chapter focuses on the actual research approach to the study.

2.5 *Unsupervised Technique for intrusion detection in Wi-Fi and its outcomes*

Data Mining is one of the approaches which are recently used in networking area. Data mining provides machine learning techniques to perform analytical techniques. Data mining is up-and-coming with the key features of much security inventiveness. Both the private and public sectors are currently increasingly using the data mining. Many application domains use data mining. Data mining applications initially were used as a means to detect fraud and waste, but have grown to also be used for purposes such as measuring and improving program performance. Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. The Data Mining tools can include statistical models, mathematical algorithms, and machine learning methods.

In this research work the wireless network log is taken from the various access points deployed at three different sites. The log traced for every day from March 2010 to December 2012. There are three files for each access point and for each day. These files contain the information regarding the SNMP Packet transmission in wireless network, the information regarding SNMP interface and the SNMP users' information. File names start with the name of the access point and suffix name of the file gives the information about the information it stored. For ex. The access point name is amp01 then the file name for SNMP packet information is amp01.snmp, for interface information regarding file name is amp01-interfaces.snmp and for users information its name is amp01-users.snmp. So for the further processing on the data, one file is generated with merging all the attributes which are very much helpful to accomplish this research aim to identifying the anomalous activities. Approximate the file is generated with more than 5, 00,000 instances. It contains the attributes like Site name, day, Name, System up time, the moment at which packet is transmitted, number of input and output packets from IP, SNMP, TCP and UDP.

After preparing the log for further processing of the analysis techniques first statistical techniques are applied to the same file. The statistical analysis is used to define normality and to define the prepared log is valid for the further analytical techniques. In Most of circumstances normalized data can be extremely difficult or impossible to obtain. Intrusion detection system requires the efficient unsupervised algorithms for intrusion behavior analysis and for better efficiency of unsupervised algorithm normalized data required. The main purpose of factor analysis is to identify important input features to build IDS that is computationally efficient and effective development of classification techniques for unsupervised anomaly detection.

Data mining provides most accurate model but the model accuracy totally depends on the data. It is very much helpful before developing model, first preprocess the data. Data mining also provides a way to develop any analytical model first perform data preprocessing which contain different processing such as data summarization, data cleaning and data reduction. The following figure shows the proposed model for this research work.

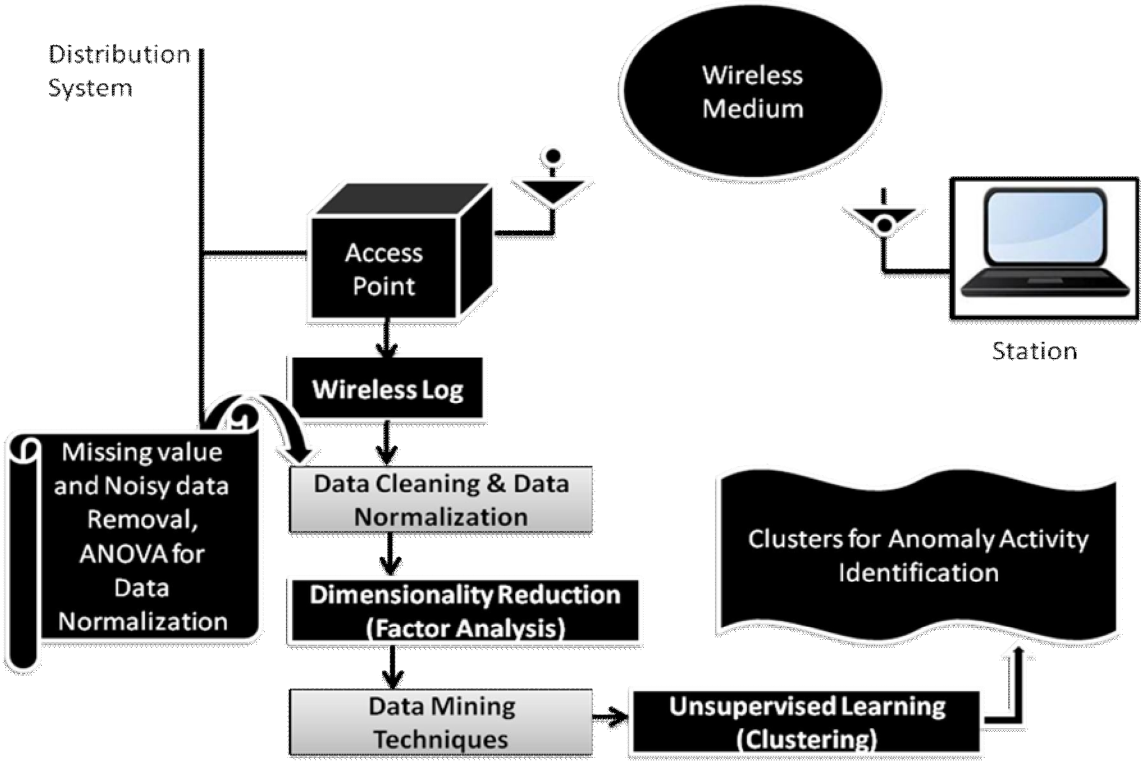
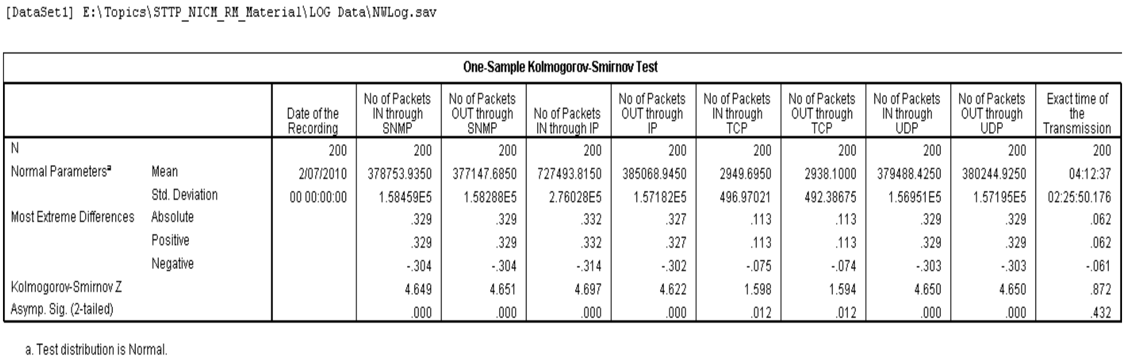

Distribution System
Wireless Medium
Access Point
Station
Wireless Log
Missing value and Noisy data Removal, ANOVA for Data Normalization
Data Cleaning & Data Normalization
Clusters for Anomaly Activity Identification
Dimensionality Reduction (Factor Analysis)
Data Mining Techniques
Unsupervised Learning (Clustering)

Figure 4: Proposed Model

In this research work, the data summarization and data cleaning is done using the different statistical techniques. For data summarization, descriptive summarization measures like mean and standard deviation is applied. It gives the positive result. In this research work for data cleaning process data missing or noisy data avoidance also tested and ignored. This log also applied for normality test. On this log ANOVA the statistical technique is applied. ANOVA is much useful for both parameters (scoring data) and non parametric data (Ranking or ordering data). ANOVA is also useful to compare attributes of these different groups which means of measures something for a specific period of time in a single group. So in this research work, to identify some activities perform on the same network is measured. The following figure shows the results of ANOVA applied to wireless network log.



```
[DataSet1] E:\Topics\STTP_NICM_RM_Material\LOG Data\NWLog.sav
```

One-Sample Kolmogorov-Smirnov Test

		Date of the Recording	No of Packets IN through SNMP	No of Packets OUT through SNMP	No of Packets IN through IP	No of Packets OUT through IP	No of Packets IN through TCP	No of Packets OUT through TCP	No of Packets IN through UDP	No of Packets OUT through UDP	Exact time of the Transmission
N		200	200	200	200	200	200	200	200	200	200
Normal Parameters[a]	Mean	2/07/2010	378753.9350	377147.6850	727493.8150	385068.9450	2949.6950	2938.1000	379488.4250	380244.9250	04:12:37
	Std. Deviation	00 00:00:00	1.58459E5	1.58288E5	2.76028E5	1.57182E5	496.97021	492.38675	1.56951E5	1.57195E5	02:25:50.176
Most Extreme Differences	Absolute		.329	.329	.332	.327	.113	.113	.329	.329	.062
	Positive		.329	.329	.332	.327	.113	.113	.329	.329	.062
	Negative		-.304	-.304	-.314	-.302	-.075	-.074	-.303	-.303	-.061
Kolmogorov-Smirnov Z			4.649	4.651	4.697	4.622	1.598	1.594	4.650	4.650	.872
Asymp. Sig. (2-tailed)			.000	.000	.000	.000	.012	.012	.000	.000	.432

a. Test distribution is Normal.

Figure 5: ANOVA Test result

The next step is to perform data reduction. Different data reduction and dimensionality reduction methods are available and as per the objective they applied to data to identify which attributes affect more each other and which attributes are not much helpful for this research objective. In this research work feature construction is performed using factor analysis. Factor analysis is a collection of methods used to examine how basic constructs manipulate the responses on several measured variables. Factor analysis is done by two types: exploratory and confirmatory. Exploratory factor analysis (EFA) attempts to discover the nature of the constructs influencing a set of responses. Confirmatory factor analysis (CFA) tests whether a specified set of constructs is influencing responses in a predicted way. Factor analysis is performed by examining the pattern of correlations (or co-variances) among the observed measures.

Factor analysis mainly performed to identify essential attributes or factors that enlighten the correlations among a set of initial features. Factor analysis is one of the data reduction statistical methods which discovers a number of factors that enlighten the variance empirical in a large number of variables. It also used to generate hypothesis concern with fundamental causally related mechanisms to identify Co linearity so it’s prior to performing a linear regression analysis.

Factor analysis is a statistical technique used to identify a relatively small number of factors that can represent relationships among sets of many interrelated variables. It reduces the attribute space from a larger number of variables to a smaller number of factors. Factor analysis generates a table in which the rows are the observed as raw indicator variables and the columns are factors that explain as much of the variance in these variables as possible. The cells in this table are factor loadings, and the meaning of the factors must be induced from observing which variables are most heavily loaded on certain factors. The factor loadings are the correlation among the variables and factors. [39]

In this research work factor analysis is performed on this log for factorize the attributes that means to find out which attributes are closely related so the effectiveness of the attributes for each other is controlled. So after applying factor analysis it gives two factors this denotes that which attributes are dependent. This gives the results as the number of input output packets from which protocols are closed related to the time.

The factor analysis is done on this using the KMO statistics. The results which are shown in figures getting by performing feature construction methods gives information that the network traffic using TCP differs than other protocols IP, UDP and SMTP. So the packet transmission using with these three may be Co related. So it's useful for further analysis using unsupervised learning for anomaly detection. The main result of the factor analysis is measured by scree plot and the result of the scree plot for this log is represented by following figure.

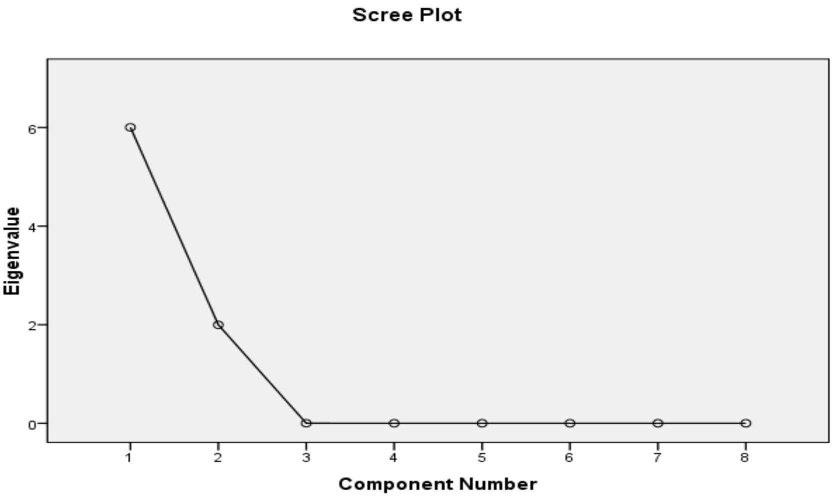


Figure 6: Factor Analysis Result (Scree Plot)

The scree plot is a useful way of establishing how many factors should be retained in an analysis. Eigenvalues represent the amount of variance accounted for by each factor. As per the primary analysis, the scree plots and the eigenvalues over 1 lead you retain the same number of factors then continue with the analysis with proper direction. At this stage, the applied network log is statistically proved to apply for further analytical techniques of data mining to develop analytical model.

Data mining exploits a discovery approach, in which algorithms can be used to scrutinize several multidimensional data relationships concurrently, discovering those that are unique or frequently represented. Many Organizations provide data mining tools to survey different user work oriented information and gives analytical results to interpret so these tools reduce fraud and waste of time to assist in developing algorithms for research. The ideal model is given by the algorithms as per the error rate and accuracy given by them. Some observers suggest that data mining should be used as a means to identify terrorist or intrusive activities, such as money transfers and electronic communications, and to identify and track individual terrorists or intruders themselves. [35] The data mining provides supervised learning techniques as well as unsupervised learning. Supervised learning means to provide labeled classes to further knowledge gathering process. The unsupervised learning techniques are not depending on any labelized classes and processing the data for knowledge gathering as per the behavior of the data. The classification is the supervised learning technique and clustering is the unsupervised technique of the data mining.

Anomaly detection means any significant deviations from the expected behavior are reported as possible attacks. Data mining provides various techniques to find out the knowledge from the data. Anomalies are some type of activities that would be performed by intruders. Anomaly detection is the process of finding the objects that are not related to other normal objects. Data mining provides the techniques to find out such a group or classes as per the requirement and the usage of the work. Classification is used to classify the data gathered from the different collected data. Data mining also provides another technique that is clustering. Clustering is also used for grouping the data as per the behavior of the data. So data mining techniques are useful to find out the groups or classes. These classes or groups are useful to differentiate the other dissimilar groups as per the predefined labels or the behavior of data. Unsupervised learning is the method of grouping the data as per behavior of data. Clustering works on the data directly no any predefined label are required. Clustering also executes or gives the different groups as per the user wants to generate. Clustering techniques generate the groups as per the distance criteria among the data. There are different distance measure methods are available to count the distance amount the instances.

Different clustering provider tools use different distance measure of grouping the data. The accuracy of the results depends on the algorithms that uses which types of statistics used to measure the distance between instances and the representation of the cluster characteristics.

K-Mean Clustering is one of the clustering techniques. K-Mean algorithm is developed by J. MacQueen in 1967 and then by a J. A. Hartigan and M. A. Wong around in 1975. This algorithm is used for grouping the instances based on features into k number of groups. The grouping is done by minimizing distances between data and the corresponding cluster centroids. The general process of the K-Mean is shown in below figure. [38]

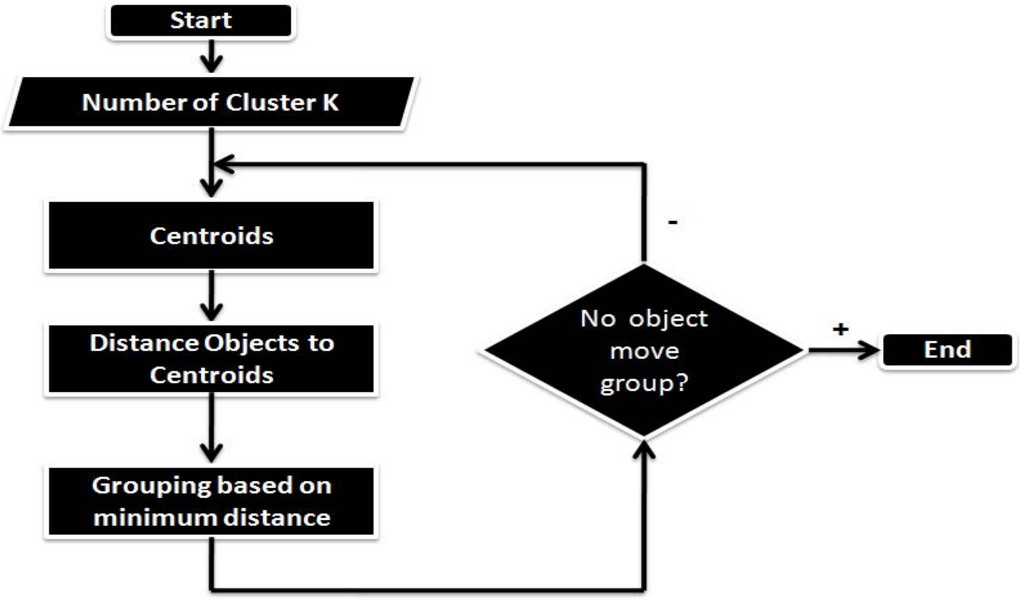

Start
Number of Cluster K
Centroids
Distance Objects to Centroids
Grouping based on minimum distance
-
No object move group?
+
End

Figure 7: Process of K-Mean Clustering

Clustering is the data mining unsupervised learning technique which is used with wireless network to detect the intrusive behavior of network activities. To apply clustering the applied wireless network log is statistically ready to perform further machine learning unsupervised technique. In this research work, Microsoft SQL Server 2008 which provides Business Intelligence Development Studio is used to perform analytical services. There are many different data mining tools available to perform data mining algorithms. For this research work, WEKA, Tanagra 3.6, and SPSS also have tried but compare to these tool BIDS of MS SQL Server provides ideal model. Among this tool is best suited for especially statistical analysis and one is specially research tool. Weka is Open source but it's can't give a better model for lengthy and complex data. The BIDS of MS SQL Server provides effective data mining algorithms. The one of is Microsoft Clustering algorithm. This detaches applied data into intelligence groupings. The Microsoft Clustering allows more flexibility because it's working with a more content type like continuous, Discrete and more types as input and configure appropriate method to create clusters. The Microsoft Clustering is not generally used for prediction purpose but it's majorly used to discover out natural groupings from the applied data. This algorithm provides more flexibility for configuring the method of grouping. The mining structure and model is one of the results given by BIDS to display the designing model and the attributes which selected for inputs or for key parameters for the algorithm.

The Microsoft Clustering provides four types of cluster views. These are Cluster Diagram, Cluster Profile, Cluster Characteristics and Cluster Discrimination.

- **Cluster Diagram:** The Cluster Diagram demonstrates all the clusters that are used in Mining model.

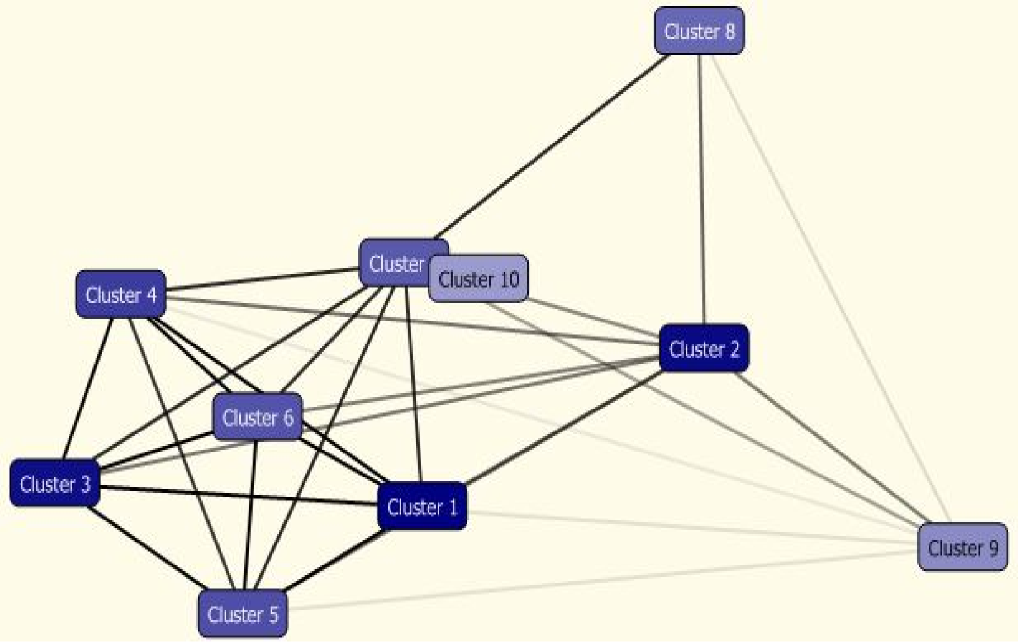


Figure 8: Cluster Diagram for Wireless Network Log

The strength of the similarity of the clusters represented by the shading of lines connected among the clusters. The light shading the clusters denotes that these clusters are not very similar.

The clustering of the BIDS is more flexible because it uses EM, K-Mean and scalable or non scalable methods of grouping. The Cluster diagram shows the characteristics of each and every cluster. The strength of the similarity of the clusters represented by the shading of lines connected among the clusters. The light shading the clusters denotes that these clusters are not very similar. So as per this model of Cluster diagram cluster number eight, nine and ten represented by light shading so they have instances that is not much similar to the others. So the instances belong to those clusters show the anomalous activities. The cluster number five six and seven represented by average shading so it’s interpreted as the instances of these clusters are suspicious. The remaining clusters are purely highlighted so they have normal behavioral instances. The model gives 15% density which is accurate by calculating the ratio of the number of instances in each cluster with the overall instances in the log. So it gives the ideal model to identify each and every instance of the log statistically.

- **Cluster Profile:** The Cluster Profile provides the information about the clusters generated by an algorithm for applying log.

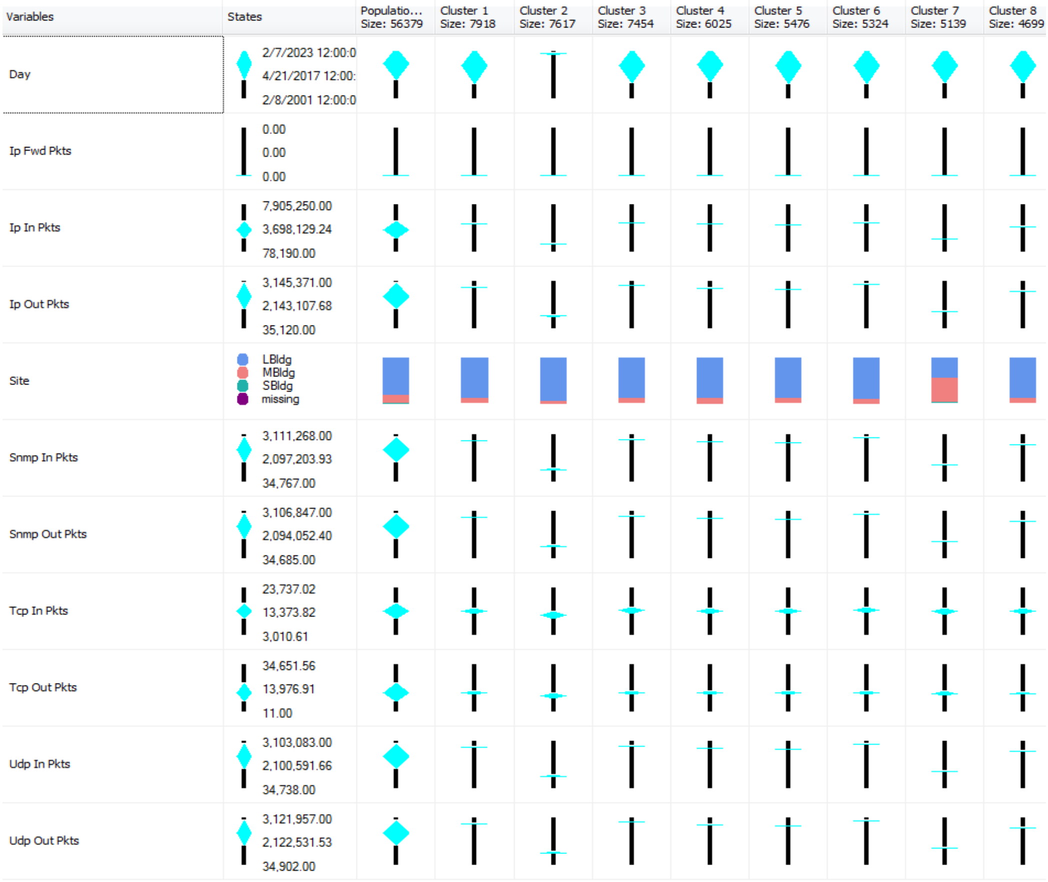
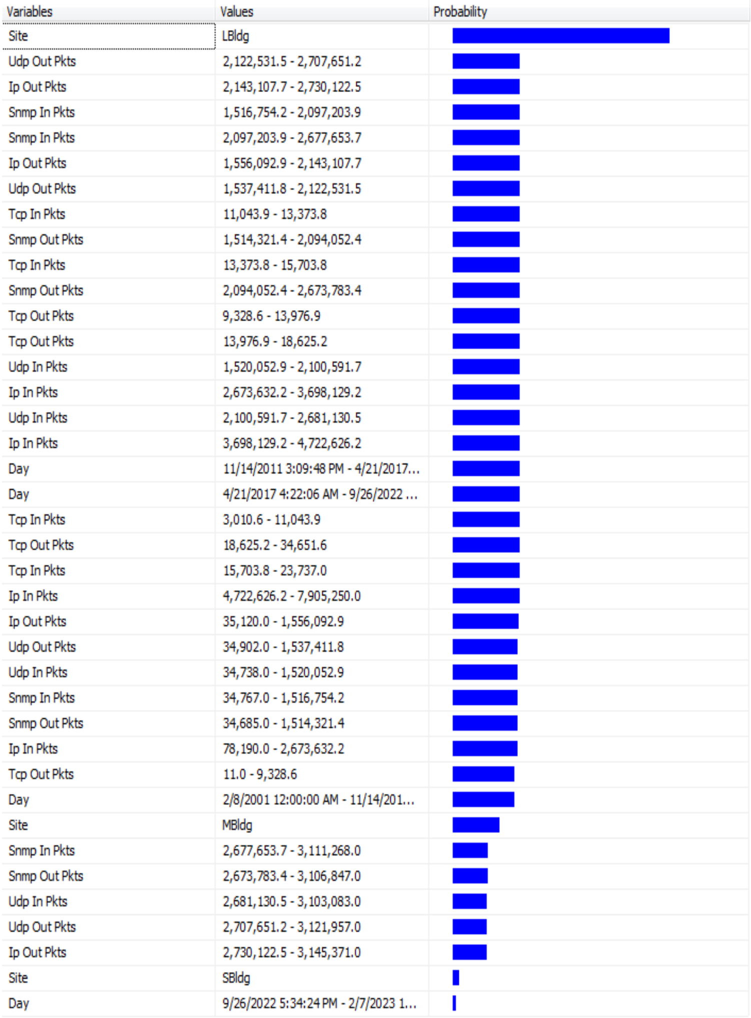


Figure 9: Cluster Profile

This profile displays each attribute, together with the allotment of the attribute in each cluster. An Information Tip for each cell demonstrates the distribution statistics, and the column heading displays the cluster population. The continuous instances are represented by diamond chart and discrete instances represented by color bars that shows the standard deviation and mean for each cluster. The Histogram bars option can be in command of the number of bars visible in histogram. The bars of highest importance are retained, and the remaining bars are grouped together into a gray bucket.

- **Cluster Characteristics:** Cluster Characteristics provides the facility to show the statistics of each cluster instance as per the selection list chosen by the user. The overall characteristics and its possibilities shown in following figure.



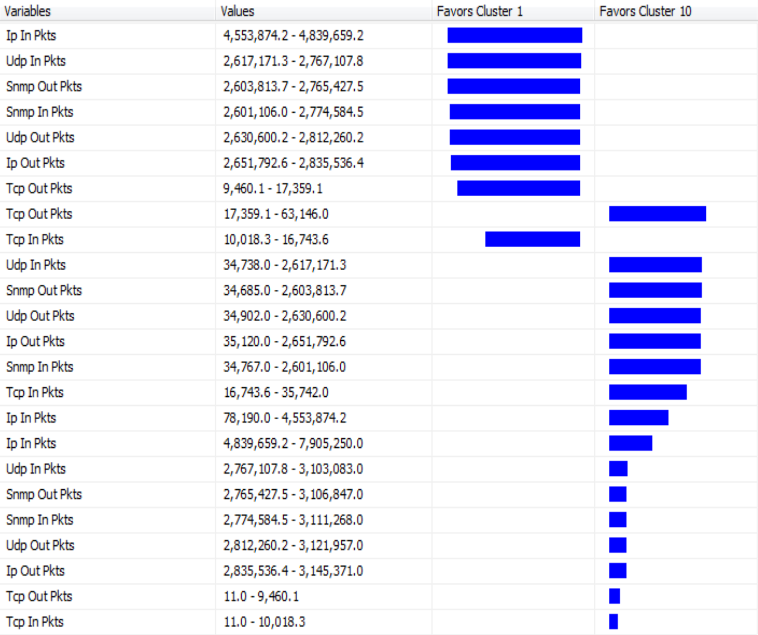
Variables	Values	Probability
Site	LBldg	
Udp Out Pkts	2,122,531.5 - 2,707,651.2	
Ip Out Pkts	2,143,107.7 - 2,730,122.5	
Snmp In Pkts	1,516,754.2 - 2,097,203.9	
Snmp In Pkts	2,097,203.9 - 2,677,653.7	
Ip Out Pkts	1,556,092.9 - 2,143,107.7	
Udp Out Pkts	1,537,411.8 - 2,122,531.5	
Tcp In Pkts	11,043.9 - 13,373.8	
Snmp Out Pkts	1,514,321.4 - 2,094,052.4	
Tcp In Pkts	13,373.8 - 15,703.8	
Snmp Out Pkts	2,094,052.4 - 2,673,783.4	
Tcp Out Pkts	9,328.6 - 13,976.9	
Tcp Out Pkts	13,976.9 - 18,625.2	
Udp In Pkts	1,520,052.9 - 2,100,591.7	
Ip In Pkts	2,673,632.2 - 3,698,129.2	
Udp In Pkts	2,100,591.7 - 2,681,130.5	
Ip In Pkts	3,698,129.2 - 4,722,626.2	
Day	11/14/2011 3:09:48 PM - 4/21/2017...	
Day	4/21/2017 4:22:06 AM - 9/26/2022 ...	
Tcp In Pkts	3,010.6 - 11,043.9	
Tcp Out Pkts	18,625.2 - 34,651.6	
Tcp In Pkts	15,703.8 - 23,737.0	
Ip In Pkts	4,722,626.2 - 7,905,250.0	
Ip Out Pkts	35,120.0 - 1,556,092.9	
Udp Out Pkts	34,902.0 - 1,537,411.8	
Udp In Pkts	34,738.0 - 1,520,052.9	
Snmp In Pkts	34,767.0 - 1,516,754.2	
Snmp Out Pkts	34,685.0 - 1,514,321.4	
Ip In Pkts	78,190.0 - 2,673,632.2	
Tcp Out Pkts	11.0 - 9,328.6	
Day	2/8/2001 12:00:00 AM - 11/14/201...	
Site	MBldg	
Snmp In Pkts	2,677,653.7 - 3,111,268.0	
Snmp Out Pkts	2,673,783.4 - 3,106,847.0	
Udp In Pkts	2,681,130.5 - 3,103,083.0	
Udp Out Pkts	2,707,651.2 - 3,121,957.0	
Ip Out Pkts	2,730,122.5 - 3,145,371.0	
Site	SBldg	
Day	9/26/2022 5:34:24 PM - 2/7/2023 1...	

Figure 10: Cluster Characteristics

This statistic shows the strength of the selected cluster. The instance of the selected cluster is represented by the variable column and the value state of the instances statistics display in value column. The last column shows the probability that the attributes that will appear in clusters.

- **Cluster Discrimination:** The Cluster Discrimination is used to compare the instances between two clusters. Select the two clusters which comparison you want to analyze. The instances which are contained in clusters are represented as variable and its value state that defines the associated instances range to define the importance and the key differences. The last two columns display the bars to represent the favor of selected cluster instances.

The following figure shows the cluster discrimination for applying log.



Variables	Values	Favors Cluster 1	Favors Cluster 10
Ip In Pkts	4,553,874.2 - 4,839,659.2	■	
Udp In Pkts	2,617,171.3 - 2,767,107.8	■	
Snmp Out Pkts	2,603,813.7 - 2,765,427.5	■	
Snmp In Pkts	2,601,106.0 - 2,774,584.5	■	
Udp Out Pkts	2,630,600.2 - 2,812,260.2	■	
Ip Out Pkts	2,651,792.6 - 2,835,536.4	■	
Tcp Out Pkts	9,460.1 - 17,359.1	■	
Tcp Out Pkts	17,359.1 - 63,146.0		■
Tcp In Pkts	10,018.3 - 16,743.6	■	
Udp In Pkts	34,738.0 - 2,617,171.3		■
Snmp Out Pkts	34,685.0 - 2,603,813.7		■
Udp Out Pkts	34,902.0 - 2,630,600.2		■
Ip Out Pkts	35,120.0 - 2,651,792.6		■
Snmp In Pkts	34,767.0 - 2,601,106.0		■
Tcp In Pkts	16,743.6 - 35,742.0		■
Ip In Pkts	78,190.0 - 4,553,874.2		■
Ip In Pkts	4,839,659.2 - 7,905,250.0		■
Udp In Pkts	2,767,107.8 - 3,103,083.0		■
Snmp Out Pkts	2,765,427.5 - 3,106,847.0		■
Snmp In Pkts	2,774,584.5 - 3,111,268.0		■
Udp Out Pkts	2,812,260.2 - 3,121,957.0		■
Ip Out Pkts	2,835,536.4 - 3,145,371.0		■
Tcp Out Pkts	11.0 - 9,460.1		■
Tcp In Pkts	11.0 - 10,018.3		■

Figure 11: Cluster Discrimination

The graphical representation of the mining model is Lift Chart which is used to analyze the scope or to make further improvement in the model. The improvement measure compared against the randomness guess and the measures the change in terms of lift score. It determines the idleness of the model which measured by the percentage of the cases in the data set would benefit from applying the model's predictions. The following figure shows the lift chart for applying log.

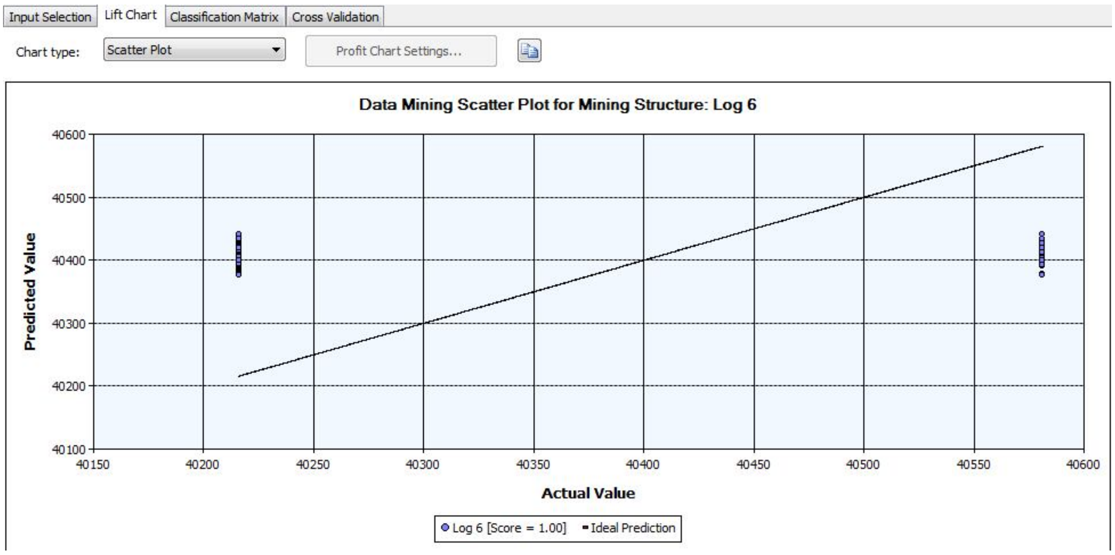


Figure 12: Lift Chart

As per the applied log lift chart, It shows the model accuracy with the statistically generated score as 1 and it's defined as this model is the ideal model for the prediction. Lift Chart is also used to compare the accuracy of predictions of multiple models that have the same instances. This shows the overall accuracy of the generated model.

2.6 ***Future Scope for the extension of research work***

The proposed research work is useful to categorize the activities performed on the wireless network. It also conveys the extension of the research work and its future scope. It shows the how to extend this current proposed work to provide model for more accurate detection of anomalized activities perform on network which is harmful. Intrusion Detection Systems (IDS) have nowadays become a necessary component of almost every security infrastructure. In this research work, data mining techniques are applied to identify the anomalized activities.

This model will develop using unsupervised learning techniques of data mining. For more specification based results, Sequential clustering also useful because it monitors the states between values. On this log for analyzing more reactive behavior of the wireless network or user activities other different analytical techniques of data mining are useful. If the time series analysis technique of data mining applied to this log then it analyzes the frequency of occurrence time of anomalous activities.

The model will extended using the different approaches to make it more effective. Swarm Intelligence (SI), a relatively new bio-inspired family of methods, seeks inspiration in the behavior of swarms of insects. The unique characteristics of SI make it ideal for this purpose. More specifically, Swarm Intelligent techniques aimed at solving complex problems by the employment of multiple but simple agents without the need of any form of supervision to exist. Every agent collaborates with others toward finding the optimal solution. [37] Clustering-based intrusion detection algorithms, clustering using a simple distance-based metric and detection based on the centers of clusters. Swarm intelligence can deal with the unknown intrusions efficiently in the real network connections. So the extended model works effectively for real time network by using this approach. [36]

By dealing with real time activities performed on a wireless network, the intrusion prevention system will developed to restrict this type of activities.

3. References

1. Mattbew S. Gast. “802.11 Wireless Networks The Definitive Guide” Published By: O’Reilly, ISBN: 0-596-00183-5 [Viewed On 10/8/2011]
2. Singh Amardeep & Singh Gurjeet (December 2010). “Vulnerabilities and Intrusion Detection in Wireless Networks” Published At: IJCST Vol. 1, Issue 2 ; ISSN : 2229-4333 (Print) | ISSN : 0976-8491 (Online)
3. Hung-Yun Hsieh & Raghupathy Sivakumar. “IEEE 802.11 over Multi-hop Wireless Networks: Problems and New Perspectives” currently in proceedings
4. Crow Brian P. , Widjaja Indra, Kim Jeong Geun & Sakai Precott T. (September 1997). “IEEE 802.11 Wireless Local Area Networks” Available At: IEEE Communications Magazine - 0163-6804/97
5. Karygiannis Tom & Owens Les (November 2002). “Wireless Network Security 802.11, Bluetooth and Handheld Devices” Special Publication by : NIST Special Publication 800-48, Available At: Computer Security
6. W@P Forum (January 2002). “Wireless Application Protocol WAP 2.0” Technical White Paper Available At: www.wapforum.org
7. Boncella Robert J. (2002). “WIRELESS SECURITY: AN OVERVIEW” Published By: Communications of the Association for Information Systems (Volume 9, 2002) 269-282
8. Banerjee Rahul. “Introduction to Computer Network” Information Technology, Birla Institute of Technology and Science
9. U.S. Robotics. “Wireless LAN Networking” White paper available at: Ready. Set. Connect.
10. Lafe Olu , PhD Student (March 7, 2005). “The Future of Wireless Technologies” Current trends in Wireless Technology, John Carroll University in Free Enterprise Event
11. Dubendorf Vern A. (2003). “History of Wireless Network ” ; Available at: 2003 John Wiley & Sons, Ltd; ISBN: 0-470-84949-5
12. Goldsmith Andrea (2005). “WIRELESS COMMUNICATIONS” Published By by Cambridge University
13. Negus Kevin J. (IEEE Member) Al Petrick (IEEE Member) (April 4, 2008). “History of Wireless Local Area Networks (WLANs) in the Unlicensed Bands”

Published At: George Mason University Law School Conference, Information Economy Project, Arlington

14. Holt Keith and Intel Corporation (2005). “Wireless LAN: Past, Present, and Future” Proceedings of the Design, Automation and Test in Europe Conference and Exhibition (1530-1591); Published At: IEEE Computer Society
15. NETGEAR Inc. Team. - Santa Clara, USA (2005). “Wireless Networking Basics” Published By: NETGEAR, Inc - Vol. Product Family 1.0
16. Chandramouli Vijay “A Detailed Study on Wireless LAN Technologies” in proceedings.
17. BROADCOM (April 2006). “802.11n: Next-Generation Wireless LAN Technology”
18. IEEE Standards Association. IEEE 802.11 Standard(1999). [Online] Available http://standards.ieee.org/getieee802/ download/802.11a-1999.pdf
19. Gast Matthew (24 May, 2002). “The top seven security problems with WLAN” An Article available at: http://www.oreillynet.com/wireless/2002/05/24/wlan.html
20. ArsTechnica. “Wireless Security Blackpaper” http://arstechnica.com/articles/paedia/security.ars
21. Rivest, R L. (1992). “The RC4 Encryption Algorithm” By RSA Data Security Inc.,
22. Goldberg Borisov N, Wagner I. D. (2001) Intercepting Mobile Communications: The Insecurity of 802.11. Proceedings of the 7th Annual Inter Conference on Mobile Computing and Networking, July 2001, 180–189.
23. Scarfone Karen & Mell Peter Mell “Guide to Intrusion Detection and Prevention Systems (IDPS)” Published By: NIST Special Publication 800-94
24. Sobh Tarek S., “Wired and wireless intrusion detection system: Classifications, good characteristics and state-of-the-art” Available At: www.elsevier.com/locate/csi; Published By: Science Direct Computer Standards and interfaces 28(2006) 670-694
25. Karygiannis Tom & Owens Les (November 2002) “Wireless Network Security 802.11, Bluetooth and Handheld Devices” , Published By: NIST Special Publication 800-48; Available At: Computer Security
26. Lim Yu-Xi & Schmoyer Tim (June 2003) “Wireless Intrusion Detection and Response” Published By: IEEE; ISBN 0-7803-7809-1/03/$17.00

27. Boncella Robert J. (2002) "WIRELESS SECURITY: AN OVERVIEW" Published By: Communications of the Association for Information Systems (Volume 9, 2002) 269-282
28. White paper on "Wireless LAN Networking" By U.S. Robotics
29. National Security Agency, Article published on "Systems and Network Analysis Center Information Assurance Directorate", Available at: http://nsa.gov./snac/
30. Gill Rupinder (July 2, 2009), Thesis on "Intrusion Detection Techniques Wireless Local Area Network"
31. Kaur Lakhwinder (April 2011), Thesis on "STUDY OF THE ENHANCEMENTS IN INTRUSION DETECTION TECHNIQUES FOR WIRELESS LOCAL AREA NETWORKS (WLAN)" , Submitted At: UNIVERSITY COLLEGE OF ENGINEERING
32. Hutchison Ken, (18 October 2004) "Wireless Intrusion Detection Systems" Published By SANS Institute Reading Room site
33. Pleskonjic Dragan, "Wireless Intrusion Detection Systems" (Refered On Date: 23 October, 2012) Available At: Dragan_Pleskonjic@conwex.net
34. Moorthy M. & Sathiyabama S.(2012) , "A Hybrid Data Mining based Intrusion Detection System for Wireless Local Area Networks" Published At: International Journal of Computer Applications (0975 – 8887) Volume 49– No.10, July 2012
35. Seifert Jeffrey W. ,A CRS Report for Congress"Data Mining: An Overview"
36. Zhong-Fu Wu ; Kai-Gui Wu ; Zhong-Yang Xiong ; Ying Zhou, (2005)"An unsupervised anomaly intrusion detection algorithm based on swarm intelligence" Published By: International conference of IEEE Xplore digital library in proceeding volume of 2005.
37. Kolias C., Kambourakis G., Maragoudakis M., (2011)"Swarm intelligence in intrusion detection: A survey"; Published By: Elsevier Computer & Security in 2011
38. Zlatan Aki Mur, (2006) K-Mean Clustering, Retrieved from: http://people.revoledu.com/kardi/tutorial/kMean/NumericalExample.htm
39. Patel A. M. & Patel A. R., "Exploratory Data Model for Effective WLAN Anomaly Detection based on Feature Construction & Reduction", (2012) Published At: IJCA, ISBN: 973-93-80869-55-7, Impact factor: 0.814

40. Damon McCoy, Doug Sicker &Dirk Grunwald, (2007) " A Mechanism for Detecting and Responding to Misbehaving Nodes in Wireless Networks" Published By: 1-4244-1268-4/07/$25.00 C) 2007 IEEE
41. George Lapiotis, Byungsuk Kim, Subir Das & Farooq Anjum (2005), "A Policy-based Approach to Wireless LAN Security Management ", Published By: 0-7803-9469-0/05/$20.00 ©2005 IEEE
42. Chirumamilla Mohan K & Ramamurthy Byrav (2003), "Agent Based Intrusion Detection and Response System for Wireless LANs", Published By: 0-7803-7802-4/03/$17.00 © 2003 IEEE
43. Daxin Tian, Qiuju Li & Songtao Chen (2008), "Anomaly Intrusion Detection Methods for Wireless LAN" Published By: Fourth International Conference on Natural Computation, IEEE Computer Society; 978-0-7695-3304-9/08 $25.00 © 2008 IEEE
44. Eric Bloedorn, Alan D. Christiansen, William Hill & Other Co Authors, "Data Mining for Network Intrusion Detection: How to Get Started";
45. Lee Wenke & Stolfo Salvatore J. (January 26-29, 1998) "Data Mining Approaches for Intrusion Detection ", Published By: The Proceedings of the 7th USENIX Security Symposium, San Antonio Texas; Available At: http://www.usenix.org/
46. Julisch Klaus (2002), Research Article on "DATA MINING FOR INTRUSION DETECTION A Critical Review"
47. Jian Pei, Upadhyaya Shambhu J., Faisal Farooq &Venugopal Govindaraju (2004), "Data Mining for Intrusion Detection: Techniques, Applications and Systems" Proceedings of the 20th International Conference on Data Engineering (ICDE'04); Published By: IEEE Computer Society
48. BRUGGER TERRY S. (2004), Research Article on "Data Mining Methods for Network Intrusion Detection"
49. Hall Jeyanthi, Michel Barbeau & Evangelos Kranakis; "Enhancing intrusion detection in wireless networks using Radio frequency fingerprinting (extended abstract)"
50. Kurt Derr & Milos Manic (2007); "Intelligent Control in Automation Based on Wireless Traffic Analysis" 1-4244-0826-1/07/$20.00 © 2007 IEEE
51. Veda Anshu, Kalekar Prajakta & Bodhankar Anirudha; "Intrusion Detection Using Data Mining Techniques"

52. Ian H. Witten, Eibe Frank, Len Trigg, Mark Hall, Geoffrey Holmes, and Sally Jo Cunningham; Article on “Weka: Practical Machine Learning Tools and Techniques with Java Implementations”
53. Thiprungsri Sutapat & Miklos A. Vasarhelyi.(2011); “Cluster Analysis for Anomaly Detection in Accounting Data: An Audit Approach“ Published By: The International Journal of Digital Accounting Research; Vol. 11, 2011, pp. 69 – 84 ISSN: 1577-8517
54. Mrs.P.Nancy & Ramani Geetha R. (2011); “A Comparison on Performance of Data Mining Algorithms in Classification of Social Network Data” Published By: International Journal of Computer Applications (0975 – 8887) Volume 32– No.8, October 2011
55. Growe Glenn A. (1999), Thesis on “Comparing Algorithms and Clustering Data: Components of the Data Mining Process”
56. Patel Ajay M. (2012), “Wi-Fi Deployments In Conjunction With Wi-Max For Next Generation Heterogeneous Network” Published At: IJRCM VOLUME NO. 2 (2012), ISSUE NO. 5 (MAY); ISSN 2231-1009