CHAPTER 2
LITERATURE REVIEW

This chapter explores the study of relevant literature; the scholar has produced since year 1990. Validation of the efforts of the present research work with its results and conclusions is only justified, if sufficient time has elapsed to the realization of various assignments in the field of outlier node detection and management of trust in MANETs. The subsequent review of research papers published by the eminent researchers is helpful in crystallizing the research plan and it provides support for designing the methodology of the present study. This review work explores the groundwork of outlier detection mechanism in MANETs and a survey on the existing techniques of outlier detection is presented. In the frequent paragraphs, there is a discussion on the topics which had been previously undertaken regarding outlier detection techniques available in Data Mining, trust management and clustering issues in MANETs. Therefore, the present review work is divided into four sections as given below.

The following is the organization of this chapter. Section 2.1 explores review on Data mining techniques used for outlier detection. Section 2.2 discusses literature on trust management in MANETs and the usage of trust in detection of misbehaving nodes. Section 2.3 examines review on various clustering approaches in MANETs that helps in effective utilization of resources. Section 2.4 presents the miscellaneous literature review which may relate to present research work directly or indirectly and Section 2.5 provides the outcome of the chapter. These sections are elaborated in the upcoming paragraphs.

2.1 Section-I

This section includes the research work carried out by many researchers in the field of outlier detection. In Data Mining, there is a huge research done regarding the topic of outlier detection. The main inspiration behind the research of outlier detection is to proficiently recognize outliers in very large databases. However in MANETs, the scenario is quite different from the large databases. In such type of networks, data is generated as well as stored in distributed nodes. The transmission of the data takes place through wireless channels, which are considered as untrustworthy and power
restricted. So, direct application of outlier detection techniques which are designed for the large-scale databases into mobile ad hoc networks is very challenging because it may result into lots of communication overhead. There are a number of outlier detection methods proposed in various application areas e.g. large scale databases, wireless sensor networks and mobile ad-hoc networks, which are reviewed in the upcoming paragraphs.

In Data Mining, an outlier detection method based on distance was at first presented by (Knorr EM and Ng RT, 1999). The authors defined an entity p in a given data set as distance based outlier denoted as DB (q, dist) if q part of the data set occurred at a distance greater than variable dist from p. The above definition generalized several statistical outlier tests. (Ramaswamy S., et al., 2000) further extended the definition given above. The authors stated that an entity p was considered as outlier if for two integers defined as k and w, if less than w objects contained higher value for Dk than p, where Dk denoted the distance of the kth nearest neighbour of the object p. All points were ranked based on the outlier score.

(C. C. Aggarwal and P. S. Yu., 2001) described cluster based outliers. All those points, which were found external to the clusters and isolated from the noise were considered as outliers. The behaviour of such points found to be different than the normal ones. The requirement of numerous passes in order to process the given data was found to be the drawback of such type of approaches.

The issue of identifying outliers in a huge database was solved by (Simon Hawkins, et al., 2002). Replicator Neural Network was used for measurement of the abnormality found in available data. A rank based mechanism was used for evaluating the performance of such type of networks.

An outlier detection approach for large dataset was proposed by (George Kollios, et al., 2003). To accelerate the task of clustering and outlier detection, biased sampling was used for reducing data. The uniform random sampling was considered as a special case of biased sampling, in which all the points in the given sample had the similar probability. However, the likelihood of adding a data point to the sample in biased sampling was unique in relation to point to point.

(Victoria j. Hodge and Jim austin, 2004) found that initially the outlier detection approaches were found not to be appropriate and presented a logically correct
approach drawn from the range of Computer Science and Statistics. The authors provided a review of existing techniques of outlier detection and recognized the stimulus behind them in a relative view.

(Ville Hautamaki, et al., 2005) presented an algorithm named as ORC (Outlier Removal Clustering) which worked on the concepts of outlier detection as well as data clustering at the same time. To improve the estimated centers of allocation, the strategy merged clustering into discovery of outliers. The authors used K-means algorithm for making clusters and discovered the vectors that were a long way from their cluster centers.

(Branch J., et al., 2006) introduced a method of outlier detection to detect the outliers in wireless sensor networks. The working of the proposed method began by calculating outlier nodes which were local to the network. After that, a few messages accommodating the neighbourhood outliers and additional relevant data were broadcasted to all nodes in the network. This process of exchanging the messages continued till generation of the similar global outlier view.

Further the concept of anomaly detection in wireless systems was elaborated by (Animesh Patcha and Jung-Min Park, 2007). The survey made by the authors highlighted various historical as well as existing techniques of anomaly detection and suggested the scope of applications of these techniques. The authors also identified various issues and difficulties around there.

By taking motivation from the work proposed by (Branch J., et al., 2006) a collaborative and gossip based approach for identifying outlier nodes in MANETs, was developed by (Wenjia Li, et al., 2008). The authors stated that, first of all, each node calculated behaviour information of the neighbouring nodes by watching their performance in the network. Based on the behaviour data, a local outlier view was generated by every node. After that, the local views get exchanged among the neighbouring nodes. Then they updated their local views if they found outlier lists from other nodes more accurate than their own. That process continued, with each node updating its neighbours whenever there is some change in current outlier view and halted when there were no more changes. In further addition (Wenjia Li and Anupam Joshi, 2009) used the Weighted Voting Method and Dempster-Shafer Theory of evidence to combine observation results from multiple nodes. The results improved
the accuracy rate of finding out misbehavior nodes.

An outlier detection technique ODMAD (Outlier Detection for Mixed Attribute Data sets) based on anomaly score was described by (Koufakou, et al., 2010). The propose score found to be useful in those approaches which mine outliers from data which contains both continuous and categorical attributes. In this, for each point, the anomaly score was computed by taking into consideration the abnormality of the continuous values, categorical values and the relations between two spaces in the data set.

Outlier detection considered as an important anomaly detection approach. (Prasanta Gogoi, et al., 2011) provided comparison of important outlier detection techniques that could be applied in wireless networks for the detection of anomalies. The defined methods helped in finding out system errors so that supervisor could take precautionary actions before they appeared.

A clustering based approach to capture outliers was defined by (Pamula R., et al., 2011). In this approach, the clusters are generated by using K-means algorithm of clustering. The points that lie close to the center of the cluster were not considered as possible candidates for outliers. For remaining points, the authors calculated an outlier score based on distance. Due to the filteration of some points, the computation complexity to generate the outlier score was decreased greatly. In this approach, topmost n number of points having highest value of score were selected as outliers.

(Naveen N.C., et al., 2012) identified the scope of using change point outlier detection technique of intrusion detection in real time. The authors introduced a novel and integrated approach, which embedded an algorithm of change detection into Data Mining for the analysis of intrusions in MANETs. Detection of the variations observed in collected data on different times considered as the most essential application of security as it provided analysis of short interval on global scale.

An outlier detection strategy for lightweight MANETs was proposed by (Adarsh Kumar, et al., 2013). The authors defined work to select and avoid exceptional data which was identified from lightweight mobile nodes. They modified the existing outlier detection model for wireless networks by adding the concepts of event detection and group key management.

(Nauman Shahid, et al., 2013) presented various features of outlier detection
methodologies for implementation in harsh environments. A mechanism was developed to recognize the features an outlier detection method should have in such type of environments. These features were e.g. input data type, type of outliers, spatio-temporal correlations, types of approach (centralized or distributed), outlier recognition, susceptibility to dynamic topology etc. The authors also discussed feasibility study of various outlier detection mechanisms in such environment.

(Dylan McDonald, et al., 2013) generated various solutions in WSN for outlier detection to find outliers by using various approaches such as centralized and distributed approaches. They identified that for limited range of data sets, distance based solutions provided more optimal results. Density based algorithms were slower in execution as compared to distance based algorithms. These types of algorithms should make a typical compromise between correctness and cost of communication.

An improvement in outlier detection approaches with imperfect labels was suggested by (Bo liu, et al., 2014). The authors introduced likelihood values for the input data to handle data with imperfect labels. The input data denoted the membership degree of an example with respect to regular and irregular classes respectively. Based on the local behavior of the nodes, a pseudo training dataset was generated by calculating likelihood values for all example set. In this, two approaches were used to calculate the likelihood values which were LOF (Local Outlier Factor) and k-means method of clustering.

In 2015, (Nirav J. Patel and Rutvij H. Jhaveri, 2015) used Support Vector Machine (SVM) to differentiate nodes in two classes i.e. regular or malicious nodes in MANETs. The input to SVM was trust value of the neighbouring nodes that was calculated by using control and data packets. SVM had to train data through the neighbour trust value of neighbour node, then SVM classified it into two categories i.e. normal and malicious.

(J. James Manoharan, et al., 2016) provided an outlier detection method in Data Mining. The proposed method was based on improved k-means method of for making clusters of data sets and weight based center approach. Thus in the proposed approach, two methods were used efficiently for finding out the outliers from the data set. Threshold value then calculated programmatically by taking the absolute value of minimum and maximum value of a particular cluster.
In 2017, (Zeeshan Ahmad Lodhia, et al., 2017) discussed various important techniques of outlier detection with their applicability in different application scenarios. The data provided in the survey report helped in identifying various issues in real time data sets.

2.2 Section-II

This section includes the research work carried on management of trust. In MANETs, trust management plays an important role in security of the network. This section reviews various existing approaches for trust management, specifically developed for environment of MANETs. Trust management aims for building of a reputation value of the nodes by evaluating their behaviour.

Trust is abstract by nature and it has intrinsic relationship with the other social concepts, like ethics, cooperation and commodity. In different perspectives, trust is defined in different forms. (Dasgupta, 1990) stated that trust is the belief of a person on the actions of others which affects the first person’s choice, when an action should be performed before the actions of others are known. Social psychologists considered trust as an interpersonal phenomenon as suggested by (J. G. Holmes, 1991) whereas Economists viewed trust as a means of coherent selection for increasing its own value (O. E. Williamson, 1993). (T. Beth, et al., 1994) discussed the relationship between recommendations and trust. When a person does not know either to trust a specific person or not, a third person whom he trusts is asked. Similarly, the authors described reliability of the person that was based on the recommendations of third person. According to (D. J. McAllister, 1995), trust represented a person’s willingness and belief to take action based on the actions, words and decision of the others. (McKnight and Chervany, 1996) stated trust as the extent to which an individual depends on someone in a specific condition having a feel of relative security, although it might result into negative possibilities.

The term “Trust Management” was firstly introduced by (M. Blaze, et al., 1996) and they recognized it as an essential part of applying security in networks and illustrated that it provided an integrated approach to interpret and specify various policies, credentials in security and their relations. Trust management provided solution for various issues given as the representation of trust facts, formulation of evaluation policies and rules and management of trust relations between nodes.
An attractive distinction in the field of trust field was given by (L. Rasmusson and S. Janssen, 1996). The authors introduced two terms i.e. soft security and hard security to differentiate two approaches of security. Soft security contained mechanisms of social control called as reputation systems and trust management. Hard security referred to conventional methods such as authentication and access control.

(Aifarez Abdui-Rahman and Stephen Hailes, 1997) recognized the requirement of an efficient trust management scheme to be used for distributed networks. The authors introduced a recommendations based model of trust. To generalize the concept of trust, a decentralized approach of trust management was adopted in the model.

According to (C. Castelfranchi and R. Falcone, 1998), trust can be viewed as a function of belief, a trustor found in a trustee. The term belief was defined as the confidence degree of the truth for something based on evaluation and inspection of the evidences. The authors considered the important evidence was the experience of trustor on trustee, calculated through direct observations by the trustor. Thus experience represented the knowledge gained by the trustor by doing interactions with the trustee and called as experience belief. The authors found that experience might be incomplete and observed the need of opinions of other nodes also.

(Blaze M., et al., 1999) elaborated the trust management scheme by highlighting its essential features and described two existing trust management engines which were PolicyMaker and KeyNote. These trust engines in trust management provided a language to denote trusted actions and their relationships.

(S. Marti, et al., 2000) defined two approaches such as Watchdog and pathrater to identify and filter misbehaving nodes, i.e. the nodes which refused to transmit packets. The authors considered Watchdog as responsible for detecting the misbehaviour, and Pathrater, responsible for suggesting the most secure route to get out of these nodes.

(Pietro Michiardi and Refik Molva, 2001) introduced a reputation based trust management scheme named as, CORE (COllaborative Reputation). In that scheme, all those participants having good level of trust used the common resources of the network whereas the participants having bad level of trust were steadily debarred from the network.

(S. Buchegger, et al., 2002) presented efficient approach to encourage the cooperation
among the nodes, named as CONFIDANT (Cooperation Of Nodes, Fairness In Dynamic Ad-hoc Networks). The aim of CONFIDANT was to detect and isolate misbehaving nodes by the use of recommendations and direct observations. To handle the issue of false recommendations, the strategy of personal experience was used. The recommendations were tested using the deviation test and the nodes which showed more variations than threshold values were discarded. On the basis of the results obtained from the deviation test, the value of reputation of a recommending node was modified.

(P. Michiardi, et al., 2002) described a method for identifying selfish nodes from the network named as CORE (Collaborative Reputation mechanism). As analogous to CONFIDANT, the proposed method CORE applied both an observation system and a reputation system for observing and evaluating behaviour of the nodes. The nodes in CONFIDANT were allowed to transmit positive as well as negative behaviour of their neighbouring nodes, on the other hand in CORE, only positive behaviour observations got exchanged between the nodes. Thus the efficiency of the system decreased because of preventing the exchange of bad observations generated for malicious nodes.

(Sonja Buchegger and Jean Le Boudec, 2003) investigated the impact of rumors on the detection time of malicious nodes and the robustness of the reputation system with respect to incorrect charges. The authors used a Bayesian mechanism for representation, updation of reputation and integration of views. They also presented a methodology for the detection and isolation of probable lies.

A trust distribution mechanism named as ABED (Ant-Based trust Evidence Distribution) was introduced by (Tao Jiang, John S. Baras, 2004) in MANETs. This scheme worked on the basis of the methodology of swarm intelligence, which supposed to be highly resilient to mobility. The swarm intelligence model was motivated from artificial ant colony approach for providing solution to combinatorial optimization problems. The nodes in ABED, communicated with each other by means of agents called as “ants” which deposited patterns called as “pheromones”. As a result, the agents were able to find out the best possible path for collecting evidences of trust.

The trust evidence evaluation scheme was proposed by (G. Theodorakopoulos and J.
S. Baras, 2006) in MANETs. In the proposed scheme, the process of evaluation was depicted similar to path problem of a directed graph in which nodes indicate entities and trust relations were represented by edges. The theoretical concept of semirings was used by the authors to represent the establishment of trust relations in the absence of previous interactions.

(Sun Y.L., et al., 2006) proposed a theoretical framework of information to measure the trust value and represent propagation of trust in MANETs. In the proposed structure, the value of trust followed uncertainty and represented with entropy. The authors highlighted four axioms that included a basic knowledge of trust and set of policies for propagation of trust. On the basis of these axioms, two models of trust were presented such as Probability Based Model and Entropy Based Model. Various methods of establishment and updation of trust were also presented to obtain the value of trust from direct observations. The proposed trust models and trust evaluation methods were applied in MANETs for secure communication and misbehaviour detection.

Trust management was closely linked to reputation management as suggested by (H. Li and M. Singhal, 2007). However, (J. Liu and V. Issarny, 2007) identified significant differences between trust and reputation. The authors defined trust as the belief of a node in the trust attributes of a neighbour and described reputation as the vision that neighbours create for the node. Recommendations were used as a means for measuring the values of trust or reputation. The authors further classified trust management as monitoring based trust management and evidence based trust management. Monitoring based trust management measured the level of trust, on the basis of direct observations such as packets dropping and forwarding etc. as well as indirect observations such as recommendations from other nodes. Evidence based trust management identified the facts which proved the trust relations among nodes such as public key, identity, address or any other evidence that a node can obtain for itself or other nodes by means of a challenge/response process.

(F. Yunfang, 2007) suggested two approaches for trust evaluations named as Reputation Based Trust Management and Policy Based Trust Management. In Policy Based approach, the strong security mechanism e.g. logical rules were applied. The results of Policy Based Trust Management system obtained in the form of binary
decisions which stated whether the requester was trustworthy or not and access request was granted accordingly. Because of this binary nature of Policy Based Trust Management, this approach was considered as less flexible. On the contrary, Reputation Based Trust Management applied computational and numerical mechanism for trust evaluation. In that scheme, trust was computed by means of collection, aggregation, and propagation of reputation among the entities.

(J. Li, et al., 2008) classified trust management as trust establishment framework and reputation based framework. In trust establishment framework, the evaluation of neighbouring nodes was performed on the basis of direct observations. For the nodes with the absence of previous direct links, trust relations were created by using a set of opinions from intermediary nodes. On the other hand, a reputation based framework utilized direct observations and second hand data available in the network for evaluation of other nodes.

(Wenjia Li, et al., 2009) analyzed the behaviour of malicious and faulty nodes and observed that most of the conventional misbehavior detection approaches considered similar treatment for these, without any further investigation. The authors proposed a malicious node detection approach based on the use of policies which described context information e.g. communication buffer status, channel status, and transmission power level. This information was used to differentiate misbehavior performed by malicious intent or some faults.

(J. Luo, et al., 2009) proposed a trust model called RFSTrust, on the basis of fuzzy recommendation similarity. This was developed to calculate and evaluate the reliability of nodes. The authors applied theory of similarity, in order to calculate the recommendation relations among nodes. The higher the similarity degree between the recommending node and evaluating node, the more reliable was the process of evaluation between two nodes. The proposed model considered a scenario of presence of selfish attack only and did not test for any other attacks.

To raise the level of honesty in utilizing recommendations, (R. Li, et al., 2009) combined the confidence value with trust value in the evaluation to calculate trustworthiness of the node. The authors utilized the trust value to evaluate the recommendations where a recommending node having higher trust value possessed more priority.
(J. H. Cho, et al., 2009) proposed a trust management mechanism for group communication in MANETs. The authors used multiple trust metrics by combining QoS trust with social trust of the participating nodes. They considered honesty and confidence to indicate the social trust attributes, and energy as QoS trust attributes.

The problem of proposing a composite trust metric was addressed by (L. Yu, et al., 2010). The authors presented a trust model consisting of various decision factors. In the model, quality trust and security trust were combined to evaluate the trust value of the nodes in MANETs. In order to combine these two types of trust, Analytic Hierarchy Process (AHP) was used. To measure the security trust of nodes, the model integrated energy trust and transmitting trust and to evaluate the quality trust, it used value of delay trust.

A trust model based on human trust was proposed by (Pedro B. Velloso, et al., 2010). In the proposed model, the trust was calculated by using past experience of the nodes and recommendations collected from other nodes. For the purpose of sending and receiving recommendations from neighbours, a recommendation exchange protocol (REP) was used. A unique concept of calculating relationship maturity was introduced which indicated the time period for which the nodes were in relation to each other. Recommendations of nodes having higher value of relationship maturity were given highest priority. Thus relationship maturity was evaluated based on a single term which considered the period of relationship only.

(Wenjia Li, et al., 2010) defined a multi-dimensional trust management framework for the evaluation of the trust value of the nodes in a better way in MANETs. In contrast to the conventional trust management methods, the trustworthiness of the nodes was evaluated using different views. Each view of trust was obtained from different sets of misbehaviors depending on their nature.

(J. Cho, et al., 2011) provided a review of various schemes of trust management specifically developed for MANETs and discussed various potential attacks, classifications, trust metrics and performance metrics in MANETs. The authors highlighted different concepts of trust and derived some distinctive properties of trust representing social ideas of trust in MANETs.

(H. Yu, et al., 2011) proposed a clustering mechanism to differentiate trustworthy and untrustworthy recommendations. The method followed the majority law in which the
cluster having highest recommendations was taken as the most trustworthy. The authors evaluated their models by testing against two attacks e.g. ballot stuffing and bad mouthing.

To detect malicious nodes in MANETs, SVM classifier was used by the Automated Trust Management (ATM) system proposed by (Wenjia Li, et al., 2011). The ATM scheme was flexible enough to accept the challenges by malicious nodes to vary their pattern of misbehavior with time. Trust value for each and every node was detected by direct and indirect observation only, but malicious nodes were detected by SVM (Support Vendor Machine). SVM performed learning process like bridge learning. With the learning knowledge SVM detected the malicious node.

TRUNCMAN was a trust based routing method developed by (G. Thanigaivel, et al., 2012). The mechanism was used to filter out dishonest nodes during the process of path discovery, to protect from network layer attacks such as greyhole and blackhole attacks. The method was divided into two phases such as the Suspicion Phase and the Detection Phase. The first phase searched for the route request flooding and acknowledgement and the second phase provided explanation of the identification of non-cooperative nodes.

(Aravindh S, et al., 2013) described a trust mechanism to identify and separate the malicious node in mobile ad hoc networks. The authors used trustworthiness values to forward packets and maintained a value of trust counter for every node. If the value of trust counter fall below the trust threshold, then corresponding intermediary node was called as malicious and separated from the network which enhanced the network performance.

A multi level trust architecture based on context awareness in MANETs was proposed by (A. Rajesh and N. Mohan Kumar, 2014). In the proposed architecture, a trust model was enclosed with three models of trust defined as low-level, medium-level and high-level trust model that was based on context sensitive security method. The low-level trust model used direct observations to satisfy the essential security requirements. In addition, the average level of security was ensured in medium-level trust model by using direct observations as well as recommendations. The high-level trust model offered a very high secured system having more complexity and cost of computation.
A new trust-based information sharing model in MANETs was developed by (Khalid Zaman Bijon, et al., 2014). It was a recommendation-based trust management scheme, in which Dempster Shafer Theory was used to aggregate recommendations from various nodes. To reduce the number of recommendations, a new strategy buffering on-the-fly was introduced in which trust values were stored in intermediate nodes.

A distinctive approach using trust for mitigating misbehavior in MANETs was presented by (Vijayan R and Jeyanthi N., 2015). The authors claimed that a powerful network was created by taking the unique parameters such as security, quality of service, and mobility. On the basis of available energy, every node was assigned with trust values. The centralized system monitored the trustworthy nodes and misbehaving nodes to guarantee the exchange of certificate only to the trusted nodes.

(Vijayan R. and Jeyanthi N., 2016) surveyed various trust management approaches in MANETs to offer researchers with various perspectives on trust properties, metrics, and trust evaluation. This provided better understanding of trust management concepts such as trust propagation, aggregation, and prediction. The authors suggested some research directions such as composite trust metric which considered different facets of social networks and communications, and their respective trust management, trust distribution, and trust measurement.

In 2017, (S. Venkatesh Babu and C. Kezi Selva Vijila, 2017) surveyed various trust management techniques to provide security based on trust values. The trust values were computed through various methods by using the value obtained by direct or indirect methods. The authors also discussed some attacks to provide insight for designing of security algorithms.

2.3 Section-III

This section includes the research work carried on clustering in MANETs. In clustering, nodes with similar behavior grouped together to form clusters in network. There are two approaches of clustering defined as centralized and distributed algorithms. In centralized method of clustering, every node transmits its data to the central or gateway node, which then executes data clustering. In this approach, there is high communication overhead due to transmission of data of all the nodes. However, in a distributed clustering, each node is capable of generating clusters of mobile data vectors and then sends explicit parameters of clustered data to gateway.
node in order to reduce the communication complexity. The outliers are identified by measuring distance from the closest cluster.

(Jane Y. Yu and Peter H. J. Chong, 2005) classified clustering into six types on the basis of their main objectives as follows: Mobility-aware clustering, Dominating-Set-based (DS-based) clustering, energy efficient clustering, low maintenance clustering, combined metrics-based clustering and load-balancing clustering. The authors also partitioned the cost of clustering into five types such as the ripple effect of re-clustering, explicit control message transfer, constant computation round, the stationary assumption, and communication overhead.

(Wonchang Choi and Miae Woo, 2006) presented a distributed weighted clustering algorithm. The working of the proposed algorithm was similar to WCA (Weighted Clustering Algorithm). In the proposed method, the power management and distributed cluster set up was performed by localizing configuration and reconfiguration of clusters. The authors observed that the battery power was an important measurement than the aggregate time for which the mobile node worked as a CH. The lifetime of the topology might be increased by shifting the CH responsibilities to a normal node if the node had sufficiently large battery power.

(A. Abbasi and M. F. Younis, 2007) arranged catalog of essential attributes in three types such as Clustering process, cluster properties and CH features. The authors explained the complexity, features and the impact of the network model over existing schemes and also provided a summary of some significant schemes by defining their importance and weakness.

A novel technique of clustering based on mobility named as Mob-HiD was introduced by (Konstantopoulos, et al., 2008). The objective of the approach was to make prediction of the mobility patterns of the nodes to generate more stable clusters. A unique mobility prediction scheme was employed by the authors, which ensured the possibility of having same neighbours for a long time. Thus the mobility metric calculated the unpredictability of the neighbourhood of the node with respect to time. This measurement specified the appropriateness of the host for the selection of CH.

A comprehensive survey of various clustering methods was given by (Ratish Agarwal and Mahesh Motwani, 2009). The authors also highlighted the relevant issues with respect to clustering in MANETs, e.g. the energy utilization of mobile nodes, the
structure stability of cluster, the control overhead of constructing and maintaining the clusters, the equality of serving as cluster head for a mobile node and the distribution of traffic load in the clusters.

(Wonchang Choi and Miae Woo, 2010) illustrated an improved weighted clustering algorithm named as EWDCA (Efficient Weighted Distributed Clustering Algorithm). The proposed algorithm was mainly concerned with providing scalability in MANETs. The algorithm considered various parameters e.g. residual battery power, connectivity, distance between nodes and average mobility. With the help of these parameters, cluster head were selected which resulted into high scalability.

A novel clustering method named as NSLOC (Novel Stable and Low-maintenance Clustering) was given by (Conceicao, et al., 2010). The main objective of the proposed method was to provide high stability and low maintenance overhead. The algorithm is fully distributed in nature and did not depend upon cluster head in comparison to other clustering approaches.

In the environment of ad-hoc networks, different approaches of clustering had proposed that were based on various criteria. A comparative analysis of various clustering techniques was presented by (Sheetal Mehta, et al., 2011). The authors stated that every clustering algorithm must be implemented with low maintenance overhead to achieve high efficiency. The structure of network should not change with the movement of nodes.

(M. Anupama and B. Sathyanarayana, 2011) classified the clustering algorithms into five categories such as Weight Based Clustering, Identifier Neighbour Based Clustering, Mobility Based Clustering, Topology Based Clustering and Energy Based Clustering. The authors also evaluated their cost and performance by highlighting their importance and limitations.

A location based distributed clustering algorithm was introduced by (Kuldeep Sharma, et al., 2012). The authors used location metrics for cluster formation and divided cluster into layers to secure the ordinary nodes. It provided solution for two main issues of cluster stability and manageability. The algorithm relieved networks from the overhead due to broadcasting of control messages, particularly in static networks.

An energy efficient approach was described by (Yogendra Kumar Jain and Rakesh
Kumar Verma, 2012). The approach utilized node energy in an efficient way. The authors observed energy as the limited and the most significant resource in the mobile ad hoc networks. Idle consumption of energy had a great impact on the whole energy consumption in mobile networks. Hence the approach considered various parameters of energy such as receiving power, transmission power, sleep power and ideal power.

A Node Based Cluster Routing Algorithm (NBCRA) was produced by (Chaitali Uikey, 2013). The author developed a scheme for improving the stability of clusters which enhanced the network performance by choosing efficient cluster head. Each node in the proposed algorithm observed and measured its movements by itself and the obtained information was used to choose cluster head. In addition the proposed approach improved the CH stability.

(N. Gupta, et al., 2013) developed a weight based clustering approach which defined various system parameters as weight metrics. The aggregate value of weight was calculated by using the degree difference, battery power, mobility and transmission range. In addition, the proposed algorithm increased the trust value of cluster formation by removing the malicious nodes from clusters subsequently.

(Shivangi Singhal and A. K. Daniel, 2014) identified a cluster head selection algorithm derived from fuzzy logic. The authors used three parameters e.g. goodness factor, node degree and the competent level for selecting CH. The proposed algorithm reduced the re-construction time of CH and improved the network performance by minimizing the network overhead.

Cluster head choice considering neighbourhood commitment and normal least power was proposed by (S. Balaji and V. Priyadharsini, 2015). The parameters considered for weight calculation contributed to the stability of the clusters. By avoiding the nodes which had less number of neighbours and energy compared to the other nodes helped in reducing the overhead in selecting the cluster head.

(M. Ashwin, et al., 2016) defined a weighted clustering trust model in MANETs. In the model, clustering was followed by trust evaluation. Cluster heads were selected by combining various system variables such as mobility, ideal node degree, communicative power and battery life of the nodes. Trust election considered various security parameters such as packet misrouting ratio, packet dropping ratio, packet altering ratio and packet falsely injected ratio. If node’s trust value was more than the
relative threshold, then the node was assumed to be malicious. Therefore, the node was disqualified as a cluster member in the network. In the other case if a node’s trust value was found to be less as compared to the relative threshold, then the node was assumed to be non-malicious and was allowed to perform as cluster member node of the network.

(Abinaya. N and Ajikumar.K, 2017) classified clustering schemes into four types on the basis of their objectives and features such as Mobility Based Clustering, Energy Based Clustering, Trust Based Clustering and Topology Based Clustering. The authors described various objectives, characteristics, mechanisms and performance criteria in MANETs that were suggested by different clustering methods. The authors analyzed that the cluster formation and maintenance overhead, stability of cluster structure etc. were the significant matters to be paid attention by recent clustering methods.

2.4 Section-IV

This section includes reviewed papers which are not suited to above three sections but helped in understanding the objectives of this research. There are various Data Mining based security approaches in MANETs. Research study of misuse detection centered over classification of network intrusions by making use of a variety of standard Data Mining algorithms which are Cost Sensitive Models, Association Rules Models and Rare Class Predictive Models as defined by (Y Zhang and W Lee, 2000). A key benefit of misuse detection approach was found to be high level of accuracy in the detection of well-known attacks and their deviations. The weakness observed in such type of systems was the failure to identify novel attacks for which signature had not been defined.

(Aleksandar Lazarevic, et al., 2003) classified Data Mining based intrusion detection methods into two classes such as Anomaly Detection and Misuse Detection. In misuse detection, every instance of a data set was marked either as ‘ordinary’ or ‘intrusive’. The approach used learning algorithm which was trained using the labeled data. Both methods were capable of retraining intrusion detection systems using different input data which included new types of attacks.

The importance of using threshold cryptography in MANETs security was stated by (Saxena N, et al., 2007). The threshold cryptography aimed to issue trust to all nodes
such that the number of trusted nodes should not be smaller than a given threshold of trusted nodes. In such types of approaches, the Certification Authority (CA) issued certificates to the nodes entering in the network. In the proposed approach, a threshold signature protocol was applied by the authors to issue certificates.

To stimulate packet forwarding in self organized MANETs, an approach named SIP (Secure Incentive Protocol) was proposed by (Zhang Y, et al., 2007). Each node in SIP marked a print on every forwarded packet as an evidence of forwarding. On the basis of these, packets relays were rewarded and packet sources and destinations were remunerated with credits accordingly. The incentive schemes such as SIP must be effective and less complex as not to interrupt other regular functioning of the network e.g. routing etc.

In the last decades, researchers had considered various approaches such as Information Theory (Yan Lindsay Sun, et al., 2006), Game theory (Trinh Anh Tuan, 2006), Probabilistic Estimation (Jie Li, et al., 2008), Fuzzy theory (H. Dai, et al., 2009) and Grey theory(Liang Hong, et al., 2010) to design trust models. Grey theory was broadly deployed in different areas e.g. aerographs, environment science and agriculture. (Liang Hong, et al., 2010) described the basic idea of Grey theory, according to the authors, it observed the relations of various factors on the basis of similarity degree between samples of data. The main contribution of the method was that it did not demand a high quantity of data samples. In addition, the data consistency was not required for producing effective results, that were consistent with quality analysis.

(Farhan Abdel-Fattah, et al., 2010) proposed a novel intrusion detection method named CPDOD by using two anomaly methods such as Conformal Predictor k-nearest neighbour and Distance based Outlier Detection algorithm. A number of experimental results show that the proposed method was able to identify anomalies effectively with high detection rate and low false positive rate.

(J. Guo, et al., 2011) produced a new trust management framework having a large number of input factors to evaluate the trust values, thus making it more complicated for any malicious node to copy all of them. In addition, it used various weight vector sets to generate different trust values of a node. This helped in differentiating node’s behaviour as abnormal, as compared to other neighbour nodes. The approach was also
capable of finding out behaviour strategies of selfish nodes.

Honesty as a social term, was applied for evaluating the nodes behaviour. It was considered as an indicator of positive or negative behaviour. In the model proposed by (N. Vastardis and K. Yang, 2013), the honesty metric was utilized in a different way by joining it with confidence to form friendship relations among nodes. The authors evaluated the honesty degree of the evaluating node $i$ about the evaluated node $j$ based on the direct observation or recommendation by other nodes from the network.

(Wei, et al., 2014) used the probability theory to obtain direct trust values as a kind of uncertain reasoning. The authors used reasoning on the basis of uncertainty taken from Artificial Intelligence as a result of development in probability theory approach. In the context of security in MANETs, these theories claimed to be appropriate to evaluate trust based on interpretation of the trust in the research.

To implement and measure the performance of protocols different simulators are used. (M.Bheemalingaiah, et al., 2017) surveyed various simulators in MANETs such as Glomosim (Global Mobile Information System simulator), Network Simulator 2, NETSIM (Network Based Environment for Modeling and Simulation) etc. They also reviewed various mobility models available in MANETs and represented a hierarchical classification of different mobility models. The study provided by them was helpful for researchers in simulating the security algorithm in MANETs.

### 2.5 Outcome and Research Gap

On the basis of the literature reviewed, challenges faced by the outlier detection methodologies in MANETs are highlighted. In literature, different approaches to overcome the challenges/issues are also explored.

Outlier detection strategies that are commonly used in data mining applications are discussed in the literature. Out of them, some are useful in detection of anomalies in distributed networks like MANETs as given by (Prasanta Gogoi, et al., 2011). A trust based SVM (Support vector machine) is used by (Nirav J. Patel and Rutvij H. Jhaveri, 2015) to classify nodes in two different classes either regular or misbehaving nodes.

Literature study reveals that trust management has a great scope in finding out misbehaving nodes in MANETs. In literature, several models have been proposed for misbehavior detection using trust with different computation methods, such as
CONFIDANT (S. Buchegger and J. Y. Le Boudec, 2004), CORE (P. Michiardi and R. Molva, 2002) to evaluate trust among nodes. The aim of these models is to increase security in such environments by enabling nodes to evaluate their neighbours directly or indirectly through recommendations from other nodes in the network.

Outlier detection in MANETs is a broad term which aims for detection of anomalies, intrusions as well as misbehaviors. Intrusion detection, anomaly detection, misbehavior detection are the common mechanisms used in MANETs to provide security. Each of the mechanism protects nodes from specific threats in MANETs. There is very limited work done regarding outlier detection in MANETs. A Gossip Based Outlier Detection Framework was proposed by (Wenjia Li, et al., 2008) which uses trust as a base for identifying outlier nodes from the MANET. The framework provided a distributed approach to generate trust among nodes in MANETs which causes a lot of communication overhead. In literature there is a solution to overcome this problem by dividing the network into clusters. A study of clustering schemes and their issues are discussed in literature, which will help in generating “A Hybrid Approach for Outlier Detection using Data Mining Techniques in MANETs”

However, none or little work has been done in the field of outlier detection in MANETs security using data mining techniques. This has been identified as a research gap, and the work carried on, in the present research will try to fill up this gap.