A COLLABORATIVE INTRUSION DETECTION SYSTEM FOR MANET USING DATA MINING TECHNIQUE

S. B. Ninu[1] and S. Behin Sam[2]
[1]Department of Computer Science, Bharathiar University, Coimbatore, India
[2]Department of Computer Science, Government Arts and Science College, Perumbakkam, Chennai, India
E-Mail: behinsam@gmail.com

ABSTRACT

Mobile Ad hoc Networks (MANETs) are vulnerable to various kinds of threats due to their dynamic nature and lack of a central point of control. Intrusion Detection System (IDS) which can act in collaboration with other IDS nodes in the nertwork is getting popularity due to its faster adaptability to the changes in the behavior of network traffic. A standalone node in MANET will feel very difficult to set any predefined rule for identifying correctly attack traffic since there is no major difference between normal and attack traffic. Hence, in this paper we have proposed an intelligent collaborative model based on data mining for intrusion traffic detection system that can detect attacks. Here we find and deploy friendly nodes in the network that continuously monitors the behavior of other nodes to find nodes or set of nodes exhibiting anomalous behavior. NS-2 simulations were carried out to anlyze the performance of the proposed system. We evaluated the performance of our proposed collaborative IDS scheme with various other existing IDS models. The results clearly showed that the proposed intrusion detection system considerably reduces the false positive rate, thereby proving that the proposed technique is capable of identifying anomalies in network better than other existing system.

Keyword: MANET, anomaly detection, data mining, IDS, collaborative IDS.

INTRODUCTION

With the rapid development of technology, wireless communication networks have appeared in many forms. Mobile Ad hoc Networks (MANETs) have self-configuration and self-maintenance capabilities (Nekovee et. al., 2010). In MANETs, each node works as a router and can communicate with other nodes directly or indirectly with the help of its neighbors. MANETs can be deployed in disaster areas to collect critical information, in battlefields to allow for communication among soldiers, and in hazardous areas in the form of sensor networks. Due to the lack of a central point of control, it is more likely that malicious nodes can join the network and launch various types of attacks (Muhammad Imran *et al,* 2014, Hamed Janzadeh *et al*., 2009, Humaira Ehsan *et al*., 2012). An attack can be launched by a single node or multiple nodes in a cooperative manner. The attacker node can be external (node outside the network) or internal (compromised node inside the network), with the internal attackers being the more dangerous and difficult to detect of the two. In some attacks, multiple attackers synchronize their actions to disrupt a target network.

According to Ghosh *et al*., 1998, Vigna and Kemmerer, 1998, intrusion detection systems have been widely implemented in many networks aiming to defend against a variety of attacks. IDS have already become an essential component for current defense infrastructure (Scarfone and Mell, 2007). However, network intrusions have become much more sophisticated and hard to detect (Vasilomanolakis *et al*., 2015). Since conventional IDSs are not scalable to resolve this issue, IDS collaboration is considered as an effective way to enhance the detection capability of a single ID. They consist of several monitoring components that collect and exchange data. Depending on the specific collaborative intrusion detection system (CIDS) architecture, central or distributed analysis components mine the gathered data to identify attacks. Resulting alerts are correlated among multiple monitors in order to create a holistic view of the network monitored.

Motivated by this, collaborative intrusion detection systems have been developed, with the purpose of strengthening a single IDS by collecting knowledge and learning experience from other IDS nodes. According to (Wu *et al*., 2003), collaborative IDS is expected to enhance the overall detection accuracy of intrusion assessment and will also improve the possibility of identifying novel attacks. Hence, the main objective of this paper is to design a robust collaborative intrusion detection system that can effectively evaluate the trustworthiness of each node within the network and identify the intrusions in the network.

The collaborative IDS pproposed in this paper uses data mining techniques for detecting attacks. Here we select and deploy friendly IDS nodes based on their trust, that continuously monitor the behavior of other nodes in the network for any intrusions. The proposed IDS was evaluated using NS-2 simulations which showed that the proposed system considerably reduced the false positive rate compared to other existing IDS, thereby proving that it is better than other existing system.

RELATED WORK

Shakshuki *et al*., 2013 proposed an IDS named Enhanced Adaptive Acknowledgment (EAACK) for MANETs. Their scheme requires all acknowledgment packets to be digitally signed by its sender and verified by its receiver. They used DSA and RSA as digital signatures and showed that their scheme is able to detect wide range of attacks. However, the drawback of their scheme is the requirement to digitally sign all the acknowledgments which increases computational overhead.

Marti *et al.*, 2000 proposed an IDS scheme for MANET which consists of two different modules, viz. the Watchdog and the Pathrater. In this scheme, the Watchdog acts as an IDS for the MANET and detects malicious node behaviors in the network by promiscuously listening to its next hop's transmission. If the Watchdog notices that its immediate next node fails to forward the packet within a given period of time then it increments the node's failure counter. If the failure counter of the monitored node exceeds a threshold value then the Watchdog reports the node as misbehaving. The Pathrater is then employed to inform the routing protocol to avoid the reported nodes for further data transmission. The drawback of this scheme is that it requires continuous monitoring by the Watchdog for detecting intrusions.

Liu *et al.*, 2007 proposed a TWOACK MANET IDS scheme which requires every data packets transmitted over three consecutive nodes along the source to the destination path to be acknowledged. Every node along the route has to send back an acknowledgment packet to the node that is two hop counts away from it in the route. The arrival of TWOACK packet at first node X (in the three consecutive nodes along the route) indicates a successful transmission of packet from node X to node Z via the intermediate node Y. However, if this TWOACK packet is not received within a given predefined time interval, both nodes Y and Z are reported as malicious. The drawback of this scheme is that it introduces a routing overhead due to frequent TWOACK packet generation.

Misra *et al.*, 2010 proposed a distributed self learning, energy aware and low complexity protocol for intrusion detection in wireless sensor network. Their protocol uses the stochastic Learning Automata (LA) on packet sampling mechanism to obtain an energy efficient IDS. They showed that their approach was successful in detecting and removing malicious packets from the WSN. The drawback of this scheme is that the LA needs multiple rounds of learning before it becomes efficient.

Haddadi and Sarram, 2010 proposed a hybrid IDS model for Wireless Local Area Network (WLAN) that uses both misuse and anomaly based IDS sub-modules to detect intrusion. The drawback of this approach is that the response times of the misuse based and anomaly based IDSs are different. It also introduces significant computational overhead due to processing of the same data traffic by two different IDSs.

A light weight, energy efficient and non-cryptographic intrusion detection solution against the gray hole attack in MANET is proposed by Mohanapriya and Krishnamurthi, 2014. However, their scheme requires the IDS to operate in a promiscuous mode to detect intrusions, which results in high power consumption for operating the IDS.

Liu *et al.*, 2006 proposed a game theoretic framework to analyze the interactions between pairs of attacking/defending nodes using a Bayesian formulation in wireless Ad-hoc Networks. They suggested a Bayesian hybrid detection approach for the defender, in which a less powerful lightweight module is used to estimate the opponent's type, and a more powerful heavyweight module acts as a last line of defense. They analyzed the obtainable Nash Equilibrium (NE) for the attacker/defender Bayesian game in both static and dynamic settings and concluded that the dynamic approach is a more realistic model, since it allows the defender to consistently update its belief about the maliciousness of the opponent player as the game evolves. The drawback of their work is that it is difficult to determine a reasonable prior probability about the maliciousness of the attacker player.

Liu, 2003 proposed a general incentive-based method to model attacker's intent, objectives and strategies (AIOS) based on game theoretic formalization. The author developed an incentive-based conceptual framework for AIOS modeling which can capture the inherent inter-dependency between AIOS and defender objectives and strategies in such a way that AIOS can be automatically inferred. The AIOS modeling enables the defender to predict which kind of strategies are more likely to be taken by the attacker than the others, even before such an attack happens. The AIOS inferences lead to more precise risk assessment and harm prediction. The drawback of the scheme is that it assumes the complete information game.

Chen *et al.*, 2001 proposed a framework that applies two game theoretic schemes for economic deployment of intrusion detection agent. In the first scheme, the interaction between an attacker and the intrusion detection agent is modeled and analyzed within a noncooperative game theory setting. The mixed strategy Nash Equilibrium solution is then used to derive the security risk value. The second scheme uses the security risk value derived by the first scheme to compute the Shapley value of the intrusion detection agent while considering the various threat levels. This allows the network administrator to quantitatively evaluate the security risk of each IDS agent and easily select the most critical and effective IDS agent deployment to meet the various threat levels to the network. The drawback of this scheme is the computational overhead involved for calculating the Shapley values of the intrusion detection agents.

PROPOSED METHODOLOGY

We propose a framework for detecting intrusion in MANET with the help of other specialized nodes in the network. The proposed IDS work in collaboration with other designated IDS nodes in the network. Mostly the friendly nodes are designated as IDS nodes. All the IDS nodes continuously monitor the behaviour of other nodes in the network.

Each IDS node maintains a list called *helpers* that have details of other friendly IDS nodes with which it currently collaborates with. Each IDS node in the network has the freedom to choose its *helpers* based on their own interest. The communications between collaborating IDS nodes are requests for intrusion alert evaluation and their corresponding feedbacks. An IDS node initiate a cooperative intrusion detection procedure under any of the two conditions. The first condition is that the IDS node cannot detect locally an intrusion or anomaly and the second condition is that the evidence available locally is inconclusive and warrants broader investigation. The overall procedure works by propagating the intrusion detection state information among neighbouring IDS nodes.

The helpers are found by their trustworthiness. To evaluate the trustworthiness of a friendly IDS node, an IDS node can send a challenge to this target periodically using a random generation process (i.e., sending time is not fixed, but random). When receiving the feedback from the target node, the IDS node can give a score to reflect its satisfaction level. Since we define two types of trust including feedback-based trust (T_{fd}) and packet-based trust (T_{pt}), we develop a single metric called overall trust (T_{total}) to facilitate the trust evaluation as follows:

$$T_{total}= W_1 \times T_{fd}+ W_2 \times T_{pt} \qquad (1)$$

where W_1 and W_2 are weight values and W_1 + W_2 = 1. For the feedback based trust $T_{fd}^{i,j}$of node i according to node j, we can compute it by using the equation described as below:

$$T_{fd}^{i,j} = w_s \frac{\sum_{k=0}^{n} F_k^j \lambda^{tk}}{\sum_{k=0}^{n} \lambda^{tk}} \qquad (2)$$

where $F_k^j \epsilon[0.1]$is the score of the received feedback k and n is the total number of feedbacks. λ is a forgetting factor that assigns less weight to older feedback responses and emphasizes the impact of the latest responses. w_s is a significant weight that depends on the total number of received feedback, if there is only a few feedback under a certain minimum m, then $w_s = \frac{\sum_{k=0}^{n} \lambda^{tk}}{m}$otherwise w_s= 1.

Additionally, in this work, we integrate a type of packet-based trust, in which the trust of node i according to node j can be computed based on our previous research as below:

$$T_{pt}^{i,j} = \frac{k+1}{N+2} \qquad (3)$$

where k is the number of received benign packets and N is the total number of received packets.

To evaluate the trustworthiness of a node j, we can use a weighted majority method as follows:

$$T_j = \frac{\sum_{T\geq r} T_{total}^{i,j} D_i^j I_s^i}{\sum_{T\geq r} T_{total}^{i,j} D_i^i} \qquad (4)$$

where r is a threshold in which node j requests alert ranking to those nodes whose trust values are higher than this threshold. $T_{total}^{i,j}\epsilon[0.1]$is the overall trust value of node i according to node j. $D_i^i\epsilon[0.1]$is a measure of hops between these two nodes and is anti proportional to the hops between the nodes in the number of grid steps. $I_s^i \epsilon[0.1]$is the intrusion sensitivity of node i. Each IDS node can consult other nodes and can thus evaluate the intrusion sensitivity of nodes accordingly.

Once the *helpers* are found the IDS nodes uses a system that contains an ensemble of four methods for detecting intrusion. These methods work as follows:

a) Support vector machine (SVM)

SVM (Vapnik, 2000) is a supervised machine learning method for classification and regression analysis. The principle of SVM is to find the best hyperplane to separate the data into two parts. A SVM model consists of the samples represented as points in space. The samples of the different categories are divided by hyperplane. This hyperplane always maximizes the margin between those two regions or classes. The margin is defined by the farthest distance between the samples of the two classes and computed based on the distances between the closest samples of both classes, which are called supporting vectors. Test samples are then mapped into the same space. Based on which sides of the hyperplane they fall on, test samples are predicted to belong to the corresponding categories.

b) Classification and regression tree (CART)

CART is a method based on the Gini index. It usually uses a topdown approach when CART constructs a decision tree. Decision tree (Qinlan, 1986) is a categorization model that recursively partitions the training data into a tree structure. In the experiments, we first put all the training samples at the root node. We then search the best partition of the root node so that the Gini impurity can reduce to minimum. Gini impurity represents the possibility that a randomly selected sample is classified into the wrong subset. When all the samples of a node belong to one class, Gini impurity equals to zero. We use the best partition to divide root node into two parts, each of which is seen as a new node. This process is then repeated on the new nodes.

c) Naive Bayes (NB)

Naive Bayes is a probabilistic classifier based on Bayes theorem (Bayes, 1763). Given a test sample, we need to calculate the probabilities of the appearance of various categories under the condition of the appearance of a test sample. The sample belongs to the category whose probability is the largest.

d) K-Nearest Neighbor (K-NN)

K-NN (Fix, 1952) algorithm is a non-parametric statistical methods for categorization and regression. It classifies a test sample by measuring the distance between the training samples and test sample. We need to choose k nearest samples and use majority voting to predict which category the sample belongs to.

To exert the advantage of each algorithm and to further improve the accuracy of detection and categorization, we employ the ensemble of multiple classifiers with majority voting after obtaining the classification results of the five algorithms described above. When a test app is given as input, each base classifier predicts its classification. All the five prediction results will then vote to generate a final prediction.

RESULTS AND DISCUSSIONS

We have implemented our proposed model in the network simulator NS2 installed on ubuntu 12.04 running gcc version 4.6.3. We restrict the movements of mobile nodes to a predefined flat grid area. Table-1 lists the various parameters used for our simulation.

Table-1. Parameters used for NS2 simulation.

PARAMETERS	VALUE
Simulation	8000-15000s
Number of Nodes	12-30
Simulation Area	600 x 600 m2
Transmission Range	150m
Mobility	Random Way Point
Routing Protocol	DSR
MAC Layer	DCF of IEEE 802.11
Max. Node Movement Speed	20 m/s
Pause Time	500s
Traffic Type	CBR/UDP
Data Rate	20kbps
Packet Size	512 Bytes
MAC Protocol	IEEE 802.11
Sampling Interval	3s

TCL scripts were used to generate which contain data related to normal profile and simulated attacks.The sampling rate of 3 s is used to record the values. The rules extracted from these traces are then used to build the normal and abnormal behaviour of the network. In order to compare efficiency of proposed algorithm, the test traces were used to compute the classification accuracy using the following parameters:

- Accuracy
- True Positive Rate (TPR)
- False Positive Rate (FPR)
- Precision
- Recall
- F-meausre

The detection results of the proposed IDS using each classifier are shown in Table-2. It is seen that the TPR of ensemble are the highest among the five methods, achieving 98.25%. This method also achieves the accuracy of 99.39% in the detection of intrusion after employing ensemble of the four classifiers with majority voting mechanism. In general, our method outperforms SVM, CART, NB and K-NN.

Table-2. The detection results of proposed IDS model with four base classifiers as well as with ensemble of classifiers.

Classifier	**Accuracy** (%)	**TPR** (%)	**FPR** (%)	**Precision** (%)	**Recall** (%)	**F-measure** (%)
SVM	98.82	96.07	3.39	92.79	95.09	93.93
CART	99.23	95.83	0.52	93.31	95.83	94.55
NB	76.46	90.92	24.63	21.73	90.92	35.08
K-NN	97.95	76.69	0.45	92.73	76,69	83.95
Ensemble	99.39	98.25	0.15	97.94	93.25	95.54

We have evaluated the performance of our proposed collaborative IDS scheme with various other models like SRPDBG (Kaliappan *et al.*, 2015), CrossLayer (Shrestha *et al.*, 2010), SPF (Tseng *et al.*, 2003),Watchdog (Marti *et al.*, 2000), TWOACK (Liu *et al.*, 2007)and EAACK (Shakshuki *et al.*, 2013). The following metrics were used for evaluation of the proposed IDS scheme with other IDS schemes:

- Packet delivery ratio (PDR) refers to the ratio of the number of packets delivered to the destination node against the number of packets generated by the source node.
- Routing overhead (RO*)* refers to the overhead involved in transmission due to introduction of additional routing control packets.

Figures 1 and 2 show the *PDR* and *RO* of the various IDS schemes under varying percentage of malicious nodes. It can be observed from these figures that all the four schemes (TWOACK, EAACK, SRPDBG and proposed IDS) have higher *PDR* than the simple Watch Dog scheme. The *PDR* of our proposed IDS scheme is comparable to that of EAACK and CrossLayer schemes, while it outperforms the TWOACK and SRPDBG schemes. On the other hand, the Watchdog scheme has the least *RO*, as it does not use any acknowledgment scheme to detect misbehaving nodes.

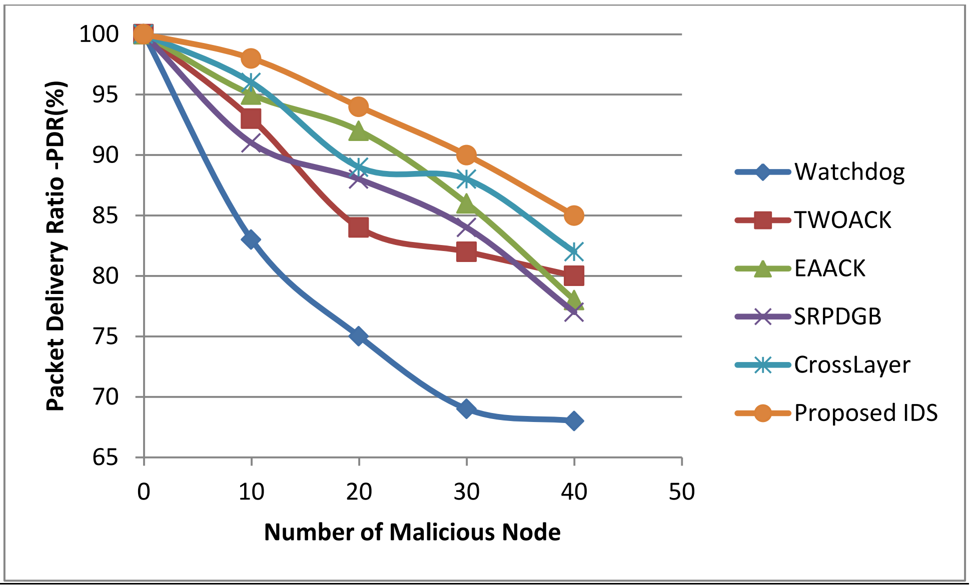

Figure-1. Packet delivery ratio.

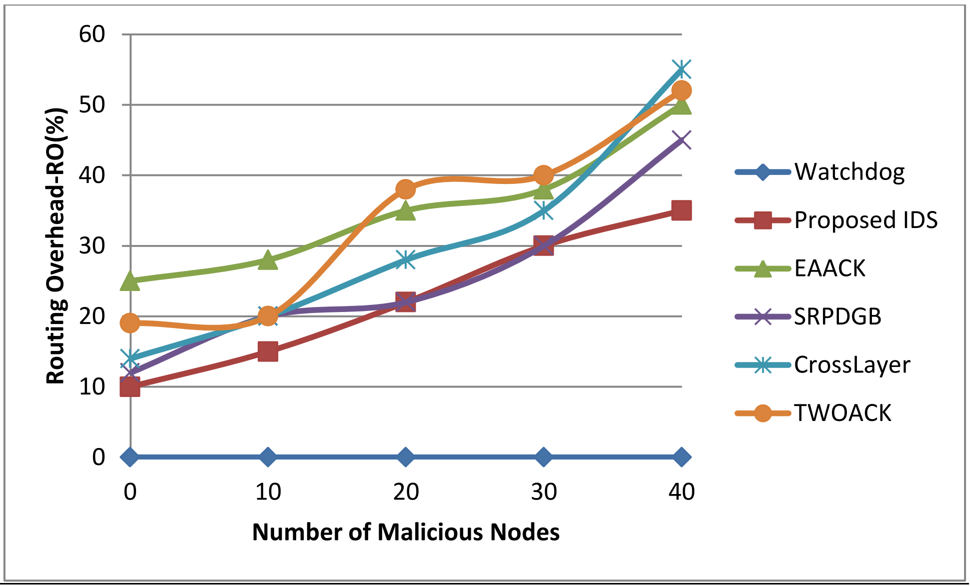

Figure-2. Routing overhead.

The *RO* of the proposed IDS is less than the TWOACK, EAACK and Cross Layer schemes but higher than the SRPDBG scheme. The *RO* of the proposed IDS scheme is also better than the other existing IDS primarily due to fewer exchanges of control messages for detecting intrusion.

CONCLUSIONS

In this work, we propose a collaborative framework for IDS in MANET to detect intrusion packets and normal packets with ensemble of four classifiers. Given a packet, our collaborative IDS framework will set an alarm if the packet is identified as malicious. Otherwise, it will be categorized as a normal packet. We employ ensemble of four classifiers, namely, SVM, CART, NB and K-NN with majority voting for the detection of intrusion and the normal packets. The experimental results show that our ensemble method is more robust than the other four base classifiers in the detection. In the experiments of intrusion detection, our ensemble method achieves the detection accuracy as 99.39%. The collaborative architecture is achieved with the deployment of friendly IDS nodes in the networks which are selected based on their trustworthiness.

In future work, we plan to explore the problem trust management in a better way thereby improving the performance of the entire intrusion detection mechanism. Designing more effective ensemble algorithms can also be investigated.

REFERENCES

Bayes T. An essay towards solving a problem in the doctrine of chances, Philos. Trans. R. Soc. 53(1763): 370-418.

Chen Y.-M, D. Wu, C.-K. Wu. 2010. A game theoretic framework for multi-agent deployment in intrusion detection systems, in: Security Informatics, vol. 9, Annals of Information Systems, Springer. pp. 117-133.

Fix E, J.L. Hodges. Discriminatory analysis: Nonparametric discrimination: Small sample performance, Technical Report Project 21-49-004, Report Number 11, 1952.

Ghosh A.K., Wanken J., Charron F. 1998. Detecting anomalous and unknown intrusions against programs. In: Proceedings of the 1998 Annual Computer Security Applications Conference (ACSAC). pp. 259-267.

Haddadi F, M. Sarram. 2010. Wireless intrusion detection system using a lightweight agent, in: Second International Conference on Computer and Network Technology. pp. 84-87.

Hamed Janzadeh, Kaveh Fayazbakhsh, Mehdi Dehghan, Mehran S. Fallah. 2009. A secure credit-based cooperation stimulating mechanism for MANETs using hash chains, Future Gener. Comput. Syst. 25(8): 926-934.

Humaira Ehsan, Farrukh Aslam Khan, Malicious AODV: 2012. Implementation and analysis of routing attacks in MANETs, in: Proceedings of 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom. Liverpool, UK. pp. 1181-1187.

Kaliappan M., B. Paramasivan. 2015. Enhancing secure routing in Mobile Ad Hoc Networks using a Dynamic Bayesian Signalling Game model, Comput. Electr. Eng. 41: 301-313.

Liu K, J. Deng, P.K. Varshney, K. Balakrishnan. 2007. An acknowledgment-based approach for the detection of routing misbehavior in MANETs, IEEE Trans. Mob. Comput. 6(5): 536-550.

Liu P. 2003. Incentive-based modeling and inference of attacker intent, objectives, and strategies, in: Proceeding of the 10th ACM Computer and Communications Security Conference. pp. 179-189.

Liu Y, C. Comaniciu, H. Man. 2006. A Bayesian game approach for intrusion detection in wireless ad hoc networks, in: Proceedings of the 2006 Workshop on Game Theory for Communications and Networks, ACM.

Marti S, T.J. Giuli, K. Lai, M. Baker. 2000. Mitigating routing misbehavior in mobile ad hoc networks, in: Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, ACM. pp. 255-265.

Misra S, P. Krishna, K. Abraham. 2010. Energy efficient learning solution for intrusion detection in Wireless Sensor Networks, in: Second International Conference on Communication Systems and Networks. pp. 1-6.

Mohanapriya M, I. Krishnamurthi. 2014. Modified DSR protocol for detection and removal of selective black hole attack in MANET, Comput. Electr. Eng. 40(2): 530-538.

Muhammad Imran, Farrukh Aslam Khan, Haider Abbas, Mohsin Iftikhar. 2014. Detection and prevention of black hole attacks in mobile ad hoc networks, in: Proceedings of Security in Ad Hoc Networks (SecAN) Workshop, 13th International Conference on Ad-Hoc and Wireless Networks, Ad Hoc Now 2014, Benidorm, Spain.

Nekovee M, R.S. Saksena. 2010. Simulations of large-scale WiFi-based wireless networks: Interdisciplinary challenges and applications, Future Gener. Comput. Syst. 26(3): 514-520.

Quinlan J. 1986. Introduction of decision tree, Mach. Learn. 1(1): 81-106.

ShresthaR, K.-H. Han D.-Y. Choi S.J. Han. 2010. A novel cross layer intrusion detection system in MANET, in: 24th IEEE International Conference on Advanced Information Networking and Applications (AINA). pp. 647-654.

Scarfone K., Mell P. 2007. Guide to Intrusion Detection and Prevention Systems (IDPS), NIST Special Publication. 800-94, February.

Shakshuki E.M, N. Kang, T.R. Sheltami. 2003. EAACK - a secure intrusion-detection system for MANETs. IEEE Trans. Ind. Electron. 60(3): 1089-1098.

Tseng C.-Y, P. Balasubramanyam, C. Ko, R. Limprasittiporn, J. Rowe, K. Levitt. 2003. A specification-based intrusion detection system for AODV, in: Proceedings of the 1st ACMWorkshop on Security of Ad Hoc and Sensor Networks. pp. 125-134.

Vapnik V. 2000. The Nature of Static Learing Theory, Springer.

Vasilomanolakis E., Karuppayah S., Muhlhauser M., Fischer M. 2015. Taxonomy and survey of collaborative intrusion detection. ACM Comput. Surv. 47(4): 55.

Vigna G., Kemmerer R.A. 1998. Netstat: a network-based intrusion detection approach. In: Proceedings of the 1998 Annual Computer Security Applications Conference (ACSAC). pp. 25-34.

Wu Y. S, Foo B., Mei Y., Bagchi S. 2003. Collaborative intrusion detection system (CIDS): a framework for accurate and efficient IDS. In: Proceedings of the 2003, Annual Computer Security Applications Conference (ACSAC). pp. 234-244.