Data Mining in Intruder Detection in MANET: A Survey

[1]Shalu Verma, [2]Mrs. Nisha Charaya

[1]M.Tech Scholar, Computer Science and Engineering, Om Institute of Technology and Management, Juglan (Hisar)

[2]Assistant Professor, Computer Science and Engineering, Om Institute of Technology and Management, Juglan (Hisar)

Abstract: MANET attack is increasing day by day. The intruder affects the data and losses are increasing. To avoid it several researchers have worked in this direction. Some has applied algorithms on node end without any database, some used special databases of previously compromised history and used soft computing algorithms to mitigate that. In this paper we focused to review on latest work done in this field. We kept more papers based on data mining in MANET as these are using artificial intelligence for it.

I. INTRODUCTION

Conventional cellular wireless mobile networks that consider in depth infrastructure to support quality, MANETs don't want high-priced base stations or wired infrastructure. The absence of a set infrastructure needs mobile hosts in MANETs to work with one another for message transmissions. to create such a cooperative self-configurable surroundings, each mobile host is meant to be a friendly node and is willing to relay messages for others to their final destinations. world trustiness altogether network nodes is that the main basic security assumption in MANETs. However, this assumption isn't continually true actually. the character of MANETs makes them terribly liable to malicious attacks starting from passive eavesdropping to active meddlesome. Most routing protocols solely target providing economical route discovery and maintenance practicality and pay very little attention to routing security. only a few of them specify security measures from the terribly starting. the character of MANETs makes them terribly liable to malicious attacks compared to ancient wired networks, owing to the employment of wireless links, the low degree of physical security of the mobile nodes, the dynamic topology, the restricted power provide and therefore the absence of central management purpose. Some environments (such because the military plan of action operations) have terribly rigorous needs on security, that build the preparation of security-related technologies necessary. Intrusion hindrance measures, resembling coding and authentication, will be employed in MANETs to cut back intrusions, however cannot eliminate them. for instance, a physically captured node that carries the personal keys could permit the defeat of the authentication safeguards. The history of security analysis has incontestable that regardless of what percentage intrusion hindrance measures ar used, there ar continually some weak points within the system. in an exceedingly network with high security needs, it's necessary to deploy intrusion detection techniques. MANET IDSs, serving because the second wall of defence to shield MANETs, ought to operate beside hindrance mechanisms (authentication, coding etc.) to ensure associate surroundings with high secure needs. they ought to complement and integrate with alternative Manet security measures to produce a high-survivability network. However, most of today's Intrusion Detection Systems (IDSs) target wired networks. The dramatic variations between MANETs and wired networks build it unsuitable to use ancient wired ID technologies on to MANETs. Manet doesn't have a set infrastructure. whereas most of today's wired IDSs, that consider period of time traffic dissect, filter, format and analysis, sometimes monitor the traffic at switches, routers, and gateways. the dearth of such traffic concentration purpose makes ancient wired IDSs unsuitable on Manet platforms. every node will solely use the partial and localized communication activities because the offered audit traces. There also are some characteristics in Manet resembling disconnected operations, that rarely exist in wired networks. What's additional, every mobile node has restricted resources (such as restricted wireless information measure, computation ability and energy provide, etc.), which suggests Manet IDSs ought to have the property to be light-weight. All of those imply the irrelevance of wired IDSs on the Manet platform. what is more, in MANETs, it's terribly tough for IDSs to inform the validity of some operations. for instance, the explanation that one node sends out falsified routing info can be as a result of this node is compromised, or as a result of the link is broken thanks to the physical movement of the node. of these recommend that associate IDS of a special design must be developed to be applicable on the Manet platform.

II. LITERATURE REVIEW

P. Tao et. al [1](2018) proposed the feature selection along with classification to detect the intruder in MANET. Genetic algorithm was used for the feature selection and reduction purpose followed by SVM classification. This increase the accuracy and decrease the error rate. The optimal feature selection process is depicted as in figure.

Z. Ullah*et al. [2] (2016)* proposed a fuzzy-based scheme to detect and isolate non-cooperative nodes in MANETs. In the proposed scheme, actions of the neighbours are detected by every node in the network and compute the trust of the observed neighbours. After computing these trust values are passed on to a fuzzy function which is mapped into different classes. Trust levels of the observed nodes are shown by resulting classes. To detect packet drop or decaying time, fuzzy logic is proposed. The proposed scheme has low false error rate. Packet delivery ratio and throughput is also enhances by this scheme.

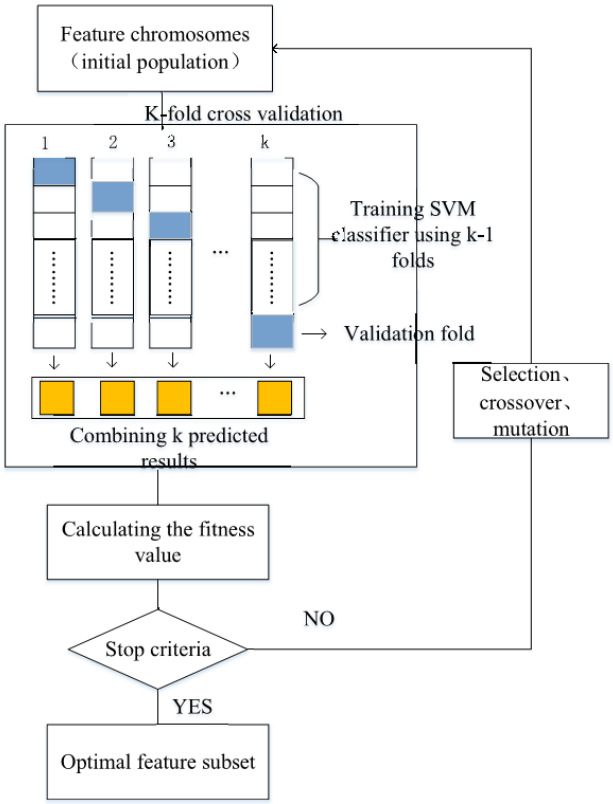

Figure: Feature Selection Algorithm using GA and classification by SVM [1].

K. Majumder*et al. [3] (2016)* proposed a new scheme based on game theory has been developed that will detect selfish nodes and avoid malicious nodes of the system for packet transmission. During data transmission if any node of the selected path moves out of the radio range, then an alternate backup route will be formed from that position so that data transmission never stops. So, this scheme guarantees secure routing and constructs alternate backup route for guaranteed packet transmission. This scheme also ensures minimum amount of idle time.

Manjula C *et al. [4] (2016)* Studied classification and predictive models for intrusion detection are built by using machine learning classification algorithms namely Logistic Regression, Gaussian Naive Bayes, Support Vector Machine and Random Forest. These algorithms are tested with NSL-KDD data set. Experimental results show that Random Forest Classifier out performs the other methods in identifying whether the data traffic is normal or an attack. Machine Learning algorithms are used to build accurate models for clustering, classification and prediction.

D. A. Varma *et al. [5] (2016)* used the modified polynomial reduction algorithm to detect the malicious nodes in MANET network. the neural network in the polynomial algorithm studies the characteristics of the abnormal node behaviour. NS2 simulator was used for the work.

Bandana Mahapatraa*et al. [6] (2016)* suggested the intrusion detection by using the self adaptive fuzzy logic method. They used Neural network to train and changes the fuzzy logic membership functions so that it can adapt to changing attacks properties. NSL-KDD dataset has been used for it.

M. A. Abdelshafy*et al. [7] (2015)* studied the performance of DSR and its flow-state extension routing protocols in the presence of black hole, gray hole, selfish and flooding attacks. DSR does not support security of routing messages and it is a reactive MANET routing protocol. The performance of flow-state DSR is better than DSR in the presence of all attacks. All the standard performance of the network is affected by flooding attacks and black hole attacks decrease the packet delivery ratio in a static network using unmodified DSR. So the overall delay of the system is increased.

C. Alociou*set al.* [8] *(2015)* studied possible MAC layer DoS attack strategies that are driven by the MAC layer malicious/selfish nodes on IEEE 802.11 protocol. Such DoS attacks could caused by malicious and selfish nodes by violating their backoff timers associated with the MAC protocol. The performance and stability in the MANETs are analysed and evaluated by these attacks. These attacks also affect the network throughput and data packet collision rate. This paper concludes that DoS attacks with selfish/malicious intend can obtain at least 50% larger throughput by denying well behaved nodes to obtain deserved throughput.

A. Quyoom*et al. [9] (2015)* proposed a Malicious and Irrelevant Packet Detection Algorithm (MIPDA) which is used to analyse and detect the Denial-of Service (DoS) attack. (DoS) attacks are major threat to the information economy. The attack is avoiding wasteful attack traffic overloading the network infrastructure. It also reduces the overhead delay in the information processing, which increases the communication speed and also enhances the security in VANET.

A. M. Shabut*et al. [10] (2015)*proposed a Recommendation based trust in the literature as a mechanism to filter out the misbehaving nodes while searching for a packet delivery route. However, building a trust model that adopts recommendations by other nodes in the network is a challenging problem due to the risk of dishonest recommendations like bad-mouthing, ballot-stuffing, and collusion. A recommendation based trust model with a defence scheme, utilises clustering technique to dynamically filter out attacks related to dishonest recommendations between certain time based on number of interactions, compatibility of information and closeness between the nodes.

P. Aggarwal *et al. [11] (2015)* presented the analysis using KDD dataset and analysed the results using metrics; Detection rate (DR) and false Alarm Rate (FAR). They used Weka tool for random tree classifier for analysis. The outcome in terms of false alarm rate is shown in figure.

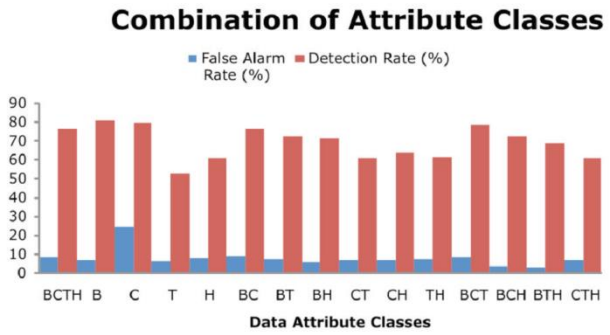

Figure: Detection rate and False alarm rate for 15 combinations of attribute classes [11]

The data has four categories: basic, Content, Host and Traffic. Analysis showed that basic class has low DR, traffic class data has high FAR, data in content class has low FAR and Host class data has low DR but decent FAR.

T. Poongothai*et al. [12] (2014)* studied about intrusion detection systems (IDS) for securing MANET. They used features selection approach before classification to reduce and optimise the data set so that features which don't contribute in accuracy can be removed. Rough set theory (RST) was proposed for this purpose followed by SVM classification algorithm to detect the attack. Here is the flow chart proposed in this paper.

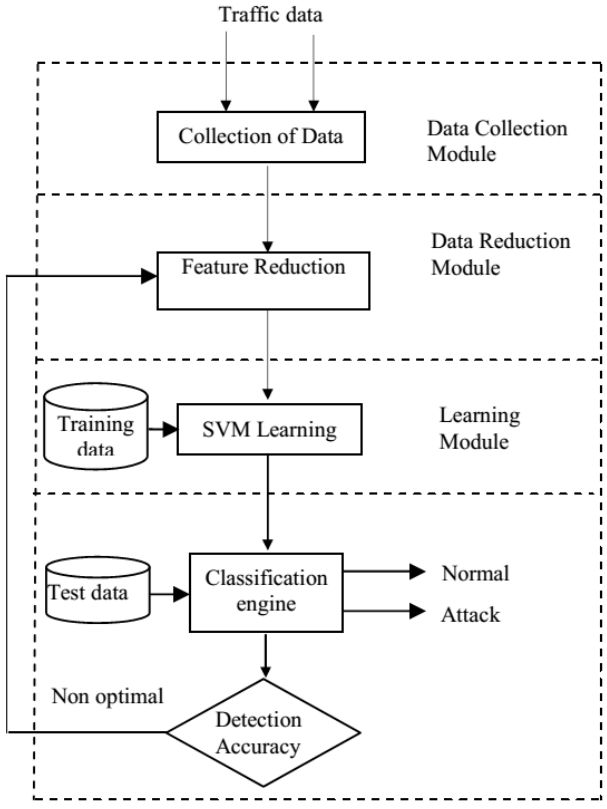

Figure: Proposed Flow chart for intrusion Detection [12]

A. M. Kurkure *et al.[13] (2014)* proposed credit based system in MANET network using AODV protocol to avoid the selfish nodes which can amend the data coming to that node. That node has to be detected and removed from the network. For the evaluation three different metrics: no of selfish node detected, end to end delay and success ratio are used.

M. S. Pervez et.al [14] (2014) also focused on feature selection and classification approach for intruder detection. They used the old NSL-KDD 99 dataset. For the feature selection approach, rather than opting any particular method, a loop for all number of features for SVM classification was used. The classification rate vs number of feature are plotted to get the idea for optimal number of features in figure.

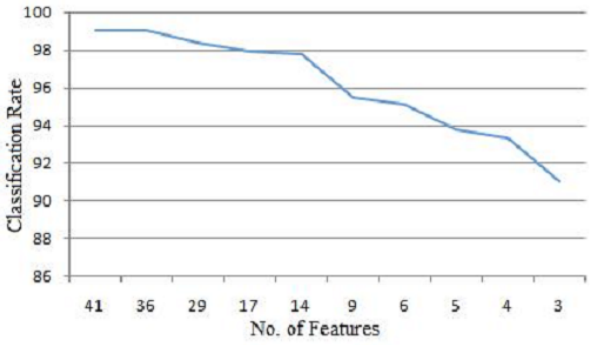

Figure: Classification accuracy vs Number of features for 100% training data [14]

A. Menaka Pushpa *et al [15] (2013)* analyzed a mesh-based multicast routing protocol PUMA against various types of protocol-dependent internal attacks. Also proposed a solution for two particular internal attacks; watchdog-based data packet dropper attacker identification and validating parent's information from core using newly added control packet MA-to-Core for MA packet modification attacks. This research work focused on security vulnerabilities of PUMA routing protocol and evaluated our proposed solutions for internal attacks using simulation results.

M. A. Abdelshafy*et.al [16] (2013)* presented a vulnerability analysis of AODV. It simulated four routing attacks to analyse their impacts on AODV protocol using NS-2 network simulator. These attacks are black hole, gray hole, selfish and flooding attacks. The black hole and flooding attacks have a severe impact on the network performance while the selfish and gray hole attacks have less significant effect on the network performance.

S. Biswas*et al. [17] (2012)* proposed a trust based model in MANET network to avoid the loss of data in case of attack. Checkpoints are used to keep the information of data and encrypt at those points so that in case of data loss, recovery can be easy. Ant colony optimisation is proposed to select the check points for that purpose. These checkpoints has negligible failure and high available battery power. ACO method select

the nodes with full of these resources. A working example of network in this paper is shown in figure.

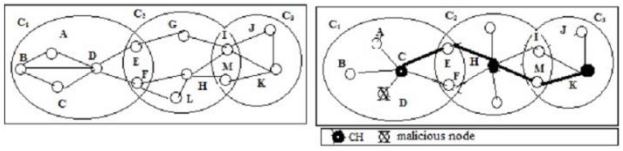

Figure: (a)working example of the network considered in our work;(b) formation of cluster after evaluating trust value of nodes [17]

C. Thomas *et al. [18] (2012)*provided supporting facts for the use of the DARPA IDS evaluation dataset. The two commonly used signature-based IDSs, Snort and Cisco IDS, and two anomaly detectors, the PHAD and the ALAD, are made use of for this evaluation purpose and the results support the usefulness of DARPA dataset for IDS evaluation.

P.Natesan*et al. [19] (2012)* proposed a design of multi stage filter which is an efficient and effective approach in dealing with various categories of attacks in networks. The first stage of the filter is designed using Enhanced Ada-boost with Decision tree algorithm to detect the frequent attacks occurs in the network and the second stage of the filter is designed using enhanced Ada-boost with Naive Byes algorithm to detect the moderate attacks occurs in the network. The final stage of the filter is used to detect the infrequent attack which is designed using the enhanced Ada-boost algorithm with Naive Bayes as a base learner.

M. Tavallaee *et al. [20] (2009)*analysed the weakness in detecting novel attacks by signature based methods for intruder detection in MANETs. They use most famous and publically available dataset KDDCUP'99 for the system's evaluation. Two major issues was detected by author in the dataset: one is the dataset is not experimentally validated for false alarms characteristics, another issue found is the lacking in the attack definition; for example probe attack in the dataset is not the attack unless iterations don't exceed the threshold. To solve this issue he removed the records from the dataset which are not contributing to validation and called this dataset as NSL-KDD dataset. This dataset has very less rate of false alarms.

III. CONCLUSION

The work on Intruder detection in MANETS is done in several ways. We have studied the papers from 2013-2018 with maximum publications in time period in 2014-2016. Papers are chosen which are using data mining techniques for IDs. We have seen NSL-KDD dataset has been widely used for this purpose on which machine learning algorithms work to learn the intruders behaviour and detect the upcoming attack. Very few researchers have used feature selection approach before classification. The feature selection approach improves the accuracy in detection as well reduce the computation time. Our future work will be based on these steps. On NSL-KDD dataset we use heuristic optimisation to select the more relevant features along with SVM classification technique.

IV. REFERENCES

[1] P. Tao, Z. Sun and Z. Sun, "An Improved Intrusion Detection Algorithm Based on GA and SVM," in *IEEE Access*, vol. 6, pp. 13624-13631, 2018.

[2] Z. Ullah, M. S. Khan, I. Ahmed, N. Javaid and M. I. Khan, "Fuzzy-Based Trust Model for Detection of Selfish Nodes in MANETs," *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, Crans-Montana, 2016, pp. 965-972.

[3] D. Das, K. Majumder and A. Dasgupta, "A game-theory based secure routing mechanism in mobile ad hoc network," *2016 International Conference on Computing, Communication and Automation (ICCCA)*, Noida, 2016, pp. 437-442.

[4] Manjula C. Belavagi and BalachandraMuniyal, "Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection," 12th International Multi-Conference on Information Processing-2016 (IMCIP-2016).

[5] D. A. Varma and M. Narayanan, "Identifying malicious nodes in Mobile Ad-Hoc Networks using polynomial reduction algorithm," *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, Chennai, 2016, pp. 1179-1184.

[6] Bandana Mahapatraa and Prof.(Dr) Srikanta Patnaik, "Self Adaptive Intrusion Detection Technique Using Data Mining concept in an Ad-Hoc Network," 2nd International Conference on Intelligent Computing, Communication & Convergence(ICCC-2016)

[7] M. A. Abdelshafy and P. J. B. King, "Dynamic source routing under attacks," *2015 7th International Workshop on Reliable Networks Design and Modeling (RNDM)*, Munich, 2015, pp. 174-180.

[8] C. Alocious, H. Xiao and B. Christianson, "Analysis of DoS attacks at MAC Layer in mobile adhoc networks," *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Dubrovnik, 2015, pp. 811-816.

[9] A. Quyoom, R. Ali, D. N. Gouttam and H. Sharma, "A novel mechanism of detection of denial of service attack (DoS) in VANET using Malicious and Irrelevant Packet Detection Algorithm (MIPDA)," *International Conference on Computing, Communication & Automation*, Noida, 2015, pp. 414-419.

[10] A. M. Shabut, K. P. Dahal, S. K. Bista and I. U. Awan, "Recommendation Based Trust Model with an Effective Defence Scheme for MANETs," in *IEEE Transactions on Mobile Computing*, vol. 14, no. 10, pp. 2101-2115, Oct. 1 2015.

[11] PreetiAggarwala and Sudhir Kumar Sharmab, "Analysis of KDD Dataset Attributes - Class wise For Intrusion Detection," 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).

[12] T. Poongothai and K. Duraiswamy, "Intrusion detection in mobile AdHoc networks using machine learning approach," *International Conference on Information Communication and Embedded Systems (ICICES2014)*, Chennai, 2014, pp. 1-5.

[13] A. M. Kurkure and B. Chaudhari, "Analysing credit based ARAN to detect selfish nodes in MANET," *2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)*, Unnao, 2014, pp. 1-5.

[14] M. S. Pervez and D. M. Farid, "Feature selection and intrusion classification in NSL-KDD cup 99 dataset employing SVMs," *The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014)*, Dhaka, 2014, pp. 1-6.

[15] A. Menaka Pushpa and K. Kathiravan, "Resilient PUMA (Protocol for Unified Multicasting through Announcement) against internal attacks in Mobile Ad hoc Networks," *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Mysore, 2013, pp. 1906-1912.

[16] M. A. Abdelshafy and P. J. B. King, "Analysis of security attacks on AODV routing," *8th International Conference for Internet Technology and Secured Transactions (ICITST-2013)*, London, 2013, pp. 290-295.

[17] S. Biswas, P. Dey and S. Neogy, "Trusted checkpointing based on ant colony optimization in MANET," *2012 Third International Conference on Emerging Applications of Information Technology*, Kolkata, 2012, pp. 433-438.

[18] Ciza Thomas, Vishwas Sharma and N. Balakrishnan, "Usefulness of DARPA Dataset for Intrusion Detection System Evaluation,2012

[19] P.Natesan and P.Balasubramanie, "Multi Stage Filter Using Enhanced Adaboost for Network Intrusion Detection," International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.3, May 2012

[20] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Ottawa, ON, 2009, pp. 1-6

	Author	Year	Technologies Used	Advantages
1	P. Tao, Z. Sun and Z. Sun	2018	Genetic algorithm to reduce features with SVM classification	More accuracy in intruder detection than only SVM with all features
2	Z. Ullah, M. S. Khan, I. Ahmed, N. Javaid and M. I. Khan	2016	Fuzzy logic based trust on neighbourhood node is calculated	Low false error rate and more packet delivery ratio
3.	D. Das, K. Majumder and A. Dasgupta	2016	Proposed game theory based approach to block the malicious nodes in network	A continuous communication is guaranteed even if a malicious node in a route is detected.
4	Manjula C. Belavagi and Balachandra Muniyal	2016	Compared Logistic Regression, Gaussian Naive Bayes, Support Vector Machine and Random Forest algorithms for NSL KDD dataset	Random forest method found more suitable
5.	D. A. Varma and M. Narayanan	2016	used the modified polynomial reduction algorithm with neural network	NA
6.	Bandana Mahapatraa and Prof.(Dr) Srikanta Patnaik	2016	Used self adaptive fuzzy logic tuned by neural network	More accurate in detection than fuzzy logic based intruder detection system
7	. A. Abdelshafy and P. J. B. King	2015	Studied the security of MANET using DSR	performance of flow-state DSR is better than DSR in the presence of all attacks
8	C. Alocious, H. Xiao and B. Christianson	2015	Focused MAC layer DOS attack	Found that 50% of throughput can be stopped by this attack
9	A. Quyoom, R. Ali, D. N. Gouttam and H. Sharma	2015	Malicious and Irrelevant Packet Detection Algorithm (MIPDA)	Reduced the overhead delay in the MANET
10	A. M. Shabut, K. P. Dahal, S. K. Bista and I. U. Awan	2015	Used a recommendation based trust model to communicate with nodes in MANET based on their	Nodes which don't behaved honestly in past are out pass from network

			previous behaviour	
11	PreetiAggarwala and Sudhir Kumar Sharmab	2015	Used Weka tool classifier to detect on the KDD cup dataset	Attacks with different sub categories are well classified with random forest in weka
12	T. Poongothai and K. Duraiswamy	2014	Rough set theory (RST) was used to reduce features along with SVM classifier on KDD dataset	Selected features reduced overhead and improved accuracy in detcetion
13.	A. M. Kurkure and B. Chaudhari	2014	Used credit based system to detect the selfish node/malicious node	Removed the corrupted nodes from the network
14	M. S. Pervez and D. M. Farid		Feature selection approach based on using one by one feature to check the accuracy	Accuracy is higher with more number of features in KDD-99 dataset