ISSN: 0975-766X
CODEN: IJPTFI
Available Online through ***Research Article***
www.ijptonline.com

DETECTION OF ATTACKS ON MANET USING SEQUENTIAL PATTERN MINING WITH FEATURE SELECTION

Fidalcastro.A [1], Baburaj.E[2], and Saleem Babu.M[3]
[1] Research Scholar, Correspondent Author, Sathyabama University, Department of Computer Science and Engineering, Chennai, India.
[2] Professor, Narayanaguru College of Engineering, Department of Computer Science and Engineering Kanyakumari Dist,.Nagercoil, India.
[3]Associate Professor, Jaya Engineering College, Department of Computer Science and Engineering, India.
Email: fidalcastro@gmail.com

Received on: 20.10.2016 *Accepted on: 25.11.2016*

Abstract

Challenges in designing a MANET is protecting from various attacks in the network. A novel approach using Sequential pattern mining algorithm with feature selection used to detect attacks in IDS. **S**equential pattern mining rules are generated to detect the misbehavior activity in the network using Intrusion Detection System. The sequential pattern mining algorithms GSP, SPAM is used to generate Sequential pattern rules to determine the sequences with attacks. Most important features selected using decision trees from the intrusion log. Its performance on simulated networks with varying traffic and mobility patterns and uses advanced form of Decision tree called decision tree based feature selection which improves the accuracy of detection of attacks. Many existing mining algorithms perform well for dense and large sequences, having costly scans using 'generate candidate and test' approach. Current mining algorithms can generate large number of candidates and very slow to generate rules or generate few results, omitting some interesting and valuable information with large amount of infrequent candidates. To overcome the above problems we propose a novel structure Feature Selection using Decision Tree (FSDT) for selection of candidates. This paper explains how FSDT is useful to generate candidates and avoids infrequent candidates. Our experimental results shows: 1) FSDT candidate generation and pruning is effective, 2)FSDT with GSP, SPAM is compact and outperforms the state of the art algorithms, and 3)The FSDT-GSP, FSDT-SPAM provide more accurate patterns that detect the attacks in an environment which provides massive amount of data. Our aim is to produce better throughput and scalability than the original algorithms. Experimental results show that the algorithm has higher scalability, better accuracy and faster than the state of art algorithms in detecting the attacks.

Keywords: Sequential Pattern Mining, Feature Selection, Intrusion Detection System, MANET, Decision trees.

1. Introduction

A Mobile Ad-hoc Network is an infra-structure less network of mobile devices linked by wireless nodes. The mobile nodes in MANET are recurrently moving in Parallel with the network, the nodes are self-configurable and the data must be routed via intermediate nodes, the nodes can act as a router/host for data transmission, it doesn't have a centralized monitoring system. A routing protocol are shared the neighbor information among instant neighbors, and then pass the information throughout the network. This way, routers gains knowledge of the topology of the network[1], Multi cast Ad hoc On Demand Distance Vector (MAODV) is a multi cast expansion of AODV protocol, and tree based functionality to route data among the nodes in MANET is provided. The route discovery is based on Route Request and Route Reply process.

The applications of MAODV are military operations, rescue operations, etc. The most important feature in MANET is mobility, represents the movement of nodes, location and other related features to the network. To overcome the security concerns in MANET, Intrusion Detection System (IDS) is implemented in each host to monitor the system activities for malicious activities or policy violations and alerts the system. A software application or a device that keep tracking the network or system activities to detect the malicious activities is known as Intrusion Detection System (IDS). The intrusion detection system is devised into two types anomaly detection and misuse detection and it is classified into two types namely network based (NIDS) and host based (HIDS) intrusion detection systems. The false alarms often rises due to the anomaly detection is able to identify new types of attack. To identify new types of attacks misuse detection is unable, but it has a high correct-detection rate for known types. Performance of IDS generally depends on the length of dataset having a lot of features used to identify attack or normal type. So, the reduced dataset with relevant features help an intrusion detection system to perform its' task easily with a better use of time and memory and to find the hidden relationship among the features and datasets, Decision tree based feature selection and reduction approaches can be used. The decision tree provides a smooth transition between member and non-member of a set; therefore, there are fewer boundary elements being excluded. In our research work, By FSDT the greatest need was to reduce the amount of data needed for processing and the false alarm rate. Here to improve the performance of a system and try to replace current intrusion detection methods with a feature selection and data mining approach. Data mining is the process of getting useful information from the large datasets. Sequential Pattern mining is finding statistically relevant patterns between data examples where the values are delivered in a sequence. Sequential pattern mining is one of the special cases of structured data mining. Intrusion logs

grows large and it needs large amount of space and processing such huge data requires lots of time and resources and the intrusions to be found in time. The sequential mining algorithms generate large amount of sequences and it is impossible to find intrusions with the large sequences. So we need the interesting rules and non redundant rules which will find the intrusions.

GSP Algorithm (Generalized Sequential Pattern algorithm) is an algorithm used for sequence mining. The algorithms for solving sequence mining problems are mostly based on the a priori (level-wise) algorithm. The algorithm is used for multiple passes are made over the data. Seed set is formed in the first pass by GSP. In our work decision tree feature selection is used for formulating the seed set. Seed set is specific to IDS and more specific features are found which increases the detecting ability. The seed set found in the decision tree is used to generate candidate sequences in subsequent passes. At the end of the pass, the algorithm determines which of the candidate sequences are really frequent. These frequent candidates become the seed for fore coming passes. The algorithm terminates when there are no frequent sequences or no candidate sequences generated. For mining sequential patterns SPAM first uses depth-first search strategy.

SPAM generates sequential patterns of different length which is an additional salient feature. It is done with three main steps Lexicographic Tree for Sequences, Depth First Tree Traversal, Pruning (S-step Pruning, I-step Pruning). In our work in first step while generating the tree for sequences, we apply seed set generated by the decision tree feature set. In the first stage of generating the tree sequences the candidates are compared with seed set. If the candidates are found in seed set then the nodes are generated otherwise discarded. So, unrelated sequences are avoided to found the actual attack sequences in IDS..

In this paper, we present a solution to this issue based on the study of item occurrences of most used features. Our contribution is. First, to find the feature set information, we introduce a Decision tree based Feature selection for finding the features (FSDT). FSDT is a compact and accurate structure, which can be built with a single database scan. Second, we propose a generic candidate pruning mechanism for vertical sequential pattern mining algorithms based on the FSDT data structure. We describe how the pruning mechanism is integrated in three state-of-the-art algorithms GSP, SPAM. Third, we perform a wide experimental evaluation on our datasets generated through various attack scenario generated. Results show that the modified algorithms (1) prune a large amount of candidates using feature set, (2) and faster than the corresponding original algorithms (3) These two modified algorithm applied and compared with the original algorithm. The rest of the paper is organized as follows. Section 3 defines the Decision

tree selection FSDT Section 4 defines the problem of sequential pattern mining and reviews the main characteristics of GSP, SPAM. Section 5 describes the FSDT structure, the pruning mechanism, and how it is integrated in GSP, SPAM. Section 6 presents the modified algorithm compared with GSP, SPAM with our experimental study. Finally, Section 6 presents the conclusion.

2. Related Work

In the proposed algorithm[1] which generates the Generalized Sequential Patterns. The number of data sequences scales linearly with GSP, and consists of average data sequence size with respect to very good scale up properties. In a proposed[2] an efficient algorithm called Sequential Pattern Mining (SPAM) that contributes into a practical algorithm by integrating a variety of old and new algorithms. SPAM assumes that it fit into main memory for all data structures used for the algorithm and the entire database completely. Memory-resident is due to the size of current main memories reaching gigabytes and growing, many moderate-sized to large databases. For mining sequential patterns SPAM is one of the best depth first search strategy. The two step process is I-step process and S-step process which is based on Candidate generation. .

In a work proposed[3] a discovery algorithm for Sequential Patterns .This algorithm with respect to the a number of other database parameters and number of input-sequences has linear scalability, and a parallel algorithm for fast discovery of frequent sequences in large databases is named as pSPADE. Each class can be solved in main-memory using simple join operations and efficient search techniques. The observer uses new structure for storing co-occurrence information named CMAP (Co-occurence MAP) [4]. To candidates can be pruned in three state-of-the-art vertical algorithms, namely SPADE, SPAM and ClaSP CMAP can be used. An extensive experimental study with three real-life datasets shows that (1) squeezed together (CAMP) (2) co-occurrence-based pruning is effective and that (3) the resulting algorithms outperform state-of-the-art algorithms for mining sequential patterns (GSP, PrefixSpan, SPADE and SPAM) and closed sequential patterns (ClaSP and CloSpan). In a proposed a methodology[5], to detect intrusions with the help of Intrusion Detection System using k-means clustering, which is employed in MANET This kind of Data Mining approach helps in evaluating the detection rate, improving the performance of network and control false dismissals and false alarm rate [5] .

In a work reported[6] the node features are analyzed for intrusion detection in MANET. Principal Component Analysis technique and Profile based monitoring techniques are used and implemented in the network environment to measure the performance of network.

Based on the routing protocol used in the MANET the performance is varied. In different scenarios the performance is evaluated. In Scenario 1, the state of the network operation is presented normally, so it does not come under attacks, whereas different traffic patterns of network are generated. In next scenario in Ad-hoc Network the Denial-of-Service Attack is deployed, it falls under two categories. In first attack, the intruder first attacks other nodes in the network and interrupts its services. The second type is Distributed DoS (DDoS); here more than one intruder attacks the network. The flooding of Route REQuest (RREQ) packets from a compromised node is the reason for the attacks [6]. In a work proposed by [7] to target the security of AODV protocol designed for MANET an Intrusion Detection System (IDS) is implemented. Here Multiple Static Agent is designed with IDS to monitor the route establishment process in the network. In a work [8] defines a hybrid data mining approach for Intrusion Detection System which coordinates the feature selection, filtering, clustering, divide and merge and clustering. The detection rate is improved and decrease the false alarm rate is attained. The intrusion detection is processed based on clustering analysis technique. The intrusion detection collects the features, and then it is filtered to remove the noise and outliers for cluster formation based on the k-means and divide the nodes as normal and malicious nodes, with these able to detect false alarm rate and detection rate.

In a work [9], an approach for handling Intrusion Detection System (IDS) of association rule mining to IDS alerts based on the features. A fuzzy association rules are integrated to detect the misbehavior activity in the network using Intrusion Detection System. The Apriori Algorithm is used to generate association rules to determine the features, based on the features creating a cluster using K-means clustering technique. The k-means algorithm is an evolutionary algorithm, where k is input value defined by user based on the number of clusters formed. The k-means clustering is designed using fuzzy rules for automatic generation of fuzzy rules to provide effective learning.

These rules are to find intrusion in the network. The observer uses a protocol which uses a specification based Intrusion Detection System to identify the misuses happening in routing the messages. The IDS is processed as a two step process. The First step is using the clustered networking monitoring selection algorithm in network monitoring nodes[10], and the second step is the NM nodes will run the monitoring protocol its responsibility is to observe the flow of data. The experiment is carried on two environments, where in the first script the nodes are static, and the second script is mobile, which is to measure the performance of the network. Among these defect prediction/faultproneness is an important issue.

It can be used in assessing the final product quality, estimating the standards and satisfaction of customers. Fault

proneness can also be used for decision management with respect to the resource allocation for testing and verification. It is also one of the quality classification tasks of software design in which prediction of fault prone modules in the early design phase emphasizes the final quality outcome within estimated time and cost [11.] The observer uses evolutionary computation techniques using genetic algorithm to implement the challenges in MANET and to analyze the power consumption using intrusion detection. The paper uses advanced form of genetic algorithm called genetic-based feature selection which improves the accuracy of the data [12]. In a work reported by Sun B, Wu K, Pooch U[13] this paper proposes the zone based intrusion detection system in which it divides the network into zones and also proposes the feature selection and it detects the flooding attack with aggregated algorithm. In a work reported[14], this paper uses advanced form of genetic algorithm called genetic-based feature selection which improves the accuracy of the data. Feature selection or feature extraction have to be employed. Much research work has been done on this dimensionality reduction[15]. The papers explains about paper ranking using feature selection, various data mining techniques, Regression and classification techniques[16 ,17, 18, 19, 20, 21, 22 .] The papers explain about Decision tree, MANET routing and various tools in data mining[23, 24 25] .

3. Proposed Work

3.1 Architecture Diagram

Figure 1. Architecture Diagram.

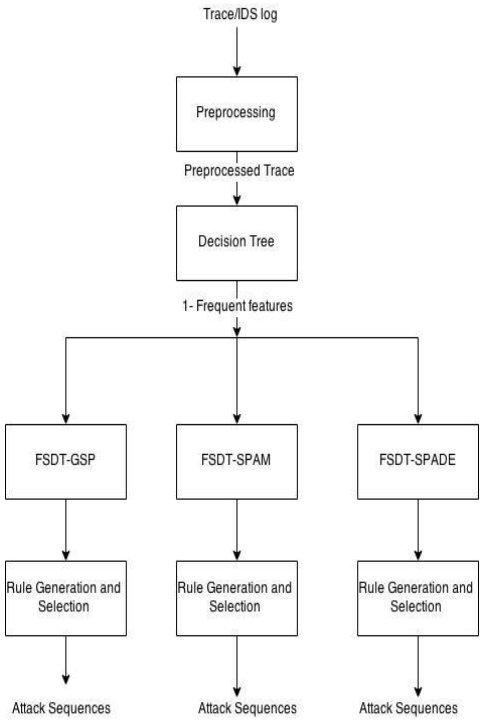

Figure 1. Explains the architecture diagram which explains how FSDT-GSP and FSDT-SPAM finds attack sequences from the IDS log.

3.2 FSDT Feature Selection with Decision Trees (FSDT)

For Feature selection we use Decision tree induction to find the relevant and important features from Intrusion Detection System dataset. Its approach is to learn from decision tree classifiers. Internal node or non leaf node represents test of an attribute. Branch node represents the output of the test. Class prediction is represented by external node or leaf node. The best attribute is chosen, and based on it Decision tree algorithm partitions the data into individual classes. The best attribute is selected by Information gain using attribute selection process.

The information gain of the attribute is calculated by

$$E(s) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

The generic algorithm of Decision Tree is used to find the features.

Generating a decision tree from training tuples of data partition D

Algorithm: FSDT

Input given: attribute_list, the set of candidate attributes.

Attribute selection method, Splitting criteria method.

Data partition, D, a set of training tuples and their associated class labels.

Output Expected: A Decision Tree

Each path from the root to leaf node creates a rule. The rule antecedent is created by ANDing each splitting criteria and the path. The leaf node has the class predictions, forming the rule consequent. The features which appeared in the extracted rules are selected as the relevant and important features for Intrusion detection. All the other features are considered as irrelevant. Both an increase in detection rate and a decrease in false positive rate are expected.

Table 1. Number of Features.

S.No.	Features of a Node
1	Number of neighbours
2	Number of added neighbours
3	Number of removed neighbours
4	Number of active routes
5	Number of invalidated routes
6	Number of added routes by route discovery mechanism
7	Number of updated routes
8	Number of added routes under repair

9	Number of received route request packets destined to this node
10	Number of received route request packets to be forwarded by this node
11	Number of broadcasted route request packets from this node
12	Number of forwarded route request packets from this node
13	Number of received route reply packets destined to this node
14	Number of received route reply packets to be forwarded by this node
15	Number of received broadcast route error packets (to be forwarded or not)
16	Number of broadcasted route error packets from this node
17	Number of received total routing protocol packets
18	Number of routes under repair
19	Number of added routes by overhearing
20	Number of invalidated routes due to expiry
21	Number of invalidated routes due to other reasons
22	Number of received route reply packets to be forwarded by this node
23	Number of initiated route reply packets from this node
24	Number of received total routing protocol packets to be forwarded
25	Number of initiated total routing protocol packets from this node
26	Number of forwarded total routing protocol packets by this node

Table 1 shows the total number of features available in MANET.

The features selected here by FSDT are as follows.

1. Number of neighbours

2. Number of added neighbours

3. Number of updated routes (modifying hop count, sequence number)

4. Number of added routes under repair

5. Number of broadcasted route request packets from this node

6. Number of forwarded route request packets from this node

3.3 FSDT GSP

Generalized Sequential Pattern algorithm is an algorithm used for sequential mining to find the attacks in the IDS log. A level-wise algorithm is used[1]. Here in FSDT-GSP the first level of seed set is the output FSDT and is given as the input to the first level to find the candidates and second and third levels follows the same.

Level wise paradigm is used here and first all the frequent items are discovered in level wise fashion. Frequent items found by FSDT seed set. The first pass of FSDT-GSP algorithm replaced and giving relevant features from the FSDT

and fore coming passes. FSDT GSP Algorithm makes multiple database passes. In first pass Using FSDT Seed set is taken and most relevant features taken into account of candidate generation. From the frequent items, a set of candidate 2-sequences are formed and 2 sequence also from the FSDT because first pass done with the same and another pass is made to identify their frequency. The frequent 2-sequences are used to generate the candidate 3-sequences and 3 sequence form FSDT, and this process is repeated until no more frequent sequences are found. There are two main steps in the algorithm.

Candidate Generation. Given the set of frequent (k-1)-frequent sequences F(k-1) from FSDT, the candidates for the next pass are generated by joining F(k-1) with itself. A pruning phase eliminates any sequence, Subsequences is not frequent and not available in seed set.

Support Counting. A hash tree–based search is employed for efficient support counting. Non frequent items are avoided by FSDT in first pass itself.

The modified algorithm

F1 = sequence from FSDT seed

k=2,

do while F(k-1)!= Null;

Generate candidate sets Ck (set of candidate k-sequences);

For all input sequences s in the database D

do

Increment count of all a in Ck if s supports a

Fk = {a ∈ Ck such that its frequency exceeds the threshold}

k= k+1;

Result = Set of all frequent sequences is the union of all Fks

End do

End do

3.4 FSDT Spam

SPAM is a sequential mining algorithm for finding all frequent sequences within a transactional database. The algorithm is specifically efficient when the databases or datasets are very long contains very long sequential patterns. A depth-first search strategy is used to generate candidate sequences, and various pruning mechanisms are

implemented to reduce the search space. It's a type of vertical sequence mining algorithm. The transactional data is stored using a vertical bitmap representation, which allows for efficient support counting and significant bitmap compression. Here FSDT-SPAM has two types of sequences the one is FSDTsequence-extended sequence and an FSDTitemset-extended sequence. An FSDTsequence-extended sequence is a sequence generated by adding a new transaction consisting of a single item to the end of its parent's sequence in the tree. In our approach sequence generated by adding a new transaction taken FSDT seed set. The first leaf nodes of empty set in the FSDT-SPAM and next nodes generated only with the FSDT seed set. [2]

FSDTitemset-extended sequence is a sequence generated by adding an item to the last itemset in the parent's sequence, such that the item is greater than any item in that last itemset. Accordingly item extended sequence generated only from the FSDT seeds. If we generate sequences by traversing the tree, then each node in the tree can generate FSDTsequence-extended children sequences and FSDTitemset-extended children sequences with all features selected from FSDT. So unimportant features are removed. Only generated sequences are contains only related with network specific attacks. Unrelated rules are avoided. We refer to the process of generating FSDTsequence-extended sequences as the FSDTsequence-extension step only with most important features are used, and we refer to the process of generating FSDTitemset-extended sequences as the FSDTitemset-extension step. Thus we can associate with each node n in the tree two sets: FSn , the set of candidate items that are considered for a possible FS-step extensions of node n, and In which identifies the set of candidate items that are considered for a possible FI-step extensions. The top element in the tree is the null sequence and each lower level k contains all of the k-sequences, which are ordered lexicographically with FSDTsequence-extended sequences ordered before FSDTitemset-extended sequences. Each element in the tree is generated only by either an FS-step or an FI-step,

FSDT-SPAM traverses the sequence tree described above in a standard depth-first manner. At each node n, the support of each FSDTsequence-extended child and each FSDTitemset-extended child is tested. If the support of a generated sequence s is greater than or equal to minSup, we store that sequence and repeat DFS recursively on s. If the support of s is less than minSup, then we do not need to repeat DFS on s by the Apriori principle, since any child sequence generated from s will not be frequent. If none of the generated children are frequent, then the node is a leaf and we can backtrack up the tree. Here the search space is huge. SPAM have s step pruning and i step pruning. But here in FSDT only takes the features set which are related to the attacks and we are considering the unrelated features and tree generated with the use of only FSDT seed set. Almost pruning is done. Here FSDT-SPAM no need of S step

pruning I step pruning.

4. Experimental Classification Results and Analysis

The Simulation environment is carried out in NS-2 simulator installed in Linux Operating System. The Scenario consists of 50 wireless nodes. For routing the data MAODV routing protocol is used. MAODV is the multicast extension of AODV protocol, whereas AODV protocol is for unicast and MAODV for multicast traffic.

The Simulation Environment is defined and Only multicast traffic exists in the simulation. After creation of ns-2 simulation environment, the scenario files to be generated for mobile node movement and CBR traffic pattern. NS2 window which contains the network with spoofing and Block hole attack present scenes. The preprocessed NS2 trace file is the input to FSDT-GSP, FSDT-SPAM algorithm to find features and optimized rules to detect the above attacks. The above techniques are developed using Java. The algorithm GSP, SPAM source code taken from http://www.philippe-fournier-viger.com/spmf/ and FSDT-GSP, FSDT-SPAM algorithms are developed and compared with preprocessed synthetic log generated by the above mentioned scenario.

Table-2. Results with the sequential Pattern mining algorithm.

Simulation	Detection Rate	False Positive Rate
high traffic, high mobility	97.63%	1.00%
medium traffic, high mobility	98.87%	0.45%
high traffic, medium mobility	98.78%	1.02%
Medium traffic, medium mobility	99.34%	0.64%
high traffic low mobility	95.67%	1.28%
medium traffic ,low mobility	97.86%	0.34%

Table 3. Results with FSDT-GSP, FSDT-SPAM.

Simulation	Detection Rate	False Positive Rate
high traffic ,high mobility	97.76%	0.91%
medium traffic ,high mobility	98.87%	0.45%
high traffic, medium mobility	98.86%	0.90%
Medium traffic, medium mobility	99.45%	0.55%
high traffic, low mobility	95.67%	1.00%
medium traffic, low mobility	97.86%	0.30%

Table 2, 3 shows that State of art algorithms GSP, SPAM are applied and compared with FSDT-GSP, FSDT-SPAM

Accuracy improved significantly.

5. Conclusion

The detection of attacks in Mobile Ad hoc networks with FSDT based algorithms is a novel kind of approach in wireless networks, and the approach helps in detecting attacks by generating optimized sequential rules based on the selection of relevant features for the nodes. The performance of mining algorithm has increased with the reduced feature set and optimized rules. Both increase in detection rate and decrease in false positive rate are observed. The FSDT with GSP, SPAM is efficient and when comparing with state of the art algorithms GSP, SPAM. It gives better performance, accuracy for MANET with spoofing, Black hole and Flooding attacks. It is advantageous over other classical sequential mining algorithms in detecting attacks with a huge amount of volatile log files like Big Data thanks to its speed and accuracy.

References

1. R. Srikant and R. Agrawal. Mining Sequential Patterns: Generalizations and Perforrmance Improvements. Research Report RJ 9994, IBM Almaden Research Center, San Jose, California, December 1995.
2. Agrawal, R., Ramakrishnan, S: Mining sequential patterns. In: Proc. 11th Intern. Conf. Data Engineering, pp. 3–14. IEEE (1995)
3. M. J. Zaki. Spade: An efficient algorithm for mining frequent sequences. Machine Learning, 42(1/2):31–60, 2001.
4. ClaSP: An Efficient Algorithm for Mining Frequent Closed Sequences Antonio Gomariz1,, Manuel Campos2, Roque Marin1, and Bart Goethals3 J. Pei et al. (Eds.): PAKDD 2013, Part I, LNAI 7818, pp. 50–61, 2013.
5. Preetee K.Karmore, Smita M. Nirkhi, “Detecting Intrusion on AODV based Mobile Ad Hoc Networks by k-means Clustering method of Data Mining”, IJCSIT Vol. 2(4),2011 PP.1774-1779
6. Peyam Kabiri and Mehran Aghaei, “Feature Analysis for Intrusion Detection in Mobile Ad-hoc Network”s, Internal Journal on Network Security, Vol 12, No.1, PP.42-49, Jan 2011
7. Hoda M.Hassan, Mohy Mahmoud, and Sherif El-Kassas, “Securing the AODV Protocol Using Specification-Based Intrusion Detection”, ACM 1-59593-486-3/06/0010, PP.33- 36, 2006.
8. Sadia patka, “Intrusion Detection Model Based on Data Mining Technique”, IOSR Journal of Computer Science 2014, PP.34-39
9. Vikas Markam and Shirish Mohan Dubey, A General Study of Association Rule Mining in Intrusion Detection

System, International Journal of Emerging Technology and Advanced Engineering, Jan 2012.

10. Karchirski O and Guha R, "Effective Intrusion Detection using Multiple Sensors in Wireless Ad Hoc Networks" In Proceedings of the 36 Hawaii International Conference on System Sciences(HICSS'03), p.57.
11. Raimund Moser, Witold Pedrycz, Giancarlo Succi, A Compariive Analysis of the Efficiency of Change Metrics and Static Code Atributes for Defect Prediction, ICSE'08, PP 181-190, May 10-18, 2008, Germany .
12. Sevil Sen and John A. Clark, "Evolutionary Computation Techniques for Intrusion Detection in Mobile Ad Hoc Networks" In proceedings of computer networks, May 10.2011.
13.] Sun B, Wu K, Pooch U (2003) Zone-based intrusion detection for mobile ad hoc networks.Int J of Ad Hoc and Sens Wirel Netw 2.
14. L. Me, Gassata, "A genetic algorithm as an alternative tool for security audit trails analysis", In proceedings of the International Symposium on Recent Advances in Intrusion Detection (RAID), Spinger, 1998, pp 1-11.
15. Almuallim, H., and Dietterich, T.G., Efficient algorithms for identifying relevant features In Proceedings of Ninth Canadian Conference on Artificial Intelligence,Vancouver,BC:Morgan Kaufmann,1992.
16. Jayshri R.Patel, "Performance Evaluation of Decision Tree Classifiers for Ranked Features of Intrusion Detection", Journal of Information, Knowledge and Research in Information Technology, ISSN: 0975 – 6698, VOLUME – 02, ISSUE – 02.
17. Reema Patel, Amit Thakkar, Amit Ganatra, "A Survey and Comparative Analysis of Data Mining Techniques for Network Intrusion Detection Systems", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012.
18. Younes Chihab, Abdelah Ait Ouhman, Mohammed Erritali, Bouabid El Ouahidi , "Detection & Classification of Internet Intrusion Based on the Combination of Random Forest and Naïve Bayes", Younes Chihab et.al International Journal of Engineering and Technology (IJET), ISSN : 0975-4024 Vol 5 No 3 Jun-Jul 2013.
19. Jen-Yan Huang, I-En Liao, Yu-Fang Chung, , Kuen-Tzung Chen Shielding wireless sensor network using Markovian intrusion detection system with attack pattern mining, Wireless sensor network 29 March 2011.
20. G.V. Nadiammai, Hemalatha, Effective approach toward Intrusion Detection System using data mining techniques Arabian Journal of science and technology 03 March 2013.
21. Boz, O. Feature Subset Selection by Feature Relevance. Submitted for the ICML 2002.
22. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, P.J. (1984). Classification and Regreession Trees

Wadsworth International Group. Belmont, CA.

23. Cardie, C. Using Decision Trees to Improve Case-based Learning. ICML 1993, pp. 25-32.
24. Shabana Asmi, P.,Justin Samuel, S. (2015), "An analysis and accuracy prediction of heart disease with association rule and other data mining techniques ",Journal of Theoretical and Applied Information Technology, Vol.79. No.2, pp. 254-260
25. M. Sudha, A. Kumaravel, "Performance Comparison based on Attribute Selection Tools for Data Mining",Indian Journal of Science and Technology,2014 Nov, 7(S7), Doi no: 10.17485/ijst/2014/v7iS7/60459.