AVAILABILITY ASPECTS THROUGH OPTIMIZATION TECHNIQUES BASED OUTLIER DETECTION MECHANISM IN WIRELESS AND MOBILE NETWORKS

Neeraj Chugh, Adarsh Kumar and Alok Aggarwal

School of Computer Science, University of Petroleum & Energy Studies, Dehradun, India

ABSTRACT

Radio Frequency IDentification (RFID) and Wireless Sensor Networks (WSN) are the two most prominent wireless technologies for implementing a complete smart environment for the Internet of Things (IoT). Both RFID and WSN are resource constraint devices, which forces us to go for lightweight cryptography for security purposes. Security in terms of confidentiality, integrity, authentication, authorization, and availability. Key management is one of the major constraints for resource constraint mobile sensor devices. This work is an extension of the work done by Kumar et al. using efficient error prediction and limit of agreement for anomaly score. This work ensures cryptographic property, availability, in RFID-WSN integrated network through outlier detection mechanism for 50 to 5000 nodes network. Through detection ratios and anomaly scores system is tested against outliers. The proposed outlier detection mechanism identifies the inliers and outliers through anomaly score for protection against Denial-of-Service (DoS) attack. Intruders can be detected in few milliseconds without giving any conflict to the access rights. In terms of throughput, a minimum improvement of 6.2% and a maximum of 219.9% is observed for the proposed protocol as compared to Kumar et al. Protocol and in terms of percentage of Packet Delivery Ratio (PDR), a minimum improvement of 8.9% and a maximum of 19.5% is observed for the proposed protocol as compared to Kumar et al. protocol.

KEYWORDS

WSN, MANET, RFID, ANOMALY, SECURITY

1. INTRODUCTION

Mobile Ad-Hoc Network (MANET) is a group of low powered computing mobile and wireless devices which co-operatively forms an infrastructure-less and decentralized network. There is an enormous use of sensor-based mobile devices in various applications like telemedicine, tele-geo processing appliances, vehicular networks, virtual navigation, military applications and household appliances. Compared to barcodes, magnetic tapes and smart cards, RFID devices are low cost with a high speed which makes them widely deployable nowadays. RFID networks consist of tags, readers, and backend storage devices. WSN consists of small, economical but low powered sensor devices with which the physical state of an object can be obtained. Identification, location tracking, and record management can be obtained from the RFID devices while the physical state of an object can be obtained from WSN. Temperature, sound, pressure, humidity could be few parameters for the physical state of an object. RFID and WSN technologies complement each other and integration of both can serve a lot in many fields like war field, animal tracking, supply chain management, and healthcare monitoring systems etc. Integration of both can be extended with the use of mobile sensor networks. A mobile RFID-WSN consists of smart mobile nodes which are constructed as an integration of mobile wireless sensor devices with RFID tag and

DOI: 10.5121/ijcnc.2018.10605

reader devices. RFID devices provide object identification, tracking & record management while mobile wireless sensor nodes provide the capability of sensing, mobility, ad-hoc, wireless and multi-hop communication. Effectiveness in terms of scalability, capability, and cost-effectiveness can be maximized with the integration of mobile sensor nodes with RFID tags. Both of these are resource constraint devices and require lightweight or ultra-lightweight cryptographic primitives for security. As a rule of thumb, approximately 30% of the total computational resources are available for cryptographic primitives. Confidentiality, Integrity, availability, authentication, and authorization are major security primitives for resource constraint mobile devices.

This work is an extension to the work done by Kumar *et al.*[1][2] Where cryptographic property, availability, is ensured for an RFID-WSN through outlier detection mechanism. Detection of unprecedented data identified from resource constraint mobile sensor devices is proposed. Through detection ratios and anomaly scores, system is tested against outliers. The proposed outlier detection mechanism identifies the inliers and outliers through anomaly score for protection against denial-of-service attack. Intruders can be detected in few milliseconds without giving any conflict to the access rights.

This paper is organized as follows. Section 2 gives details of the work done by earlier researchers and state-of-art in this area with rapid review of the work done by Kumar *et al.*[1][2]. Section 3 gives the detail about proposed approaches for ensuring availability, anomaly score error minimization and outlier detection. Section 4 presents the results and analysis. Section 5 gives the detail about the comparative analysis of the proposed protocol. Finally, Section 6 concludes the work with results and discussion.

2. LITERATURE REVIEW

Group key management is an efficient approach for user rights in a sensor based MANET where group keys are managed through different group key management protocols. These protocols are categorized as ID-based group key management; Diffie-Hellman based group key management and general group key management. In ID-based group key management, various protocols are developed to provide identification based non-repudiation [3-6]. In Diffie-Hellman based group key management, concentration has been on the reduction of communication steps and exponential calculations but these lack in providing proper authentication and non-repudiation. In General group key management protocols, concentration has been on enhancing the security level through session key, renewing procedures of session keys and non-repudiation through private identification marks like: Group Key Management Protocol (GKMP)[7-9], Group Data of Interpretation (GDoI)[10], Dunigan and Cao (DC)[8], Hao-Hua-Chu (HHC) [9],Group Secure Association Key Management Protocol (GSAKMP)[9], Burmester Desmedt Group Key Agreement (BD GKA)[10] etc.

Sensor-based ad-hoc networks consist of resource constraint devices and hence require light weight key management algorithm like in [2]for light weight devices, three group key management protocols Teo& Tan, WLH and Tseng's are identified and compared and it is found that Teo& Tan performs better in terms of delay, security and throughput compared to other two protocols.

After developing the group keys for users, permissions to access network information is controlled through access control mechanisms which ensure that the user and information interactions are authorized to enable data sharing. Level of access rights help to measure the significance of data sharing and mechanism like fine-grained access control is developed to clarify the controls. Fine-grained access control mechanisms can be classified as attribute based techniques, Identity based techniques, and Role-based techniques[1][2]. In [11], it is observed

that formal methods play an important role to check the mistakes in defining the policies that may arise due to expressiveness property of policies. In [12], authors have developed a Margrave tool to check the user-specified properties of a policy. Alloy and Margrave help to check duty constraints, roles, absence or presence, permission and behavioral response from policy members [13-15].

Outliers are the deviations of data from its regular behavior which can be classified on different categories: (i) Node & Network-based, (ii) Nearest neighbor based, (iii) Statistical based mechanisms, (iv) Bayesian network based, (v), Local, Global & Semi-global based (vi) Spectral decomposition based, (vii) Error, event or attack based, (viii) Supervised & Unsupervised based, (ix) Distance, density, machine learning or soft computing based etc. [16-20]. Security through outlier detection mechanism is one of the major advantage for such networks. There is a need to use lightweight mechanism for finding an error in sensor-based ad-hoc networks like Traag*et al.* proposed a Markov chain based technique for making a distinction between an event or error for mobile phones [21].

There are various sources of outlier or anomaly in a sensor network like the uncertainty of data [22-23] deviation from regular system pattern for security compromise [24-25], hardware/software [26], environment [27], etc. Anomalies can occur at any level like node, data or network level [28-29]. Various outlier detection techniques have been identified like nearest neighbour based, clustering based outlier [19], distance-based outlier [30-31] node, data or network-based outlier [16], neural network/fuzzy based/Markov/Hidden Markov based outlier [18], statistical/knowledge-based outlier [17], density based outlier [32-34] local/global/semi-global/distributed global/semi-global distributed outlier [35]. Ayadi*et al.* [36] identified that WSN can be affected by many anomalies which occur due to software or hardware problems. So various protocols are developed in order to detect and localize faults then distinguish the faulty node from the right one. Sensor nodes can produce erroneous measurements due to number of reasons as battery depletion, damage of device and other causes. To address the problem of outlier detection in WSN, Titouna *et al.*[37] developed a two-level sensor fusion-based outlier detection technique for WSN.

2.1. Review of work done in [1, 2]

Kumar, Agarwal & Charu (2012) have proposed an extension of Teo&Tan's [38]circular hierarchical model for a fixed number of group members. For large Ad Hoc Networks in 2005, J. C. M. Teo and C. H. Tan proposed an energy efficient key agreement protocol which performs key computation at two stages: local (inside a subgroup) and global (between subgroups) levels. At the local level, this protocol uses the Burmester Desmedt Group Key Agreement (BD GKA) protocol and at the global level, encryption/decryption process is used for computing and exchanging the keys. Teo& Tan's protocol is an efficient key management protocol for the hierarchical network. However, it suffers from an exponential increase of key messages due to dynamic topology and energy loss because the vicinity of nodes in a subgroup is high. Kumar *et al.*[2] has made changes in two categories: change in the mode of communication and the addition of virtual programmable nodes. These changes are explained in section 2.1.1 and 2.1.2.

2.1.1. Change in modes of communication

Various modes of communication are integrated with local subgroup of Teo& Tan based hierarchical network construction protocol. These modes reduce the cost of network construction and are explained as follows:

Method 1 (Serial communication rather than broadcast)

Using Dijkstra's algorithm [39], key messages are transmitted serially with minimum distance. In a local subgroup, a count is maintained for estimating the number of contributions in a subgroup and the last contributor is considered to be the subgroup controller which computes and distributes the subgroup key among subgroup members and to parent subgroup controller for generating a common hierarchical key. At a hierarchical level, the shortest path is found for distributing the key. Results show that the number of messages and exponentiation operations reduces from layer 2 onward with the major strength of the proposed approach is in the reduction of overhead on single subgroup controller of broadcasting. But the proposed approach suffers from count to infinity problem in the network.

Method 2 (Circular communication)

Key messages in a subgroup are communicated in both clockwise and counterclockwise directions. Spinrad algorithm [40] is used for forming the logical circle of mobile nodes. The key messages also carry a timestamp and a token so as to ensure the freshness and to remove the duplicity of the message on any node. Results show that communication cost got reduced if destination node immediately responds to all its subgroup nodes with a lightweight message K' where size (K') < size (K) and size (K') could be 1 bit in its best case. Another major strength of this protocol is better backward secrecy by the use of timestamp along with the use of token which avoids the duplicity of messages in the network. Despite this strength, the proposed mechanism still suffers from complete avoidance of count to infinity problem. Further, all nodes should be in transmit or receive state. If any node is in IDLE or SLEEP state then predecessor and successor nodes have to wait for activation of that node.

Method 3 (Ant's colony communication)

Ant's algorithm [41] is used in this modification which avoids count to infinity problem by leaving a sign on the node which has been covered in key message collection. Strengths of this proposed modification are: (i) Count to infinity problem can be avoided by leaving a sign of covering the node for key message collection, (ii) the unnecessary key messages are reduced in a subgroup compared to serial or circular communication, (iii) unnecessary messages in the network are reduced which results in the reduction of the delay in key establishment. Despite these strengths, this modification is unable to remove waiting delay caused due to SLEEP or IDLE state of an intermediate node, if any node is in SLEEP or IDLE state then it has to re-run the Ant's algorithm which increases the cost of communication also.

Method 4 (Using Shamir's Threshold secret sharing communication)

This modification reduces the waiting delay because of IDLE or SLEEP state of a node for which Shamir's threshold secret sharing scheme is integrated with Teo and Tan's protocol[42]. A combiner function is used to collect the shares of 'm' mobile nodes out of 'n' nodes [43] which solves the problem of deactivation of nodes and whenever any SLEEP or IDLE state mobile node starts transmitting and receiving the data then it can either use the established key or send its contribution in next update cycle of subgroup key generation. Since $m \leq n$ thus cost of message communication will be lesser as compared to Teo and Tan's protocol.

2.1.2. Addition of Virtual Programmable Nodes

Teo and Tan's protocol is extended toa variable number of group members. According to Teo and Tan protocol, a number of nodes in a subgroup are fixed in a hierarchy but if a node is not in the close vicinity then either it has to wait for a node to come closer or enforce a distant node to

become part of subgroup which results in a large amount of energy loss. The proposed modification removes the energy losses by adding virtual nodes in a subgroup which are not the real nodes but programmable node formed by a subgroup controller. Subgroup controller fixes the number of nodes in a subgroup that may contain virtual or real nodes, virtual nodes will do the same job as of existing nodes but have a difference in the joining and leaving phases to refresh a key for avoiding attacks. Virtual programmable nodes run peer to peer process with new member supposed to join a subgroup. A virtual node shall leave the space for a real node if it latter joins the network similarly, whenever an existing real node leaves a subgroup then that space is occupied by a virtual node.

2.1.3. Cost & Security Analysis

The cost comparison shows that the proposed protocol inside a subgroup is having lesser cost in terms of number of rounds required to generate a key, number of messages, number of messages sent per participant, number of messages received per participant, exponentiations per participant and total exponentiations in a subgroup. Results for Gates Equivalents (GEs) show that the proposed key management scheme uses exponentiation and multiplication operations only which requires 10113 GEs compared to the Teo and Tan protocol which uses exponentiation, multiplication and addition operations in key managements requiring 10829 GEs giving an improvement of 6.6% of GEs in key management. Three security levels are used for comparative analysis: Weak, Moderate and Strong. The proposed mechanism is having encryption/decryption, hashing and computational challenges at both subgroup and hierarchical level. Since the proposed protocol is an extension to Teo and Tan protocol hence it overcomes the weakness of this protocol and provides a strong level of security.

3. PROPOSED APPROACH

In this section, proposed approaches for ensuring availability, anomaly score error minimization and outlier detection are presented.

3.1. Proposed Approach for ensuring availability

In this section, availability property is ensured through outlier detection mechanism. A node is considered to be an outlier if it outperforms and underperforms beyond threshold limits. In order to identify these nodes, all nodes are categorized into the following categories:

Source node, a node is considered as source node if it initiates sending control or data packet.

An intermediate node, a node is considered as an intermediate node if it receives and forwards route request or data packet.

Destination node, a node is considered as destination node if it receives route request or data packets and acknowledges either through route reply or receipt of data.

Sleep node, a node is considered as a sleep node if it does not send or receive any data but sends a control packet after regular intervals in order to mark its presence. This is a power saving state and it is known as a suspended or standby state.

Idle node, a node is considered as an idle node if it remains in a state of inactivity but consumes more power compared to sleep node. This node sends control packets more frequently compared to sleep node.

Dead node, a node is considered as a dead node if it suddenly stops regular node activities.

Live node, a node is considered as a live node if it has recently joined the network and has full energy level but does not start communicating with any other node.

In this work, above-defined nodes are categorized into four categories: active, passive, dead and live. Source, destination, and intermediate are considered as active nodes since it sends control packets more frequently compared to any other node type. Whereas, passive nodes are misbehaving nodes, which consume resources by sending a large number of control packets, which includes Preamble and Physical Layer Convergence Protocol (PLCP) header, the aim of which is to check the signal strength and find the route to the destination. Although passive nodes send the control packets for finding the route but these nodes do not forward the data on these routes, as a result of which a large number of control packets consume the network resources and block the services of the network. In this work, outliers among passive nodes are identified using the anomaly score. Kumar *et al.* [1] anomaly score calculation is modified as shown inequ.1 and equ. 2.

$$AnomalyScore = \frac{\left(Total(Active) - \left(Average(ActiveandPassive)\right)\right)}{STDEV(Active)} \pm STDERR \tag{1}$$

$$STDERR = \sqrt{s^2/n} \tag{2}$$

Where, 's' is the standard deviation of the differences and '*n*' is the sample size considered for outlier detection.

Further, Limits of Agreement (LoA) is considered for anomaly score. It is the estimation of the interval value within which the proportion of the differences between two measurements (Anomaly Score + STDERR and Anomaly Score–STDERR) lies. Limit of the agreement is calculated as shown in equ. (3).

$$LoA = SystematicError(SE) + RandomError(SE) \tag{3}$$

In the proposed outlier detection approach, $SystematicError(SE)$ occurs when there is reuse of experimental setup in same way or in same case. In this approach, SE is used to pick the interval for anomaly calculation. As shown in figure 1, average systematic error in number of outliers and anomaly score variation is minimum when anomaly calculation interval is 100. Thus, an interval of 100 msec. is selected for anomaly score calculation. In equ. (4),$SystematicError(SE)$ is classified as error in number of outliers and error in anomaly score variation. Both types of SE includes parameters like: Imperfect Calibration Error (ICE), Quantity Error (QE) and Drift Error (DE) as shown in equ. 4.

$$SystematicError(SE) = \text{ICE} + \text{QE} + \text{DE} \tag{4}$$

In this work, ICE includes factors like an error due to environment conditions (temperature, pressure etc.), the distance between nodes, mathematical models used for experimentation and physical law for node movement computations. QE includes constant value added to systematic error because of increase in the number of nodes. As expected, increase in the number of nodes in a close vicinity increases overhead upon existing infrastructure, which further increases, the chances of error. It is assumed constant in this experimentation because each node adds a fixed initial overhead over network and error varies with the increase in number of nodes. DE includes variation in trends for calculating constant quantities in experimentation, which may drift to one way or another. For example, if the total number of packets sent from nodes is not equal to the number of times, nodes act as source node then resetting the experiment is a better

option for the accuracy of results.$RandomError(SE)$ occurs due to controllable factors like: sudden failure of node, nodes under black hole attack, expensive control over node behavior etc. In standard error calculation, $STDERR$-2*s and $STDERR$+2*s is acceptable standard error limits of variation and it is approximately calculated as: $\sqrt{4 * s^2/n}$ with 90% confidence intervals, n-1 degree of freedom.

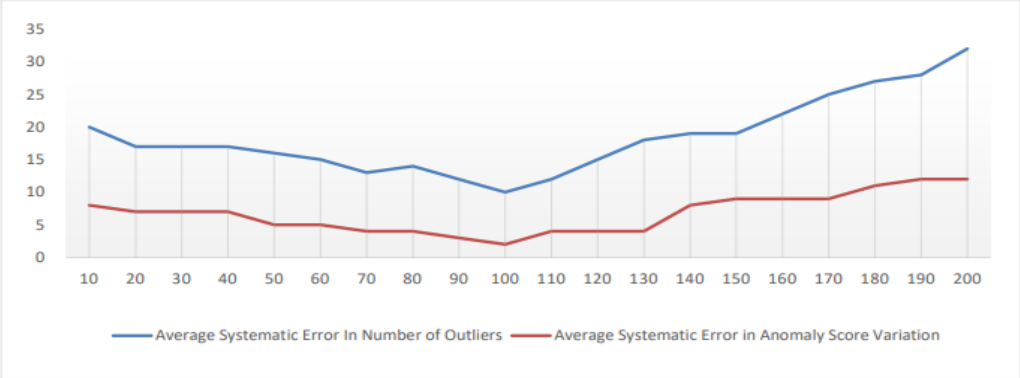

Figure 1. Variation in Systematic Error Calculation with an increase in time interval

3.2. Proposed anomaly score error minimization

In this work, the anomaly score is accepted with minimumLoA. In order to compute minimumLoA, simulatedannealing process selects anomaly score with efficient error probability as explained in pseudocode 1. In this process, best neighbouring anomaly score value is selected by selecting the minimum error different probability between actual and predicted results. In conclusion, a neighbouring point (anomaly score) is selected when probability of error in actual result is lesser than predicted result.

Pseudocode 1: Anomaly Score Error Minimization using Simulated Annealing

Premises: Let ω is the error between actual and predicted anomaly score.$\omega_{current}$represents current error value from whole set of errors, Ω represents change in current $\omega_{current}$ and previous $\omega_{previous}$ error values. Further,$\alpha$is portion of difference between minimum and maximum value of anomaly score variation.

Goal: Minimize the error between predicted and observed data

1. Error_list=NULL
2. **While** predicted_iterations≤ max_iterations **do**:

3. Compute Anomaly Score as:

$$AnomalyScore(Actual) = \frac{\big(Total(Active) - \big(Average(Active \wedge Passive)\big)\big)}{STDEV(Active)} \pm STDERR$$

4. Computer Predicted Anomaly Score as:

$$AnomalyScore(Predicted) = \frac{\big(Total(Active) - \big(Average(Active \wedge Passive)\big)\big)}{STDEV(Active)}$$

5. Computer Error (ω= $AnomalyScore(Actual) - AnomalyScore(Predicted)$

6. Error_list=Error_list.append(ω)
7. predicted_iterations = predicted_iterations +1
8. **End-while**
9. **For**$\omega_{current}$**in** Error_list**:**
10. Compute Ω=$\omega_{previous} - \omega_{current}$
11. **If** Ω= 0**then**
12. Reject$\omega_{current}$ and accept neighbor Anomaly score value
13. **Else**
14. Accept anomaly score with minimum $\omega_{current}$ value with probability ($\omega_{current}$/anomaly score)
15. **End-for**
16. Computer Anomaly Score (final)= anomaly score*α, Here, α is portion of difference between minimum and maximum value of anomaly score variation in n-experimentations and $0 < \alpha < 1$

3.3. Proposed outlier detection approach using an anomaly score

Pseudocode 2 explains the outlier detection process through anomaly score calculation. In this process, total simulation time is divided into n-slots and outlier nodes are identified after the second slot. These outlier nodes are identified through a combination of active and passive nodes. Those passive nodes are considered as misbehaving and outlier nodes that are sending unwanted messages beyond a threshold limit. The proposed outlier detection process is a continuous process of identifying outlier nodes and stopping them from participating in network activities thereafter. A common set of outliers among all outlier lists are sent in the trained dataset for feature extraction and comparison with test dataset.

Pseudocode 2: Outlier Detection through Anomaly Score

Premises: S, D, I, SL, DN, LN, IN, U, A and P represents source, destination, intermediate, sleep, dead, live, idle, undefined, active and passive node. A threshold anomaly score value is decided at implementation time and it varies from experiment to experiment.

Goal: Divide the simulation time among n-slots, compute anomaly score after each slot starting from second slot onward and count number of outliers and inliers after every slot using anomaly score

1. Compute list of active nodes as: A= [S, D, I] and passive nodes as: P= [U].
2. Divide total simulation time into *n*-slots
3. Set *iteration* =1
4. Outlier_list=NULL
5. **While**$iteration \leq n$**do**:
6. **If***iteration* ==1 **then**
7. Compute total number of A and A+P nodes
8. **Else-if***iteration*≥1 **then**
9. **Compute** Total number of A, average of A from (*iteration*-1)th and *iteration*thslots, total number of A and P nodes, average of number of A+P nodes, average of number of A+P nodes from (*iteration*-1)th and *iteration*thslots, standard deviation (STDEV) of A using (*iteration*-1)th and *iteration*thslots and anomaly score
10. Plot anomaly score of each node
11. Put those nodes in *Outlier_list* which are having anomaly score greater than a *threshold* value

12. **For***node***in***Outlier_list*
13. Discard *node* for communicating in a network
14. **End-for**
15. *Iteration=iteration+1*
16. **End-while**
17. Put nodes common in all Outlier_list in trained dataset with their features

4. RESULTS AND ANALYSIS

4.1. Simulation setup

Simulation analysis of 50 to 5000 nodes, distributed randomly over 1000 m x 1000 m area, is performed for analyzing proposed protocol. Table 1 shows the other simulation parameters taken for analysis. Table 2 shows the datasets considered for outlier detection. Eight datasets with a different number of nodes and traces are considered for analysis. Table 2 also shows the number of clusters considered at different time slots for outlier identification. The number of clusters is increasing withan increase in the number of nodes or simulation time as more number of nodes is participating in group activities.

Table 1. Simulation Parameters

Parameters	Value
Number of nodes	50 to 5000
Channel Type	WirelessChannel
Radio Propagation Model	Ray Tracing
Network Interface	WirelessPhy
MAC Type	802.11
Interface Queue	Priority Queue
Antenna	OmniAntenna
Max Packets in Queue	50
X dimension of the topography	1000 meters
Y dimension of the topography	1000 meters
Mobility Model	Random WayPoint Mobility
Data Rates	5 packets/second
Packet Size	512 bits
Simulator	ns-2 [44]
Simulation Time	2000sec
Number of slots assigned to reader at stretch (Δ)	1
Time of each slot	10 msec.
Velocity (Minimum to Maximum)	0.3 m/s to 5 m/s

4.2. Outlier detection approach

In the simulation, 50 to 5000 nodes are randomly deployed in a geographical region for 2000 seconds. Simulation time of 2000 seconds is divided into ten equal slots each of 200 seconds for observation. Outliers are identified after the second slot i.e. 200 seconds using anomaly score. Table 2 and table 3 shows an example of outlier detection for 1000 nodes.

Table 2. Datasets and Clusters

Sr. No.	Dataset	No. of Traces	No. of clusters (during different time slots)									
			upto T_1	T_1 upto T_2	T_2 upto T_3	T_3 upto T_4	T_4upto T_5	T_5 upto T_6	T_6 upto T_7	T_7 upto T_8	T_8 upto T_9	T_9 upto T_{10}
1	50-nodes	24312	3	5	13	18	26	30	32	32	32	33
2	100-nodes	46714	5	8	10	16	22	24	36	36	37	37
3	500-nodes	65920	4	5	9	11	16	26	31	38	39	39
4	1000-nodes	97441	6	8	9	14	17	22	30	34	43	40
5	2000-nodes	117237	4	8	13	18	20	26	31	34	37	43
6	3000-nodes	142652	4	9	14	21	22	34	39	39	38	44
7	4000-nodes	186121	5	15	17	24	30	32	40	47	50	51
8	5000-nodes	212618	4	19	27	28	45	46	47	48	52	55

T_1=200 sec., T_2=400 sec., T_3=600 sec., T_4=800 sec., T_5=1000 sec., T_6=1200 sec., T_7=1400 sec., T_8=1600 sec., T_9=1800 sec., T_{10}=2000 sec.
*No. of clusters calculation is the average value of five executions
**In each time slot, the maximum value of the number of clusters is considered for analysi

Table 3. Counts of Active or Passive State of a Node in the First Slot

Node	S(A)	D(A)	I(A)	SL	U(P)	DN	LN	IN	Total (A)	Total (A+P)
N1	17	17	28	18	15	2	2	1	62	77
N2	21	16	15	20	22	2	1	1	52	76
..	..	..	..	..	..				..	..
N1000	22	20	20	12	22	1	2	1	65	84

*S: Source, D: Destination, I: Intermediate, SL: Sleep, DN: Dead node, LN: Live Node, IN Idle Node, U:Undefined, A: Active, P: Passive

The active presence of nodes in the current slot using anomaly score is shown in equ. 1. A node is considered as an outlier if it shows the much low value of anomaly score which results in a proportionate low-value presence in the active state. Anomaly score calculations for each node after 400 seconds is shown in Table 4. Second slot onwards, total anomalies are calculated after every 200 seconds. Total anomalies after completion of the second slot are 97 as shown in Table 5. The plot of anomaly score of 1000 nodes is shown in Fig. 2(d). Out of these 1000 nodes, 97 nodes are the outlier nodes based on anomaly score which means that these nodes are sending a large number of control packets for unnecessary consumption of network resources and denial of service attack. Similarly, the plot of anomaly score of 50 to 5000 nodes are shown in Fig. 2. Outliers identified in 50, 100, 500, 1000, 2000, 3000, 4000 and 5000 nodes network are 18, 28, 57, 97, 187, 247, 297 and 356 respectively.

Table 4. Counts of Active and Passive State of a Node and Outliers in Second Slot

Node	S (A)	D (A)	I (A)	SL	U (P)	DN	LN	IN	Total (A)	Average(A) Slot1 and Slot 2	Total (A+P)	Average (A+P) Slot1 and Slot2	STDEV (A)	Anomaly Score
N1	20	20	24	14	17	2	1	2	66	64	83	82.50	2	-7.75
N2	26	17	15	20	17	2	2	1	61	57.5	78	77	3.5	-3.27
N3	25	18	12	22	16	2	2	1	55	54.5	71	77	3.5	-5.25
N4	26	17	15	20	17	2	2	1	61	57.5	78	77	3.5	-4.37
N5	24	19	15	18	19	2	2	1	61	57.5	76	77	3.5	-3.16
N6	27	14	15	20	17	2	2	1	61	57.5	78	77	3.5	-3.47
..														
N1000	18	15	20	18	25	2	1	1	57	61.5	82	85	4.5	-5.18

Table 5. Total Anomalies calculations after second slot

Node	Anomaly Score
N1	-7.75
N2	-3.27
..	
N1000	-5.18
Average	-7.56
STDEV	6.1
Total Anomalies	97

Inliers and outliers in subgroups are identified after identifying the outlier nodes in the hierarchical network. Fig. 3 shows the randomly deployed nodes on their (x, y) locations. These randomly deployed nodes form groups or subgroups based on their close vicinity as shown in Fig. 4. Two processes of identifying the inliers and outliers in a network are shown in Fig. 5. In process 1, inliers and outliers are identified from each subgroup of a set of nodes in close vicinity. In other process, inliers and outliers are identified from the whole set of subgroups that are constructed based on close vicinity of their nodes. In this work, both processes are used for identifying the inliers and outliers among groups or subgroups in a hierarchical network. Figure 6 shows the hierarchical clustering dendrogram formation for 500 nodes using Python and sci-kit learn [45]. Number of clusters formed using hierarchical clustering dendrogram formation are shown in Table 2.

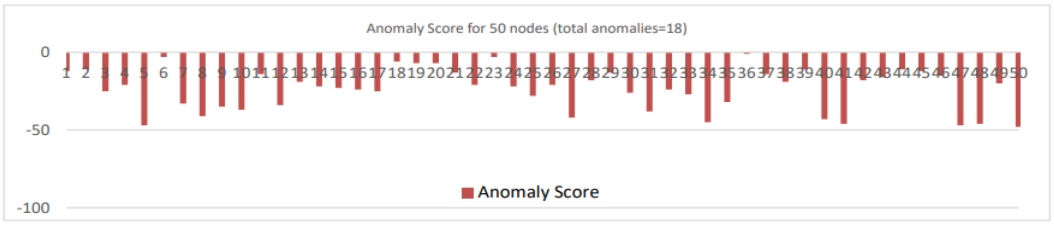

Figure 2 (a).Anomaly score for 50-nodes

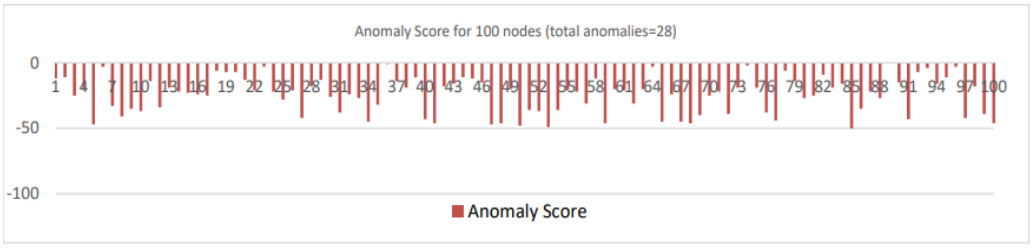

Figure 2 (b). Anomaly score for 100-nodes

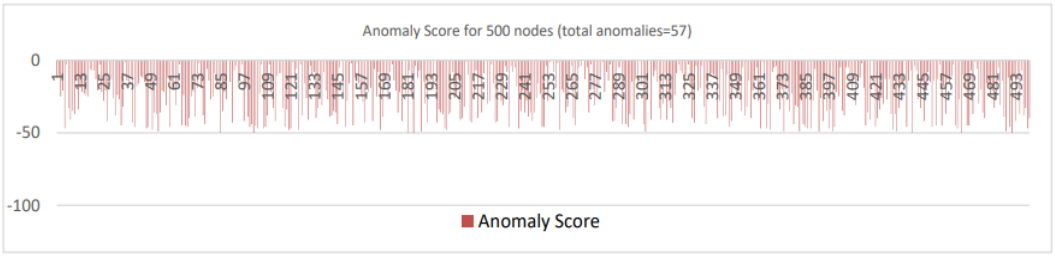

Figure 2 (c). Anomaly score for 500-nodes

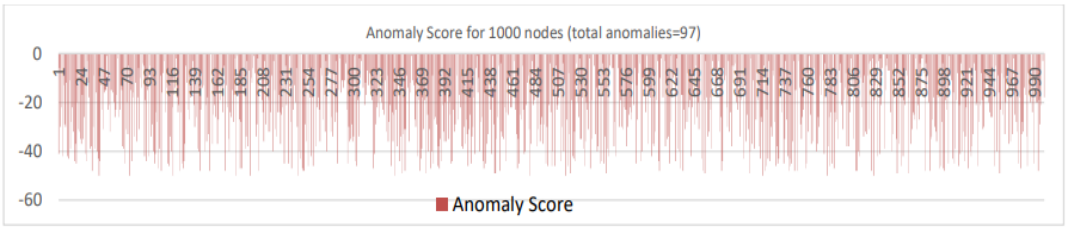

Figure 2 (d). Anomaly score for 1000-nodes

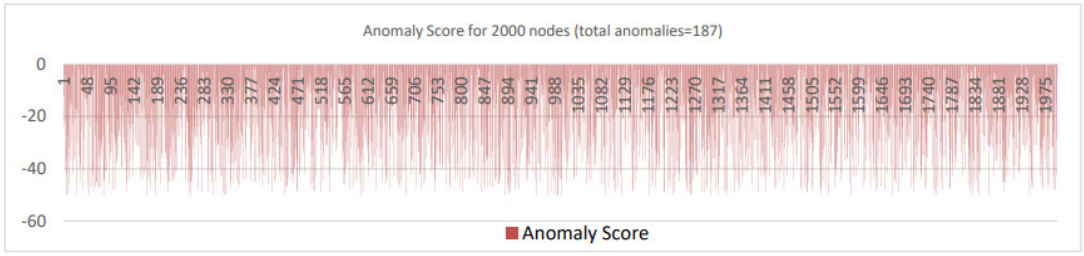

Figure 2 (e). Anomaly score for 2000-nodes

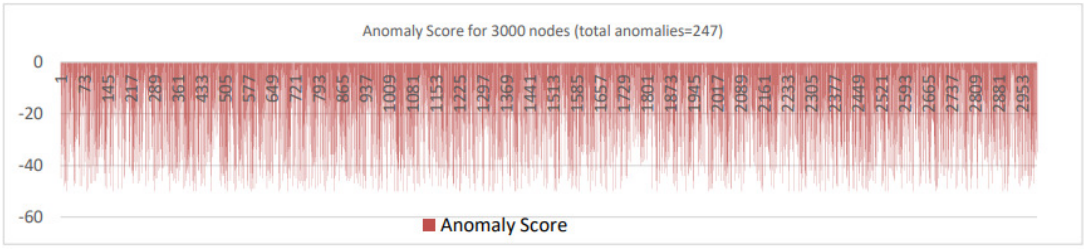

Figure 2 (f). Anomaly score for 3000-nodes

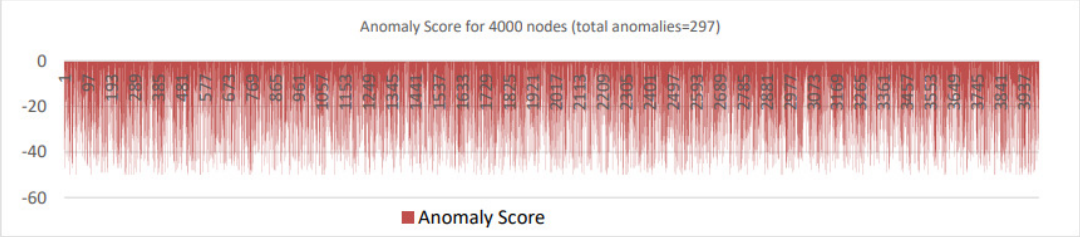

Figure 2 (g). Anomaly score for 4000-nodes

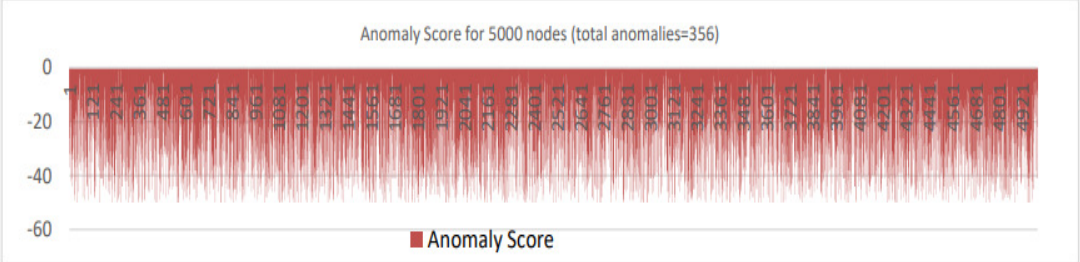

Figure 2 (h). Anomaly score for 5000-nodes

Figure 2**.** Anomaly score variations for 50 to 5000 nodes networks

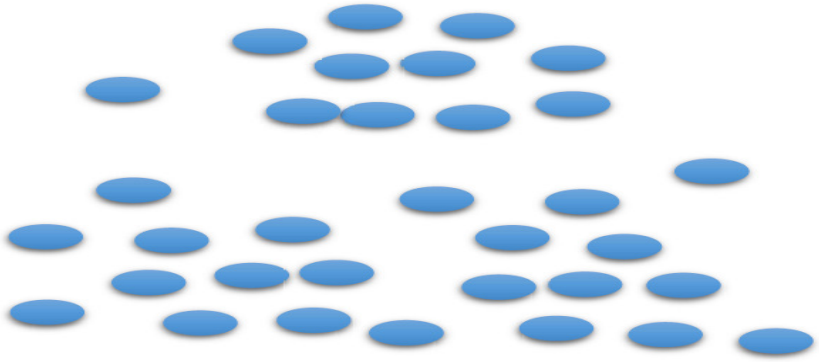

Figure 3. Randomly deployed nodes at their (x,y) locations

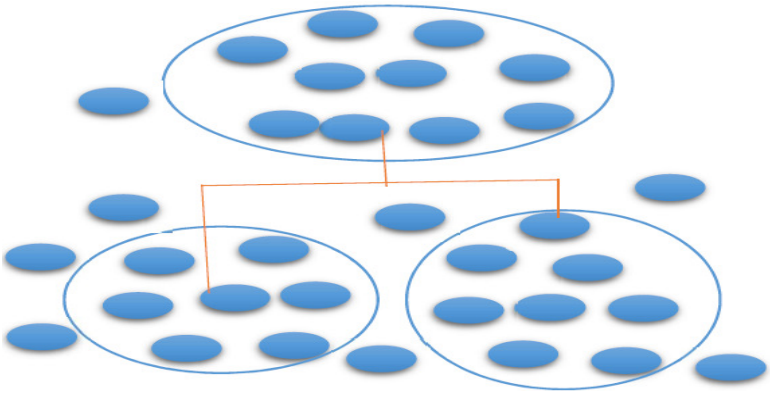

Figure 4. Group of closely randomly deployed nodes at their (x,y) locations and group connections in an hierarchy

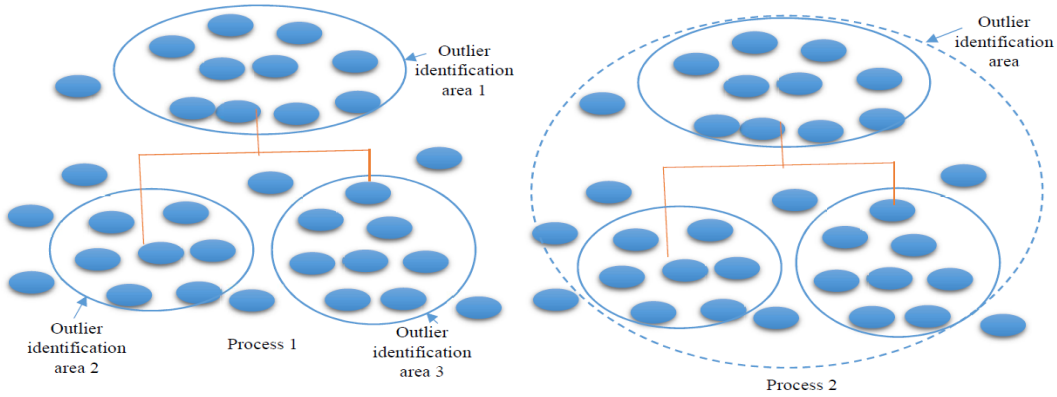

Figure 5.Two processes of identifying inliers and outliers within the group or subgroup

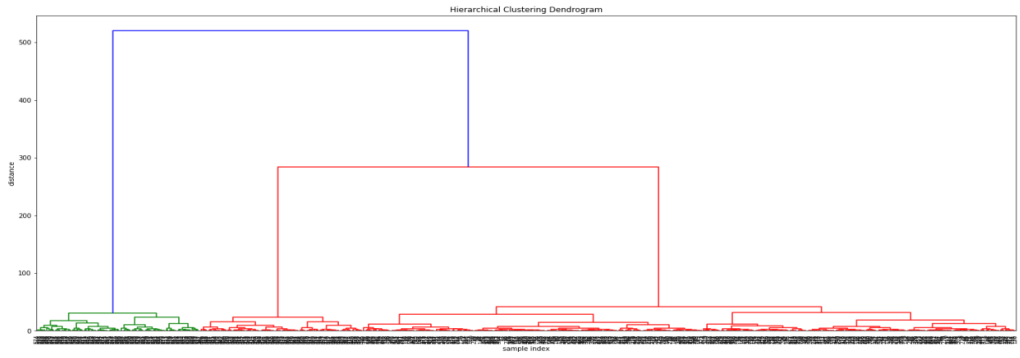

Figure 6. Hierarchical clustering dendrogram for 500 nodes

5. COMPARATIVE ANALYSIS OF PROPOSED PROTOCOL

Fig. 7 and Fig. 8 shows a comparative analysis of throughput and packet delivery rate (PDR) for 50 to 5000 nodes with and without the presence of outliers. Figure 9 shows throughput analysis and it is observed that throughput increases with increase in the number of nodes as more number of nodes are participating in network activities. For the proposed network scenario, throughput without outliers lies within threshold limits whereas throughout with outliers is below the lower acceptable threshold limit. Hence, if any group of nodes is showing throughput below lower threshold limit then it is considered as an outlier group. Similarly, comparative analysis of the percentage of packet delivery rate (PDR) is shown in Fig. 8 and Fig. 10. Results show that the percentage of PDR without outlier lies between threshold limits whereas it is below the lower threshold limit for a network with outliers. Thus, if any group of nodes shows percentage PDR below lower threshold limit then those nodes are considered as outliers. Fig. 9 shows comparative throughput analysis of the proposed protocol with Kumar *et al.* protocol [1]. Results show that the proposed protocol is having better throughput compared to Kumar *et al.* protocol [46-53] because of connected hierarchical network construction and consideration of small-scale (50 to 500 nodes) to large scale (3000 to 5000 nodes) networks. Fig. 10 shows the comparative percentage of PDR analysis for proposed protocol with Kumar *et al.* protocol [1-2]. Results show that the proposed protocol is having a better percentage of PDR compared to Kumar *et al.* protocol because of connected hierarchical network construction and refined outlier detection process.

Figure 7. Comparative analysis of throughput comparison with and without presence of outlier nodes

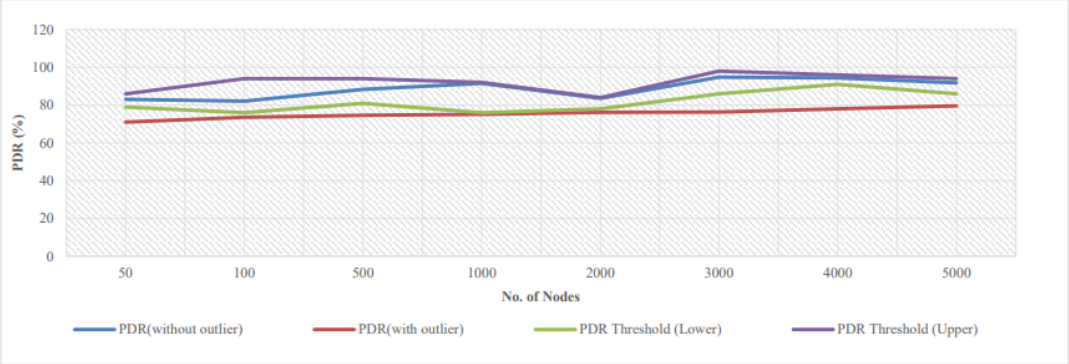

Figure 8. Comparative analysis of packet delivery rate (PDR) with and without presence of outlier nodes

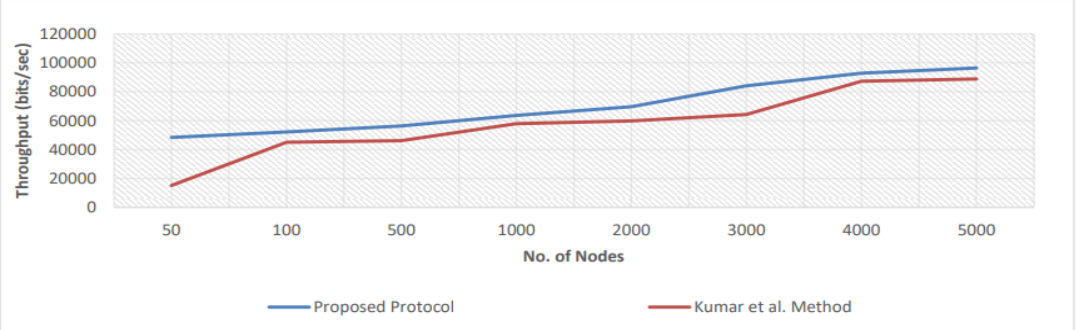

Figure 9. Throughput comparative analysis of proposed protocol with the Kumar *et al.* method

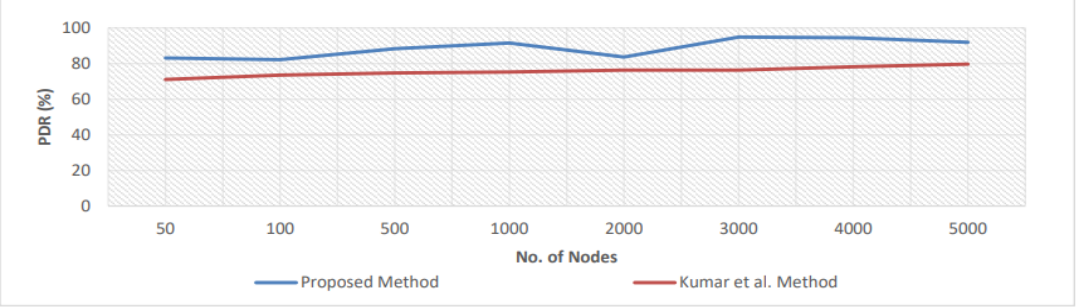

Figure 10. PDR comparative analysis of proposed protocol with the Kumar *et al.* method

6. CONCLUSION

In this work, cryptographic property, availability, is ensured through outlier detection mechanism. A node is called an outlier if it deviates from its regular behavior. Nodes are classified as active, passive, dead and live. Networks of 50 to 5000 nodes are used for analysis and it is found that 50, 100, 500, 1000, 2000, 3000, 4000 and 5000 nodes network consists of 18, 28, 57, 97, 187, 247, 297 and 356 outliers respectively, which means that these nodes are sending a large number of control packets for unnecessary consumption of network resources and denial of service attack. Though detection ratios and anomaly scores system is tested against outliers. The proposed outlier detection mechanism identifies the inliers and outliers through anomaly score for protection against denial of service attack. Intruders can be detected in few milliseconds without giving any conflict to the access rights. Comparative analysis of throughput and percentage of PDR shows that the performance of the network improves after detection of outliers and lies within threshold limits. Further, upper and lower threshold limits are computed for identifying outliers in subgroups using threshold-based outlier detection mechanism. Comparative analysis of throughput and percentage of PDR shows that the proposed protocol is better than Kumar *et al.* protocol. In terms of throughput, a minimum improvement of 6.2% and a maximum of 219.9% is observed for the proposed protocol compared to Kumar *et al.* protocol. In terms of percentage of PDR, a minimum improvement of 8.9% and a maximum of 19.5% is observed for the proposed protocol compared to Kumar *et al.* protocol.

REFERENCES

[1] A. Kumar, K. Gopal and A. Aggarwal, "Outlier Detection and Treatment for Lightweight Mobile Ad Hoc Networks," in In International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, Greater Noida,India,11-12 January 2013, pp.750-763.

[2] A. Kumar, A. Agarwal and Charu, "Efficient Hierarchical Threshold Symmetric Group Key Management Protocol for Mobile Ad Hoc Networks,"Inter. Conf. on Contemporary Computing – IC3 2012, Noida,India,6-8 August 2012, pp. 335-346.

[3] D. Bonch and M. Franklin, "Identity-based encryption from weil pairing," Advances in Cryptology-Crypto 2001,Santa Barbara, California, USA,August 2001, pp. 213-229.

[4] Merwe, J. V. D., D. D. and M. S, " A survey on peer-to-peer key management for mobile ad hoc networks," ACM computing surveys (CSUR), Vol. 39, No. 1, 1, April 2007, pp. 1-45.

[5] H. Deng, A. Mukherjee and D. Aggarwal, "Threshold and identity based key management and authentication for wireless ad hoc networks," in International conference on information technology: Coding and Computing (ITCC's 04), Las Vegas, Nevada, April 2004, pp. 1-5.

[6] Y. Zhang, W. Liu, W. Lou and Y. Fang, "Securing mobile ad hoc networks with certificateless public keys," IEEE Transaction on Dependable and Secure Computing, Vol. 3, No. 4,Dec. 2006, pp. 386-399.

[7] H. Harney, C. Muckenhirn, "Group key management protocol (GKMP) architecture", Network Working Group, July 1997.[Online].Available:https://www.rfc-editor.org/info/rfc2094. [Accessed: Jan. 1, 2018].

[8] H. Harney, C. Muckenhirn, "Group Key Management Protocol(GKMP) Specification" ,Internet Request for Comments 2093," July 1997.[Online]. Available:https://www.rfc-editor.org/info/rfc2093. [Accessed: Jan. 1, 2018].

[9] H. Harney ,U. Meth, A. Colegrove and G. Gross, "Group Secure Association Key Management Protocol(GKMP)", Internet Request for Comments 4535," June 2006. [Online]. Available:https://www.rfc-editor.org/info/rfc4535. [Accessed: Jan. 1, 2018].

[10] B. Weis, S. Rowles and T. Hardjono, "The Group Domain of Interpretation(GDOI)",Internet Request for Comments 6407, Oct.2011.[Online]. Available:https://www.rfc-editor.org/info/rfc6407. [Accessed: Jan. 1, 2018].

[11] Bryans, J. W., Fitzgerald and J. S., "Formal engineering of XACML access control policies in VDM++," in International Conference on Formal Engineering Methods, Florida, USA, Berlin, Heidelberg, Nov. 2007, pp. 37–56.

[12] K. Fisler, S. Krishnamurthi, L. A. Meyerovich and M. C. Tschantz, "Verification and change-impact analysis of access control policies," in Proc. of 27th International Conference on Software Engineering, MO, USA,May 2005, pp. 196-205.

[13] D. Jackson, , Software Abstractions: Logic, Languages, and Analysis, MIT Press, ISBN: 978-0-262-10114-1, 2006. .

[14] D. Jackson, "Micromodels of Software: Lightweight Modelling and Analysis with Alloy," MIT Lab, Jan. 2002. [Online]. Available:https://courses.cs.washington.edu/courses/cse503/04sp/readings/alloy-ref.pdf. [Accessed: Jan. 1, 2018].

[15] D. Jackson, "Alloy: a lightweight object modelling notation," ACM Trans. Soft. Eng. Methodol., Vol. 11, No. 2, April 2002, pp. 256-290.

[16] V. Chandola, A. Banerjee and V. Kumar, "Anomaly Detection: A Survey," ACM computing surveys, Vol. 41, No. 3, 2009, pp. 1-72.

[17] Y. Zhang, N. Meratnia and P. Havinga, "Outlier Detection Techniques for Wireless Sensor Networks: A Survey," IEEE Communication Surveys & Tutorials, Vol. 12, No. 2, 2010, pp. 159-170.

[18] P. Gogoi, B. Borah and D. K. Bhattacharyya, "Anomaly Detection Analysis of Intrusion Data using Supervised and Unsupervised Approach," Journal of Convergence Information Technology, Vol. 5, No. 1, Feb. 2010, pp. 95-110.

[19] P. Gogoi, D. K. Bhattacharyya, B. Borah and J. K. Kalita, " A Survey of Outlier Detection Methods in Network Anomaly Identification," The Computer Journal, Vol. 54, No. 4, April 2011, pp. 570-588.

[20] D. M. Hawkin, Identification of Outliers, London: Chapman and Hall, 1980.

[21] V. A. Traag, A. Browet, F. Calabrese and F. Morlot, "Social Event Detection in Massive Mobile Phone Data Using Probabilistic Location Interference," in SocialCom/PASSAT, 9-11 October 2011.

[22] B. Krishnamachari and S. Iyengar, "Distributed Bayesian algorithms for fault tolerant event region detection in wireless sensor networks," IEEE Transactions on Computers, Vol. 53, No. 3, March 2004, pp. 241-250.

[23] F. Martincic and L. Schwiebert, "Distributed event detection in sensor networks," in Proceedings of Systems and Network Communication, French, Polynesia, Nov. 2006, pp. 1-6.

[24] M. Ding, D. Chen, K. Xing and X. Cheng, " Localized fault tolerant event boundary detection in sensor networks," in IEEE conference of computer and communications socities, Florida, USA, March 2005, pp. 902-913.

[25] A. P. R. Silva, M. H. T. Martins, B. P. S. Rocha, A. A. F. Loureiro, L. B. Ruiz and H. C. Wong, "Decentralized intrusion detection in wireless sensor networks," 1st ACM international workshop on Quality of Service and Security in Wireles., Quebec, Canada Oct. 2005, pp. 16-23.

[26] J. Chen, S. Kher and A. Somani, "Distributed fault detection of wireless sensor networks,"Proceedings of the 2006 workshop on dependability issues in wireless ad hoc networks and sensor networks, CA, USA, Sep. 2006, pp. 65-72.

[27] X. Luo, M. Dong and Y. Huang, "On distributed fault tolerant detection in wireless sensor networks," IEEE Transactions on computers, Vol. 55, No.1, Jan. 2006 pp. 58-70.

[28] J. Raja, X. R. Wang, O. Obst and P. Valencia, "Wireless sensor network anomalies: Diagnosis and detection strategies,"Intelligence-Based Systems Engineering, Berlin, Heidelberg, 2011, pp. 309-325.

[29] W. Hu, T. Tan, L.Wang and S. Maybank, "A survey on visual surveillance of object motion and behaviors," IEEE transavtion, Vol. 34, No. 3,July 2004, pp. 334-352.

[30] D. M. Hawkins, Ident fication of outliers, London: Chapman and Hall, 1980.

[31] E. M. Knorr and R. T. Ng, "Algorithm for mining distance based outliers in large datasets," 24th international conference on very large databases, New York, USA, 1998, pp. 392-403.

[32] M. M. Breunig, H. P. Kriegel, R. T. Ng and J. Sander, "LOF: Identifying Density Based Local Outliers," ACM SIGMOD, Dallas, TX, USA, May 2000, pp. 93-104.

[33] B. Wang and W. Perrizo, "RDF: a density-based outlier detection method using vertical data representation," in Fourrth IEEE InternationalConference on Data Mining, Nov. 2004, pp. 1-4.

[34] S. Rajagopalan, R. Karwoski, B. Bartholmai and R. Robb, "Quantitative image analytics for strtified pulmonary medicine," in IEEE Int. Symposium on Biomedical Imaging (ISBI), Barcelona, Spain,May 2012, pp. 1779-1782.

[35] J. W. Branch, C. Giannelia, B. Szymanski, R. Wolff and H. Kargupta, "In-network outlier detection in wireless sensor networks," Knowledge and information systems, Vol. 34, No. 1, Jan. 2013, pp. 23-54.

[36] H. Ayadi, A. Zouinkhi and B. Boussaid, "A Machine Learning Methods: Outlier detection in WSN," in 16th international conference on Sciences and Techniques of Automatic control, Monastir, Tunisia, December 2015, pp. 722-727.

[37] C. Titouna, M. Aliouat and M. Gueroui, "Outlier Detection Approach Using Bayes Classifiers," Wireless Pers. Communications, Vol. 85, No. 3, June 2015, pp. 1009-1023.

[38] J. C. M. Teo and C. H. Tan, "Energy-efficient and scalable group key agreement for large ad hoc networks," in Proceedings of the 2nd ACM international workshop on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks, Quebec, Canada, October 2005, pp. 114-121.

[39] E. W. Dijkstra, "A note on two problems in connexion with graphs," Numerische Mathematik, Vol. 1, No. 1, December 1959,pp. 269-271.

[40] J. Spinrad, "Recognition of circle graphs," Journal of Algorithms, Vol. 16, No. 2, March 1994, pp. 264-282.

[41] W. J. Gutjahr, "A graph-based Ant System and its convergence," Future Generation Computer Systems, Vol. 16, No. 9, June 2000, pp. 873-888.

[42] A. Shamir, "How to share a secret," Communications of the ACM, Vol. 22, No. 11, November 1979, pp. 612- 613.

[43] J. V. D. Merwe, D. Dowoud and S. McDonald, "A Survey on Peer to Peer key management for Mobile Ad Hoc Networks," ACM Computing Surveys, Vol. 39, No. 1, Article 1, April 2007, pp. 1-45.

[44] "The Network Simulator - ns-2," [Online]. Available: https://www.isi.edu/nsnam/ns/. [Accessed 18 7 2018].

[45] "scipy.cluster.hierarchy.dendrogram.html,"[Online].Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.dendrogram.html. [Accessed 18 7 2018].

[46] A. Kumar, K. Gopal and A. Aggarwal," Novel Trusted Hierarchy Construction for RFID Sensor–Based MANETs Using ECCs," ETRI Journal, Vol. 37, No. 1, July 2015, pp. 186-196.

[47] A. Kumar, K. Gopal and A. Aggarwal, " Simulation and analysis of authentication protocols for mobile Internet of Things (MIoT)," 2014 IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC), JUIT, Waknaghat, India, 2014, pp.423-428.

[48] A. Kumar, K. Gopal and A. Aggarwal, "Design and Analysis of Lightweight Trust Mechanism for Accessing Data in MANETs," KSII Transactions on Internet & Information Systems, Vol. 8, No. 3, March 2014, pp. 1119-1143.

[49] A. Kumar, K. Gopal and A. Aggarwal, "Cost and Lightweight Modeling Analysis of RFID Authentication Protocols in Resource Constraint Internet of Things," Journal of Communications Software and Systems, Vol. 10, No. 3, September 2014, pp. 179-143.

[50] A. Kumar, K. Gopal and A. Aggarwal, " A complete, efficient and lightweight cryptography solution for resource contrainsts Mobile Ad-Hoc Networks", 2nd IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC), JUIT, Waknaghat, India, Feb. 2013, pp. 854-860.

[51] A. Kumar, K. Gopal and A. Aggarwal," Design and Analysis of Lightweight Trust Mechanism for Secret Data using Lightweight Cryptographic Primitives in MANETs", IJ Network Security, Vol. 18, No. 1, Jan. 2016, pp.1-18.

[52] A. Kumar, K. Gopal and A. Aggarwal,"A novel lightweight key management scheme for RFID-sensor integrated hierarchical MANET based on internet of things", International Journal of Advanced Intelligence Paradigms, Vol. 9, No. 2-3, 2017, pp. 220-245.

[53] N. Chugh, A. Kumar and A. Aggarwal, "Security aspects of a RFID-sensor integrated low-powered devices for internet-of-things", 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), JUIT, Waknaghat, India, Dec. 2016, pp. 759-763.

AUTHORS

Neeraj Chugh is an Assistant Professor in University of Petroleum & Energy Studies, Dehradun, India and enrolled in PhD (CSE) from Uttrakhand Technical University (UTU), Uttrakhand, India. He received his M. Tech. (CSE) from Kurukshetra University Kurukshetra, India in 2001. His research interests includes Database Management system, Data Mining, and Outlier/Anomaly detection and event detection in sensor networks.

Adarsh Kumar received his ME degree in Software Engineering from Thapar University, Patiala, Punjab, India, in 2005 and earned his PhD degree from JIIT university, Noida, India in 2016 followed by Post Doctoral from AIT, Ireleand during 2016-2018. From 2005 to 2016, he has been associated with the Department of Computer Science Engineering & Information Technology, Jaypee Institute of Information Technology, Noida, UttarPardesh, India, where he worked as Assistant Professor. Currently he is working with University of Petroleum & Energy Studies, Dehradun, India as Associate Professor in CSE department. His main research interests are cryptography, network security, and adhoc networks.

Alok Aggarwal received his bachelors' and masters' degrees in Computer Science& Engineering in 1995 and 2001 respectively and his PhD degree in Engineering from IITRoorkee, Roorkee, India in 2010. He has academic experience of 18 years, industry experience of 4 years and research experience of 5 years. He has contributed more than 150 research contributions in different journals and conference proceedings. Currently he is working with University of Petroleum & Energy Studies, Dehradun, India as Professor in CSE department. His main research interests are wired/wireless networks, security, and coding theory.