WWJMRD 2018; 4(4): 90-94
www.wwjmrd.com
International Journal
Peer Reviewed Journal
Refereed Journal
Indexed Journal
UGC Approved Journal
Impact Factor MJIF: 4.25
E-ISSN: 2454-6615

A Review on Intrusion Detection on Wi-Fi Network Using Hybrid Techniquesn

Maninder Jeet Brar, Dr. Sandeep Kautish

Maninder Jeet Brar
M.Tech Scholar, GKU,
Talwandi, Sabo, Punjab, India

Dr. Sandeep Kautish
Associate Professor, GKU,
Talwandi, Sabo, Punjab, India

Abstract
Presently a day there is colossal measure of Data being gathered and put away in database wherever over the globe. The security assaults can make serious disturbance information and systems. Along these lines, Intrusion Detection System (IDS) turns into a critical piece of each PC or system framework. Intrusion Detection (ID) is a system that gives security to the two PCs and systems. Highlight choice and highlight diminishment is essential region of research in interruption bearing framework. In this paper the different authors papers are reviewed and different problems are faced that are given in problem formulation. All these problems are resolved in future with the help of different methods.

Keywords: IDS, ANN, Data, security, data mining etc

I. Introduction

Now a day there is huge amount of data being collected and stored in database everywhere across the globe. The tendency is to keep increasing year after year. It is not hard to find database with Terabytes of data in enterprises and research facilities. There is invaluable and knowledge “hidden” in such database; and without automatic methods for extracting this information it is practically impossible to mine for them. Throughout the years many algorithms were created to extract what is called nuggets of knowledge from large set of data. There are several different methodologies to approach this problem: Classification, Association rule, Clustering etc.

Data mining is essential process where intelligent methods are applied extract data patterns. It is the process of discovering interesting pattern and knowledge from large amounts of data. The data source can include database, data warehouses, the web, other repositories, or data that are streamed into the system dynamically.

In data mining the data is stored electronically and the search is automated or at least augmented by computer. Data mining is about solving problems by analyzing data. It is popular due to the successful applications in telecommunication, marketing and tourism. In these days the usefulness of the methods has been proven also in medicine.

Data mining aims at discovering patterns in data which may be present in data. The process must be automatic or semiautomatic. The pattern discovered must be meaningful in that they lead to some advantage, usually an economic advantage. The data is invariably present in substantial quantities. These patterns, discovered in historical data, may be used to support future decision concerning diagnosing of new cases. Such knowledge. May also have an enormous value for decision making in treatment planning, risk analysis and other predictions. Useful patterns allow us to make nontrivial predictions on new data.

There are two extremes for the expression of a pattern: as a black box whose innards are effectively incomprehensible and as a transparent box whose construction reveals the structure of the pattern. Both are make good predictions.

Data mining is also known as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. It involves the use of sophisticated data analysis tool to discover previously unknown, valid pattern and relationships in large data set. These tools can include statistical models, mathematical algorithm and machine learning methods. Therefore, data mining consists of more than collection and managing data, it also includes analysis and prediction. Data mining is the process of extracting patterns from data.

Data mining offer promising ways to uncover hidden patterns within large amounts of data. These hidden patterns can potentially be used to predict future behaviour.

Correspondence:
Maninder Jeet Brar
M.Tech Scholar, GKU,
Talwandi, Sabo, Punjab, India

The availability of new data mining algorithms, however, should be met with caution. First of all, the techniques are only as good as the data that has been collected. Prior to the mining process it is essential to gain sufficient amount of data. Good data is the first requirement for good data exploration. This may require integrating data from multiple heterogeneous information sources and transforming it into a form specific to a target decision support application. After that the data has to be prepared for knowledge extraction. The next step is to choose the most appropriate technique to mine the data. However, there are tradeoffs to consider when choosing the appropriate data mining technique to be used in a certain application. The "best" model is often found by trial and error: Trying different technologies and the data mining process may be complex and can be divided into the following steps:

- Understanding the domain and data
- Creating the target data
- Data selection
- Data reduction and transformation
- Projection
- Attribute selection method
- Normalization
- Aggregation
- Choosing the data mining task
- Selecting the data mining algorithm/s
- Data mining
- visualization
- Interpreting mined patterns
- Consolidating discovered knowledge
- Evaluation of the result to an appropriate target.

II. Intrusion Detection System

An intrusion is an endeavor to trade off the uprightness, classification, accessibility of an asset, or to sidestep the security instruments of a PC framework or system. James Anderson presented the idea of interruption discovery in 1980 [15].It screens PC or system movement and distinguish malignant exercises that cautions the framework or system head against noxious assaults. Dorothy Denning proposed a few models for IDS. Methodologies of IDS in view of recognition are peculiarity based and abuse based interruption discovery. In inconsistency based interruption discovery approach, the framework initially takes in the ordinary conduct or action of the framework or system to distinguish the interruption. In the event that the framework goes astray from its typical conduct then an alert is delivered. In abuse based interruption discovery approach, IDS screens parcels in the system and contrasts and put away assault designs known as marks. The fundamental disadvantage is that there will be distinction between the new danger found and mark being utilized as a part of IDS for distinguishing the risk. Methodologies of IDS in view of area of observing are Network based interruption location framework (NIDS) and Host-based interruption recognition framework (HIDS) [16]. NIDS identifies interruption by checking system activity as far as IP parcel. HIDS are introduced locally on have machines and distinguishes interruptions by looking at framework calls, application logs, document framework alteration and other host exercises made by every client on a specific machine.

III. Feature Selection

Due to the large amount of data flowing over the network real time intrusion detection is almost impossible. Feature selection can reduce the computation time and model complexity. Research on feature selection started in early 60s [17]. Basically feature selection is a technique of selecting a subset of relevant/important features by removing most irrelevant and redundant features [18] from the data for building an effective and efficient learning model

Methods for Feature Selection: Blum and Langley divide the feature selection methods into three categories named filter, wrapper and hybrid (embedded) method.

(a)Filter method: Filter method uses external learning algorithm to evaluate the performance of selected features.

(b)Wrapper method: The wrapper method Wrap around‖ the learning algorithm. It uses one predetermined classifier to evaluate features or feature subsets. Wrapper algorithm uses a search algorithm to search through the space of possible features and evaluate each subset by running a model on the subset. Many feature subsets are evaluated based on classification performance and best one is selected. This method is more computationally expensive than the filter method

(c)Hybrid method: The hybrid method [19] [20] combines wrapper and filter approach to achieve best possible performance with a particular learning algorithm.

IV. Datasets

The KDD CUP 1999 benchmark datasets are used in order to evaluate Hybrid feature selection method for Intrusion detection system. It consists of 4,940,000 connection records. Each connection had a label of either normal or the attack type, with exactly one specific attack type falls into one of the four attacks categories as: Denial of Service Attack (DoS),
User to Root Attack (U2R), Remote to Local Attack (R2L) and Probing Attack.

Denial of Service Attack (DOS): Attacks of this type deprive the host or legitimate user from using the service or resources.

Probe Attack: These attacks automatically scan a network of computers or a DNS server to find valid IP addresses.

Remote to Local (R2L) Attack: In this type of attack an attacker who does not have an account on a victim machine gains local access to the machine and modifies the data.

User to Root (U2R) Attack: In this type of attack a local user on a machine is able to obtain privileges normally reserved for the super (root) users. Each connection record consisted of 41 features and are labeled in order as 1,2,3,4,5,6,7,8,9,.....,41 and falls into the four categories are shown in Table 1:

Category 1(1-9): Basic features of individual TCP connections.

Category 2 (10-22): Content features within a connection suggested by domain knowledge. **Category 3** (23-31): Traffic features computed using a two-second time window.

Category 4 (32-41): Traffic features computed using a two second time window from destination to host.

Distribution of intrusion types in datasets

Dataset	Normal	Probe	DOS	U2R	R2L	Total
(kddcup. data)	97280	4107	391458	52	1124	494020

Here the author evaluate AWID Dataset [8] as a benchmark dataset. The dataset was published in 2015 with huge and real Wi-Fi network traces. Due to its comprehensiveness and real characteristics, the AWID dataset might become the common benchmark dataset for Wi-Fi networkrelated researches. We use AWID-CLS-R-Trn and AWID-CLSR-tst for training and test dataset, respectively. There are 1,795,575 instances in the training dataset with 1,633,190 and 162,385 normal and attack instances, respectively. While the test dataset contains 575,643 instances with 530,785 and 44,858 normal and attack instances, respectively.

V. Literature Survey

Muhamad Erza Aminanto et.al.[2017] have contemplated the element weighting techniques in existing machine students and take a gander at how they could be utilized for the precise determination of the essential highlights. So as to approve our thought, we consider Wi-Fi systems since unavoidable Internet-of-Things (IoT) gadgets make immense traffics and powerless in the meantime. Identifying known and obscure assaults in Wi-Fi systems stays incredible testing assignments. We test and approve the plausibility of the chose highlights utilizing a typical neural system. This investigation exhibits that the proposed weighted-based machine learning model can beat other channel based element choice models. The trial comes about not just show the viability of the proposed demonstrate, accomplishing 99.72% F1 score, yet in addition demonstrate that consolidating a weight-based element determination strategy with a light machine-learning classifier which prompts fundamentally enhanced execution, contrasted with the best outcome detailed in the literature.[1]

Aditya Shrivastava1 et.al [2013] have proposed a half and half model for include choice and interruption identification. Highlight choice is vital issue in interruption identification. The choice of highlight in assault trait and typical movement quality is testing errand. The choice of known and obscure assault is additionally confronted an issue of order. PCNN is dynamic system utilized for the procedure of highlight choice in grouping. The dynamic idea of PCNN select characteristic on determination of entropy. The characteristic entropy is high the element estimation of PCNN organize is chosen and the property estimation is low the PCNN highlight selector diminishes the estimation of highlight determination. After determination of highlight the Gaussian piece of help vector machine is incorporated for grouping. Identification rate is high in pressure of other neural system model, for example, RBF neural system and SOM arrange. [3]

JAYSHRI R. PATEL et.al [2013] proposed a technique utilizing Decision Trees order of Intrusion location, as indicated by their highlights into either nosy or non meddling class is a broadly examined issue. Choice trees are helpful to recognize interruption from association records. In this paper, we assess the execution of different choice tree classifiers for ordering interruption recognition information. The point of this paper is to explore the execution of different choice tree classifiers for positioned interruption identification information. Information Gain is utilized to give positioning to interruption identification information. Choice tree classifiers assessed are C4.5, CART, Random Forest and REP Tree. [4]

Megha Aggarwal et.al [2013], displayed there is a sensational increment in development of Computer systems. There are different private and also government associations that store significant information over the system. This enormous development has postured testing issues in system and Data security, and identification of security dangers, regularly alluded to as interruption, has turned into an essential and basic issue in system, information and Data security. The security assaults can make extreme disturbance information and systems. In this manner, Intrusion Detection System (IDS) turns into an imperative piece of each PC or system framework Intrusion location (ID) is a component that gives security to the two PCs and systems. [2]

Venkata Suneetha Takkellapati et.al [2012] proposed as the cost of the information preparing and Internet openness builds, an ever increasing number of associations are getting to be helpless against an extensive variety of digital dangers. Most present disconnected interruption identification frameworks are centered around unsupervised and regulated machine learning approaches. Existing model has high blunder rate amid the assault order utilizing bolster vector machine learning calculation. In addition, with the investigation of existing work, include choice methods are likewise fundamental to enhance high proficiency and adequacy. Execution of various kinds of assaults discovery ought to likewise be enhanced and assessed utilizing the proposed approach. In this proposed framework, Data Gain (IG) and Triangle Area based KNN are utilized for choosing more discriminative highlights by consolidating Greedy k-implies bunching calculation and SVM classifier to identify Network assaults. This framework accomplishes high precision discovery rate and less mistake rate of KDD CUP 1999 preparing informational index. [5]

Comparison Table

Name of Author	Technique Used	Result
Muhamad Erza Aminanto et.al.[2017]	weighted-based machine learning model	accomplishing 99.72% F1 score
Ms. Rohini A. Et.al. [2016]	Intrusion Detection system	Defining architecture of IDS
Dr. S.Vijayarani et.al.[2015]	ID Life cycle	Improved security in corporate world and for network users

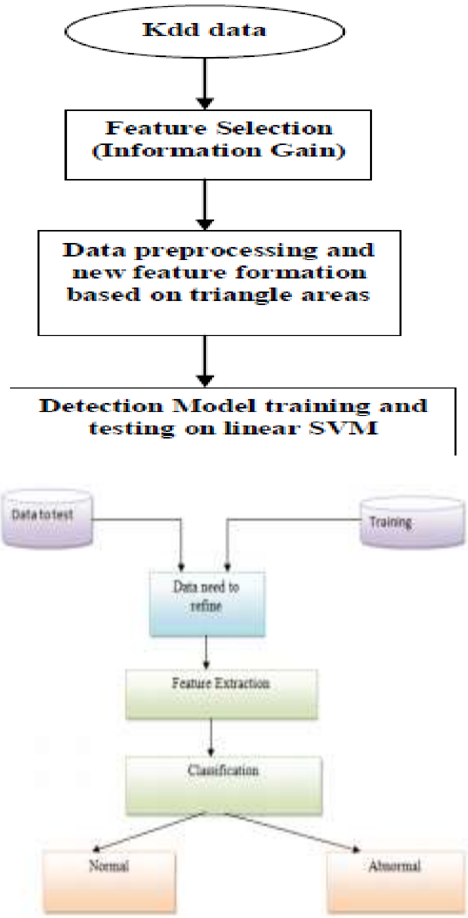
Aditya Shrivastava1 et.al [2013]	PCNN	Identification rate is high in pressure of other neural system model
Jayshri R. Patel et.al [2013]	C4.5, CART, Random Forest and REP Tree.	Controlling the intrusion
Megha Aggarwal et.al [2013],	Intrusion Detection System (IDS)	gives security to the two PCs and systems
Venkata Suneetha Takkellapati et.al [2012]	KNN	High precision discovery rate and less mistake rate of KDD CUP 1999 preparing informational index.

Discussion

This work is to detect the intrusion from network. It is based upon weka tool. There are the programmable files containing the Data about the dataset. The Intrusion detection system deals with large amount of data which contains various irrelevant and redundant features resulting in increased processing time and low detection rate. Therefore feature selection plays an important role in intrusion detection. There are various feature selection methods proposed in literature by different authors. In this a comparative analysis of different feature selection methods are presented on KDDCUP'99 benchmark dataset and their performance are evaluated in terms of detection rate, root mean square error and computational time.

As the network environment has grown rapidly, so has the problem of intrusions. MIT kdd99 dataset is currently available approaches to dealing with intrusions can be categorized.

Our proposed approach is evaluated in several experiments. First, we verify two feature selection approaches: filter-based and wrapper-based methods, which are implemented in the Waikato Environment for Knowledge Analysis (WEKA). Second, we implement the ANN classifier using MATLAB R2016a running on an Intel Xeon E-3-1230v3 CPU @3.30 GHz with 32 GB RAM.


Kdd data
Feature Selection (Information Gain)
Data preprocessing and new feature formation based on triangle areas
Detection Model training and testing on linear SVM
Data to test
Training
Data need to refine
Feature Extraction
Classification
Normal
Abnormal

Conclusion

Data mining is also known as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. It involves the use of sophisticated data analysis tool to discover previously unknown, valid pattern and relationships in large data set. These tools can include statistical models, mathematical algorithm and machine learning methods. Therefore, data mining consists of more than collection and managing data, it also includes analysis and prediction. Data mining is the process of extracting patterns from data. Intrusion Detection System (IDS) becomes an important part of every computer or network system. Intrusion detection (ID) is a mechanism that provides security for both computers and networks. Feature selection and feature reduction is important area of research in intrusion direction system. The size and attribute of intrusion file are very large. Due to large size of attribute the detection and classification mechanism of intrusion detection technique are compromised in terms of detection rate and alarm generation. There are different problems that are reviewed in the problem formulation and all these problems resolved in future with the help of different techniques.

References

1. Muhamad Erza Aminanto et.al. "Wi-Fi Intrusion Detection Using Weighted-Feature Selection for Neural Networks Classifier" IWBIS 2017.
2. Megha Aggarwal et.al "Performance Analysis of Different Feature Selection Methods in Intrusion Detection", International Journal of Scientific & Technology Research Volume 2, Issue 6, June 2013.
3. Aditya Shrivastava et.al "A Novel Hybrid Feature Selection and Intrusion Detection Based on PCNN and Support Vector Machine" Aditya Shrivastava et al, Int.J.Computer Technology & Applications, Vol 4 (6), 922-927, IJCTA | Nov-Dec 2013.
4. Jayshri R. Patel et.al "Performance Evaluation of Decision Tree Classifiers for Ranked Features of Intrusion Detection" Journal of Data, Knowledge and Research in Data Technology, ISSN: 0975 – 6698| NOV 12 TO OCT 13.
5. Venkata Suneetha Takkellapati et.al "Network Intrusion Detection system based on Feature Selection and Triangle area Support Vector Machine" International Journal of Engineering Trends and Technology- Volume3Issue4- 2012.
6. Xing, Eric P., Michael I. Jordan, and Richard M. Karp. "Feature selection for high-dimensional genomic microarray data." In ICML, vol. 1, pp. 601-608. 2001.
7. John, George H., Ron Kohavi, and Karl Pfleger. "Irrelevant features and the subset selection problem." In Machine Learning Proceedings 1994, pp. 121-129. 1994.

8. Dash, Manoranjan, and Huan Liu. "Feature selection for classification." Intelligent data analysis 1, no. 3 (1997): 131-156.
9. Panda, Mrutyunjaya, and Manas Ranjan Patra. "Network intrusion detection using naive bayes." International journal of computer science and network security 7, no. 12 (2007): 258-263.
10. Nguyen, Hai Thanh, Katrin Franke, and Slobodan Petrovic. "Towards a generic feature-selection measure for intrusion detection." In Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 1529-1532. IEEE, 2010.
11. Gong, Shangfu, Xingyu Gong, and Xiaoru Bi. "Feature selection method for network intrusion based on GQPSO attributes reduction." In Multimedia Technology (ICMT), 2011 International Conference on, pp. 6365-6368. IEEE, 2011