Research Journal of Applied Sciences, Engineering and Technology 9(6): 454-459, 2015
DOI:10.19026/rjaset.9.1425
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: October 10, 2014 Accepted: November 26, 2014 Published: February 25, 2015

Research Article

Detection of Attacks on MAODV Association Rule Mining Optimization

[1]A. Fidalcastro and [2]E. Baburaj
[1]Department of CSE, Sathyabama University, Chennai, Tamilnadu, India
[2]Department of CSE, Sun College of Engineering and Technology, Nagercoil, India

Abstract: Current mining algorithms can generate large number of rules and very slow to generate rules or generate few results, omitting interesting and valuable information. To address this problem, we propose an algorithm Optimized Featured Top Association Rules (OFTAR) algorithm, where every attack have many features and some of the features are more important. The Features are selected by genetic algorithm and processed by the OFTAR algorithm to find the optimized rules. The algorithm utilizes Genetic Algorithm feature selection approach to find optimized features. OFTAR incorporate association rules with several rule optimization techniques and expansion techniques to improve efficiency. Increasing popularity of Mobile ad hoc network users of wireless networks lead to threats and attacks on MANET, due to its features. The main challenge in designing a MANET is protecting from various attacks in the network. Intrusion Detection System is required to monitor the network and to detect the malicious node in the network in multi casting mobility environment. The node features are processed in Association Analysis to generate rules, the generated rules are applied to nodes to detect the attacks. Experimental results show that the algorithm has higher scalability and good performance that is an advantageous to several association rule mining algorithms when the rule generation is controlled and optimized to detect the attacks.

Keywords: Association analysis, black hole attack, data mining, genetic algorithm, intrusion detection system, MANET, spoofing attack, top rules

INTRODUCTION

A Mobile Ad-hoc Network is an infrastructure less network of mobile devices connected by wireless nodes. The mobile nodes in MANET are frequently moving independently in 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 immediate neighbors and then pass the information throughout the network. This way, routers gains knowledge of the topology of the network (http://en.wikipedia.org/wiki/Routing protocol). Multi cast Ad hoc on Demand distance Vector (MAODV) is a multi cast expansion of AODV protocol and it provides tree based functionality to route data among the nodes in MANET. The route discovery is based on Route Request and Route Reply process. The tree based routing protocol has the following characteristics: loop free, periodic messaging and controls the packet flooding. The applications of MAODV are military operations, rescue operations, etc.

The most important feature in MANET is mobility, describes the movement of nodes, location and other related features to the network. In the simulation scenario the mobility defines the module to propagate data packets in the network. The various scenario generation models are Random Way point Model, Random Walk Model, Random Direction Model, etc. Due to mobility of the node, it is easily poised to malicious activities and it leads to degradation of performance. 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.

Drawbacks of IDS: Current IDS are usually tuned to detect known service level network attacks. This leaves them vulnerable to novel malicious attacks, Data overload, false positives, false negatives. Data mining is used to improve intrusion detection efficiency.

A formal statement of the association rule problem is Agrawal *et al*. (1993) and Cheung *et al*. (1996).

Definition 1: Let I = {I_1, I_2, …, I_m} be a set of m distinct attributes, also called literals. Let D be a database, where each record (tuple) T has a unique identifier and contains a set of items such that TÍI An association rule is an implication of the form XÞY,

Corresponding Author: A. Fidalcastro, Department of CSE, Sathyabama University, Chennai, Tamilnadu, India

where X, YÌI, are sets of items called item sets and XY = f. Here, X is called antecedent and Y consequent.

Definition 2: The support (*s*) of an association rule is the ratio (in percent) of the records that contain XY to the total number of records in the database.

Definition 3: For a given number of records, confidence (*c*) is the ratio (in percent) of the number of records that contain XY to the number of records that contain X.

An important task in association rule mining is how to choose the thresholds to generate a desired amount of rules, because in user has limited resources in the terms of time and space to analyze the results. To analyze the results user need the certain amount of rules from the large sized rule set. The rules are to be optimized from the set of features which are relevant to the problem. Then only user can find the minimal set of rules to find the results to reduce time and space. The rules which are generated and related to the problem to be selected by selecting relevant features and avoiding irrelevant rules.

In this study, our contribution is to propose a novel association rule mining algorithm for the standard definition (with multiple items related with network domain, from a transaction database/flat files). The algorithm is to mine the Top rules with the highest support that meet a desired confidence. Features selection is by genetic algorithm differ for every attack. So the features optimized and given to OFTAR algorithm to optimize the rules in every phase of an algorithm. Mining the Top association rules is challenging because Top association rule mining algorithm cannot rely on both thresholds to prune the search space and first step process to find all rules and second to find the optimized and featured rule selection keeping the Top rules by user defined and desired feature by GA. Here investigate the selection of relevant features by using GA in order to model these complex systems.

We address the problem of association rule mining by proposing an algorithm named OFTAR. Moreover, the outcome of OFTAR is an advanced than the classical association rule mining algorithms for users to optimize the rules generated and control the generated association rules.

LITERATURE REVIEW

Karmore and Nirkhi (2011) proposed a one of the methodology of Data mining approach, which is Clustering based Data Mining, which is employed in MANET to detect intrusions with the help of Intrusion Detection System using k-means clustering. This Data Mining approach helps in improving the performance of network by evaluating the detection rate, control false alarm rate and false dismissals.

Peyam and Mehran (2011) reported the node features are analyzed for intrusion detection in MANET. Here using Principal Component Analysis technique and Profile based monitoring techniques are used and implemented in the network environment to measure the performance of network. The performance is varied based on the routing protocol used in the MANET.

Bahaya and AlAsady (2012) proposed an algorithm to prevent spoofing attacks by additional authentication process. The first authentication is processed using MAC address filtering technique followed by periodic re-authentication process, which authenticates the client periodically after a set of data frames.

Chen *et al*. (2007) implemented a method to detect and locating the malicious attackers in the network. The author used a data mining technique to locate the malicious attackers using clustering technique. The clustering approach helps to group the malicious nodes by area as well as point based by localization algorithms.

Hassan *et al*. (2006) implemented an Intrusion Detection System (IDS) to target the security of AODV protocol designed for MANET. Here the IDS are designed as a Multiple Static Agent to monitor the route establishment process 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 network monitoring nodes using the clustered networking monitoring selection algorithm (Karchirski and Guha, 2008) and the second step is the NM nodes will run the monitoring protocol its responsibility is to observe the flow of data.

Hu *et al*. (2003) defines a new type of denial of service attack called rushing attack operated on all On-Demand Ad-hoc routing protocols such as AODV, DSR, etc., are unable to find routes longer than two hops due to the presence of rushing attack.

Sadia (2014) defines a hybrid data mining approach for Intrusion Detection System includes feature selection, filtering, clustering, divide and merge and clustering coordinates. The theme of this work is to improve the detection rate and decrease the false alarm rate. The intrusion detection is processed based on clustering analysis technique.

Vikas and Shirish (2012) proposed a novel approach for handling Intrusion Detection System (IDS) alerts based on the features of association rule mining to IDS. Here integrated a fuzzy association rules to detect the misbehavior activity in the network using Intrusion Detection System.

Fournier-Viger *et al*. (2012a) describes the use of Association Analysis to generate rules and design an algorithm to determine the minimum support and minimum confidence. The rules are redundantly occurring in the list. It leads to generate the huge list of rules.

Fournier-Viger and Tseng (2012b) describes the non-redundant association rules such a generic basis, informative basis, the informative and generic basis, the minimal generic basis and minimum condition maximum consequent rules. These rule sets can be measured on multiple criteria such as theirs, compactness, the possibility of recovering redundant rules with support and confidence.

MANET ATTACKS IN MULTICAST ENVIRONMEMTS AND ASSOCIATION RULE MINING DEFINITIONS-MATERIALS AND METHODS

A Host based IDS are run on individual devices or hosts on the network. A Host IDS monitors the incoming and outgoing packets from the host only and it will alert the user of suspicious activity is detected. We are concentrating the below attacks.

Black hole attack: This attack is also called Packet drop attack; it advertises itself by having shortest path to particular node, the neighboring nodes to attract all the routing packets to them and it simply drops the packet without transmitting to the destination node.

Spoofing attack: This type of attack is attracted among hackers and widely used to exploit network vulnerabilities by gaining illegitimate advantage of other users by falsifying the data. It leads to harmful attack in future.

For routing the data across the network using Multicast Ad-hoc On-demand Distance Vector (MAODV) routing protocol is used, it is an extension of Ad-hoc On-demand Distance Vector routing protocol, its additional feature is broadcast the data across the network. The Intrusion Detection System is implemented in every host in the network, it helps to track the attackers based on the behavior of the node. Here the IDS is processed by an Data Mining technique called Association Analysis. The rules are applied on every host actively take part in routing process, if the feature varies from the generated rules it informs to the IDS, to avoid or discard the malicious node for further data communication (Pang-Ning *et al*., 2006). Here Association Analysis is processed in network model, by means of generating top association rules to process the Intrusion Detection System to check the node behavior in the network. For networking purpose it finds association among features of nodes. The strength of the associations rule can be measured in terms of support and confidence. Support determines how often a rule is applicable to a given data set, while confidence determines how frequently items in Y appear in transaction that contains X (Pang-Ning *et al*., 2006).

An important problem that has not been addressed is how the user should choose and thresholds to generate a desired amount of rules with optimized features in wireless multi cast environments.

OFTAR and feature selection through genetic algorithm: OFTAR algorithm is one of the association rule mining technique to discover association among features of nodes. In generating rules, the user has to define the number of optimized rules generated and rules relevant to the problem. Thus the rules are optimized set of features and user defined number of rules.

The generated rules are applied to Intrusion Detection System; it applies to every node in network, if the rule with any node and it informs the IDS to alert the normal nodes about intruder present in the network. The malicious node is discarded from the network for data transmission.

OFTAR mining algorithm: The algorithm first scans the IDS log to calculate transaction id for each node in the network. Then the algorithm generates all valid rules of size 1*1 by considering features of nodes, where the features must have the minimum support value (Fournier-Viger *et al*., 2012a).

The support of the rules $\{a\}\rightarrow\{b\}$ and $\{b\}\rightarrow\{a\}$ are simply obtained by dividing:

$$|\text{Tids}(a\rightarrow b)| / |T| \quad (1)$$

$$|\text{Tids}(b\rightarrow a)| / |T| \quad (2)$$

The confidence of the rules $\{a\}\rightarrow\{b\}$ and $\{b\}\rightarrow\{a\}$ is obtained by:

$$|\text{Tids}(a\rightarrow b)| / |\text{Tids}(a)| \quad (3)$$

$$|\text{Tids}(b\rightarrow a)| / |\text{Tids}(b)| \quad (4)$$

OFTARULESET (T, k, minconf, gafeatureset) R: = Ø. L: = Ø. Minsup: = 0:

- Scan the Database T once to record the log of each node
- FOR each pairs of items a, b such that $|\text{Tids}(a) \times |T| \geq$ minsup and $|\text{Tids}(b) \times |T| \geq$ minsup
- sup ($\{a\}\rightarrow\{b\}$): = $|\text{Tids}(a) \cap \text{Tids}(b)|/|T|$
- sup ($\{b\}\rightarrow\{a\}$): = $|\text{Tids}(a) \cap \text{Tids}(b)|/|T|$
- conf ($\{a\}\rightarrow\{b\}$): = $|\text{Tids}(a) \cap \text{Tids}(b)|/|\text{Tids}(a)|$
- conf ($\{b\}\rightarrow\{a\}$): = $|\text{Tids}(a) \cap \text{Tids}(b)|/|\text{Tids}(b)|$
- IF sup ($\{a\}\rightarrow\{b\}$) ≥minsup THEN
- IF conf ($\{a\}\rightarrow\{b\}$) ≥ minconf THEN
  a, b→gafeatureset THEN SAVE ($\{a\}\rightarrow\{b\}$, L, k, minsup)
- IF conf ($\{b\}\rightarrow\{a\}$ ≥minconf THEN
  a, b→ gafeatureset THEN SAVE ($\{b\}\rightarrow\{a\}$, L, k, minsup)
- Set flag expandLR of ($\{a\}\rightarrow\{b\}$) to true and CHECK gafeature

- Set flag expandLR of ({b}→{a}) to true and CHECK gafeature
- R: = RU {{a}→{b}}, {b}→{a}}
- END IF
- END FOR
- WHILE ∃r ϵ R AND sup (r) ≥ minsup DO
- Select the rule having the highest support in R
- IF rule.expandLR = true THEN
- EXPANSION-L (rule, L, R, k, minsup, minconf)
- EXPANSION-R (rule, L, R, k, minsup, minconf)
- ELSE EXPANSION-R (rule, L, R, k, minsup, minconf)
- REMOVE rule from R
- REMOVE from R all rules r ϵ R | sup (r) < minsup
- REMOVE from R all rules not belongs to gafeatureset

The SAVE procedure: The SAVE procedure is to raise minsup and update the list L when a new valid rule r is found with optimized features. The first step is to add the rule r to list. Then if list contains more than k rules and the support is higher than minsup, rules from list that exactly the support equal to minsup can be removed until only k rules are kept and choose the rules with optimized features. Finally the support of the rule in L having the lowest support and rule features set check with the GAFEATURESET. By this way, the OFTAR rules are found and maintained in list.

The expansion-R procedure: The EXPANSION-R is implemented based on the strategies (Fournier-Viger *et al*., 2012a):

- Determining items that can be expand the rule A→B to produce a valid rule and check A, B are in GAFEATURESET.
- Assessing if a frequent rule obtained by an expansion is valid and with optimized features to check with GAFEATURESET.

The expansion-L procedure: The EXPANSION-L is very similar to EXPANSION-R. The only extra step is performed compared to EXPANSION-R is that for each rule I U {c}→ J obtained by the expansion of I→J with an item c, the value Tids (I U {c}) necessary for calculating confidence is obtained by intersecting Tids (I) with Tids (c) and verify that each rule feature belongs to GAFEATURESET.

Genetic algorithm for feature selection: Genetic Algorithms is an evolutionary computation technique inspired by biological evolution. The algorithm starts with creating individuals (generally randomly) which are the candidates solutions for the problem. Traditionally, binary as strings of 0s and 1s are represented as individuals. Each individual is assigned a fitness value by fitness function which shows how the individual solves or comes close to the solution. The genetic operators (selection, crossover, mutation, reproduction and etc.) are applied on individuals based on their fitness values until the termination criteria is satisfied. The aim is to provide better individuals in the new population.

First, random individuals are created for the particular solution. These individuals represent which features are used for the OFTAR algorithm and which not. OFTAR is run for each individual and a fitness value is assigned to each individual based on the formula shown. The GA algorithm continues until a defined generation is reached:

$$\text{Fitness} = \text{Detection rate-false positive rate} \quad (5)$$

Here we used JAGA toolkit for the GA implementation for our experiments. The GA parameters are selected as follows: 90 for population size, 90 for generation size, 0.8 for crossover probability and 0.1 for reproduction probability. Other parameters used are the default parameters of the JAGA toolkit. At each generation 100 individual evaluated by creating OFTAR algorithm for each individual. Since our training dataset is huge, fitness values might not be obtained in a reasonable time.

GA algorithm is run ten times and the feature set with the highest fitness value is selected. OFTAR algorithm is run with this feature set (GAFEATURESET) in shown below. Both an increase in detection rate and a decrease in false positive rate is seen in the results. The features selected here is as follows:

- Number of neighbours
- Number of added neighbours
- Number of updated routes (modifying hop count, sequence number)
- Number of added routes under repair
- Number of broadcasted route request packets from this node
- Number of forwarded route request packets from this node
- Number of received broadcast route error packets (to be forwarded or not)
- Number of broadcasted route error packets from this node

The Attacks are detected by the IDS, after processed the OFTAR rules. The attacks found in our proposed work are: Blackhole Attack and Spoofing Attack

EXPERIMENTAL RESULTS AND DISCUSSION

Simulation setup: 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 (Yufang and Thomas, 2004).

The Simulation Environment is defined by the user based on the number of nodes:

- **Area:** 500×500 m
- **Number of nodes:** 50
- **Simulation duration:** 800 sec
- **Number of repetitions:** 7
- **Physical/MAC layer:** IEEE 802.11 at 2 Mbps, 250 m transmission range
- **Mobility model:** random way point with no pause time and node movement speed 0, 1 or 20 m/sec
- Each sender sends 2 multicast data packet per second with each packet 256 bytes long
- Each receiver is a multicast group member, but each sender is not a group member
- All receivers join multicast group at the beginning of the simulation
- 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.

The NS2 outcome is input to the GA and OFTAR algorithm to find features and optimized rules to detect the above attacks. The above two techniques developed using Java.

Result evaluation: In this section, the implementation of MAODV protocol is performed, node movement generation and traffic pattern file is generated for node communication in the network and the performance compared with and without OFTAR. By using OFTAR after the performance improved significantly.

Figure 1 shows the performance of the network is measured against intrusion in the MANET, the red line defines presence of attack in the network and the green, blue lines depicts the OFTAR algorithm detects the blackhole, spoofing attack with improved the performance.

Figure 2 represents Number of Nodes vs. Delay and it shows the delay is high at the initial stage of data transmission in the Network. The Intrusion Detection System detects the attacks using OFTAR and the delay reduces gradually till the completion of processing time of MANET for both the attacks.

Figure 3 defines the packet loss in the network during the presence of attack. At initial stage the packet loss is high due to intrusion detection system is not present and data communication is high between nodes. As the process goes on the Intrusion detection system identifies the packet loss and detect the nodes causing

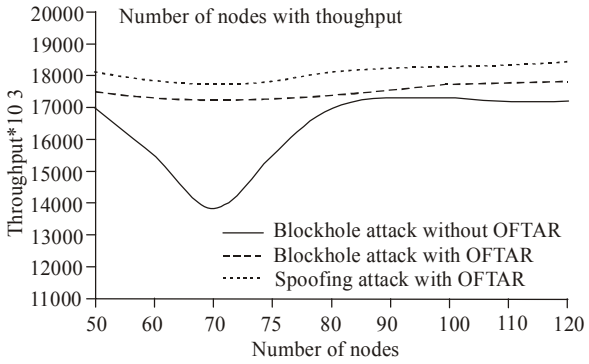

Fig. 1: Throughput graph

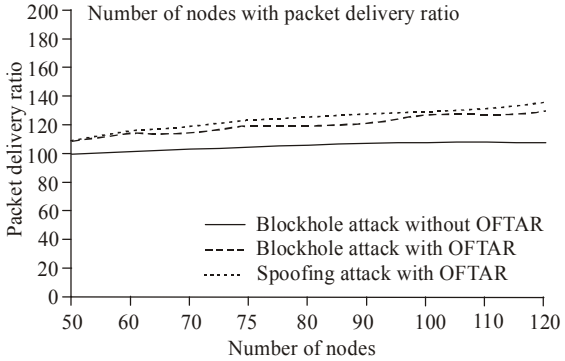

Fig. 2: Delay graph

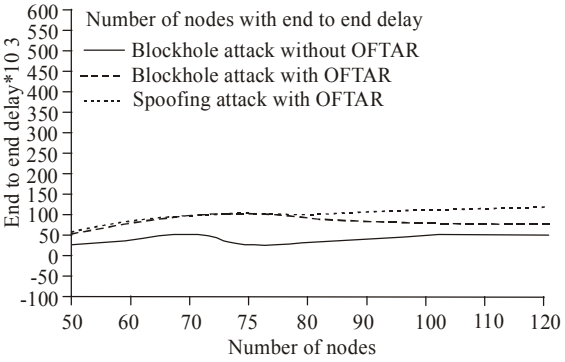

Fig. 3: Packet loss graph

packet dropping by OFTAR. It leads to prevent the nodes for routing.

CONCLUSION

The detection of attacks MAODV based Mobile Ad-hoc networks with the use of extended data mining technique OFTAR is a new kind of approach in wireless networks, the data mining approach helps in detecting attack by generating optimized association rules based on the selection of relevant features of the nodes by GA. The performance of mining algorithm has increased with the reduced feature set obtained by GA; both an increase in detection rate and a decrease in

false positive rate are observed. The association analysis is efficient when comparing with other data mining techniques. It gives better performance, delay, packet delivery ratio for MANET while spoofing and Black hole attacks. It is advantageous over other classical association rule mining algorithms in detecting attacks.

REFERENCES

Agrawal, R., T. Imielminski and A. Swami, 1993. Mining association rules between sets of items in large databases. Proceeding of the ACM International Conference on Management of Data, ACM Press, pp: 207-216.

Bahaya, W.S. and S.A. AlAsady, 2012. Prevention of spoofing attacks in the infrastructureless wireless networks. J. Comput. Sci., 8(10): 1769-1779.

Chen, Y., W. Trappe and R.P. Martin, 2007. Detecting and localizing wireless spoofing attacks. Proceeding of the 40th Annual IEEE Communication Society Conference on Sensor, Mesh and Ad Hoc Communication and Networks (SECON, 2007), pp: 193-202.

Cheung, D.W.L., J. Han, V. Ng and C.Y. Wong, 1996. Maintenance of discovered association rules in large databases: An incremental updating technique. In ICDE, pp: 106-114.

Fournier-Viger, P. and V.S. Tseng, 2012b. Mining top-K non-redundant association rules. In: Chen, L. *et al*. (Ed.), ISMIS, 2012. LNAI 7661, Springer-Verlag, Berlin, Heidelberg, pp: 31-40.

Fournier-Viger, P., C.W. Wu and V.S. Tseng, 2012a. Mining top-K association rules. In: Kosseim, L. and D. Inkpen (Eds.), Canadian AI, 2012. LNAI 7310, Springer-Verlag, Berlin, Heidelberg, pp: 61-73.

Hassan, H.M., M. Mahmoud and S. El-Kassas, 2006. Securing the AODV protocol using specification-based intrusion detection. Proceeding of the 2nd ACM International Workshop on Quality of Service and Security for Wireless and Mobile Networks (Q2SWinet'06), pp: 33-36.

Hu, Y.C., A. Perrig and D.B. Johnson, 2003. Rushing attacks and defense in wireless ad-hoc network routing protocols. Proceeding of the 2nd ACM Workshop on Wireless Security. USA, pp: 30-40.

Karchirski, O. and R. Guha, 2008. Effective intrusion detection using multiple sensors in wireless ad hoc networks. Proceeding of the 36th Hawaii International Conference on System Sciences (HICSS'03), pp: 57.

Karmore, P.K. and S.M. Nirkhi, 2011. Detecting intrusion on AODV based mobile ad hoc networks by k-means clustering method of data mining. Int. J. Comp. Sci. Inform. Technol., 2(4): 1774-1779.

Pang-Ning, T., S. Michael and K. Vipin, 2006. Introduction to Data Mining. 2nd Edn., Pearson Education, Boston, Munich.

Peyam, K. and A. Mehran, 2011. Feature analysis for intrusion detection in mobile ad-hoc networks. Int. J. Netw. Secur., 12(1): 42-49.

Sadia, P., 2014. Intrusion detection model based on data mining technique. IOSR J. Comput. Sci., 2014: 34-39.

Vikas, M. and M.D. Shirish, 2012. General study of association rule mining in intrusion detection system. Int. J. Emerg. Technol. Adv. Eng., 2(1): 347-356.

Yufang, Z. and K. Thomas, 2004. MAODV implementation for NS-2.26. System and Computing Engineering, Carleton University, Technical Report SCE-04-01.