International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 11 (2019) pp. 055-064
© MIR Labs, www.mirlabs.net/ijcisim/index.html

Received: 18 Dec, 2018; Accepted: 21 March, 2019; Publish: 17 April, 2019

An Efficient Simulated Annealing based Constrained Optimization Approach for Outlier Detection Mechanism in RFID-Sensor Integrated MANET

Adarsh Kumar[1] and Alok Aggarwal[2]

[1,2]School of Computer Science, Department of Systemics, University of Petroleum and Energy Studies
Bidholi Campus, Dehradun, India
[1]adarsh.kumar@ddn.upes.ac.in, [2]alok.aggarwal@ddn.upes.ac.in

Abstract: **Designing an outlier detection process for unknown resources is a challenging task. It may contain resourceful or resource-constrained devices. In this work, a multi-regional and multi-layered outlier detection process is proposed. Proposed approach implements MAC, routing and application layer outlier detection processes in three different regions. These regions are designed with priority of resources and importance of stakeholder taken into considerations. Similar outlier processed with different datasets is used for outlier detection in this multi-region invigilator architecture. Proposed architecture is verified through internal, external and performance based indices. Simulation results shows the cluster stability in process of data formalization and outlier detection. Internal and external indices shows that the maximum stability is possible with 5 to 500 nodes and 26 clusters for small scale network, 500 to 3000 nodes with 41 clusters for medium scale network and 3000 to 6000 nodes network with 54 clusters for large scale network.**

Keywords: Outlier, inlier, trust, indices, performance, machine learning.

I. Introduction

The term Internet of Things (IoT) is used to describe a network which connects various communication entities. These communication entities may be simple sensing devices like thermostats used for various domestic applications, printers, healthcare devices, smart phones, various electrical/electronic/mechanical devices, various control systems deployed at manufacturing units etc. During a short span of two decades ample work has been done in the field of IoT which is expending in various domains like transport & logistics, farming, wearables, smart city, banking, insurance, general security etc. IoT as standalone gives a limited applications however when clubbed with wireless sensor networks (WSNs) and mobile adhoc networks (MANETs), it gives enormous applications for societal development. WSN is a network that consists of many (few dozens to thousands) low cost tiny sensor devices which sense and collect detailed information about the physical environment over a large area. Each node in a WSN has a tiny micro-controller, radio receiver, power source and uni-or multi-type sensors like humidity, pressure, temperature, sound, vibration, heat etc. A fine-grained real time information about the physical environment is provided by a WSN which due to its cost is used in various general application discussed earlier and to few specific applications also like identifying nuclear/biological/chemical attacks, instance exploring of the battle field, learning wild life & ocean life, monitoring highway traffic, agriculture, space exploration etc. MANET is a group of low cost, low powered, economical, tiny computing mobile and wireless devices which forms a decentralized and infrastructure less network. Radio Frequency IDentification (RFID) devises are used widely now-a-days and preferred over bar codes, magnetic tapes and smart cards due to their low cost and high speed for identification, location tracking and record management purpose. RFID consists of tags, readers and backend storage devises. Physical state of an object like temperature, pressure, vibration, sound etc. can not be determined by RFID devises but by a WSN. These two technologies, RFID and WSN, complement to each other and plays a vital role in IoT.

A mobile RFID-WSN consists of smart mobile nodes with RFID tags & readers. Integration of RFID with WSN gives a better scalability, capability along with cost effectiveness. MANET-IoT based system relies on MANET routing protocols for networking purpose with WSN routing principles which has various capabilities like data sensing, handling & processing. Major challenging issues in such type of MANET-IoT systems are availability, routing and reliability. Routing protocols play a vital role in these resource constraint devices. For these systems, reliability is ensured by periodically identification of under-performing and out-performing nodes.

Data measured and collected by WSNs is often unreliable, inaccurate, susceptible to environmental effects and vulnerable to malicious attacks like DoS attack, black-hole attack and eavesdropping. Unreliability in the data measured and collected by WSNs comes from various sources like quality of data set is usually affected by noise & error, duplicated data, missing values in the data and inconsistency in the data. In WSNs, sensor nodes deployed are of low cost with low quality due to cost consideration in these type of networks. This is a major source of inaccuracy to the data measured and collected by WSNs. Probability of erroneous data grows rapidly in case battery power gets exhausted. In case of large scale and high density WSNs, number of sensor nodes goes upto few thousands which need to be deployed in harsh and unaffected environments. This is a major cause of susceptible data measured and collected by WSNs. All these factors lead to especially unreliability of sensor data. Many major events, like chemical spill, earth-quack and forest fire etc., can not be accurately detected using unreliable, inaccurate, susceptible data. Hence it becomes necessary to ensure the reliability and accuracy of the data measured and collected by WSNs before some decision making process. Outlier detection technique is one of the ways for ensuring reliability of data.

An outlier, also known as anomaly, originally stems from the field of statistics, is an observation (or subsets of observations) which appears to be inconsistent with the remainder of that set [1]. Noise, errors, malicious attacks and actual events are among the major sources of an outlier. Outlier detection is a fundamental task of predictive modelling, data mining, cluster analysis and association analysis. Statistics, machine learning, information theory, spectral decomposition, data mining are various discipline where outlier detection has been widely researched. Network intrusion, whether prediction, performance analysis, fraud detection are few application domain of outlier detection. Outlier detection mechanisms are also very useful for identifying the values which do not follow the normal pattern. Sensor data in the network and these abnormalities serve various purposes from deep analysis point of view. With respect to the availability of nodes, an outlier detection mechanism identifies those nodes which are relevant for secure communication. Majority of outlier detection mechanism adopts statistical approaches for identifying outlying nodes. Statistical properties are used for training and testing in statistical outlier detection mechanism which can be either parametric or non-parametric. Some underlying distribution such as normal or Gaussian using means and covariance is assumed in formal approaches while the later approaches are silent to the statistical properties of the data. Univariate and multi-variate analysis also plays a vital role in identifying outliers. Many researchers focused on single layered outlier detection solutions which are less efficient compared to multi-layered outlier detection solutions.

In this work, an efficient simulated annealing based multi-region and multi-tiered constrained optimization approach is proposed for outlier detection in RFID-Sensor Integrated MANET. The proposed approach is suitable for unknown devices as it has detection mechanisms with increasing computational complexity. In this work, outlier detection approach presented in [2] is extended with outlier detection, trust management and indices for multi-regional invigilation. Trust management approach applies trust computation, distribution and aggregation operation in initial phases of outlier detection. Indices based outlier detection uses internal, external and performance based indices for advanced phase of outlier detection process. In simulation, indices values are computed for 5 to 6000 nodes networks. Results indicates the ideal number of clusters required for considering the network without outliers. Performance based indices ensures QoS. Thus, proposed outlier detection approach is efficient for resourceful and resource constrained devices.

Rest of the paper is organized as follows. Section 2 gives in brief a summary of the works done by earlier researchers for different optimizations techniques of outlier detection in WSNs and particularly to RFID-sensor integrated MANET. Section 3 presents proposed multi-regional and multi-layered outlier detection approach. Section 4 evaluates the indices in order to measure the stability of structures and ensures the QoS for 5 to 6000 nodes networks. Finally, conclusion is drawn in section 5.

II. Literature Survey

During the last two decades various outlier detection mechanism have been proposed for MANETs. Outlier detection techniques for WSNs can be broadly divided into six categories, statistical based [3]-[11], nearest-neighbor based [12]-[17], clustering based [18]-[19], classification based [20]-[23], spectral decomposition based [24], density based [25]-[27] approaches etc. The earliest approaches that were used for outlier detection were based on statistical approaches in which a probability distribution is estimated for capturing the data distribution and to evaluate the data instances for estimating the fitness to the model [11][28]. Authors in [4]-[6] have presented Gaussian-parametric statistical based approaches and in [7] non-Gaussian based, [8]-[10] non-parametric statistical based approaches. In [4], two level technique has been proposed for identifying outlying sensors. For distinguishing outlying sensors and event boundary spatial correlation is taken of the existing reading among neighboring sensor nodes. Each node computes the difference in its own reading and the median reading of its neighboring reading. A node is treated as outlying node if this difference exceeds the pre-selected threshold. In [6], a spatio-temporal correlation of sensor data is used for identifying an outlier. In [7], a non-Gaussian parametric local technique is proposed which uses simple operations like average, max etc. and spatio-temporal correlation of sensor data is used.

Authors in [8]-[10] proposed a non-parametric statistical based approach of outlier detection based on histogram [8] and kernel function [9]-[10]. Shang et al. [8] has proposed a histogram-based approach which identifies global outlier in data correlation applications of sensor networks. It is shown that communication cost is considerably reduced by collecting histogram information instead of collecting raw data for centralized processing. Proposed technique however considers one-dimensional data and can not be applied to multi-dimensional data. Authors in [9] have proposed a kernel based approach which identifies outliers online in streaming sensor data. Advantage of proposed approach is that it requires no priori known data distribution however it suffers from the insufficiency of a single threshold for multidimensional data and maintain the data model built by kernel density estimator. These two problems have in addressed by the authors in [3] and [9] by extending the work proposed in [9] and proposed two global outlier detection techniques for complex applications [3] and to locally detect global outliers by having a copy of global estimator model obtained from the sink.

Nearest neighbor based [12]-[16] outlier detection techniques for WSNs analyze a data instance with respect to its nearest neighbor in the data mining and machine learning community. For WSN data collection applications, two in-network outlier cleaning techniques have been proposed in [16] in which one technique uses wavelet analysis and other one uses dynamic time warping distance based similarity comparison. Authors in [14] have proposed a nearest neighbor based approach for identifying global outliers in the WSNs where proposed approach reduces the communication cost by a set of representative data exchanges among neighboring nodes. Distance similarity is used by each node for locally identifying outliers. Later identified outliers are broadcasted to the neighboring nodes for the verification purpose. Proposed approach does not adopt any network structure and hence is not well suited for large scale networks. This limitation was covered in [15] which adopts the structure of aggregation tree.

It is also claimed in [15] that the communication cost compared to [14] is also considerably low as it prevents broadcasting of each node in the network. Each node in the tree does not send the full data but only useful parts of it and this procedure is repeated till the agreement on the global results calculated by the sink. Drawback of proposed scheme in [15] is that it considers only one-dimensional data. Authors in [18] have proposed a clustering based outlier detection technique for WSNs where communication overhead is minimized by clustering the sensor measurements and merging clusters before communicating with each other nodes. Major benefits of the proposed approach is that it does not require a priori knowledge of the data distribution and hence can be used in an incremental model which is a major drawback of the earlier approaches. Similar to [18], same authors also proposed a classification based approach in [20]. For reducing communication overhead and locally identifying outliers at each node, the proposed approach uses one-class quarter-sphere state vector mechanism where sensor data lying outside the quarter sphere is considered as an outlier. Authors in [21] have also proposed a classification based approach of outlier detection in WSNs. In this approach dynamic Bayesian network is used for identifying local outliers in environmental sensor data streams. Dynamic Bayesian networks are used so as to cover the fast track changes in the dynamic network topology of sensor networks. Several data streams can be operated simultaneously.

In a WSN, a node may behave like an outlier which is due to various reasons like environment [29], hardware/software [30]-[32], uncertainty of data [33]-[34], deviation from regular pattern of the system [34]-[35] etc. Further these anomalies may occur at various levels like node, data or network level [36]-[39]. A Margrave tool is proposed by the authors in [40] for checking the user specified properties of the policy where the tool helps to check duty constraints, permission, roles, presence, absence and behavioral response from policy members [41]-[43].

RFID-sensor integrated MANET are resource constraint devices due to cost considerations which forces these devices to go for light weight key management algorithms. Authors in [44] have compared three light weight key management protocols. Teo &Tan [45], WLH [46], and Tseng [47] and observed that light weight key management protocols proposed for RFID-sensor integrated MANET by Teo &Tan [45] outperforms the rest two protocols proposed in [46]-[47] in terms of delay, throughput and security. Work in [48] is an extension of the work done by the same researchers in [44]. Work done by the researchers in [44] suffers from the cryptographic property, availability. Availability was not ensured in [44] which was ensured in [49] for RFID-sensor integrated MANET through outlier detection mechanism. Unpresedental data is detected where precedental data is identified from resource constraint mobile sensor devices. Through anomaly scores inliers and outliers are identified for protection against DoS attack. Using threshold based outlier detection mechanism, upper and lower threshold limits are computed for outlier identification. From throughput perspective, a minimum improvement of 6.2 % and a maximum of 219.9% while from packet delivery ratio perspective a minimum improvement of 8.9% and a maximum of 19.5% is observed compared to the work presented by researchers in [44]. In [50], an outlier detection scheme is proposed for RFID-sensor integrated MANET which is multi-filtered. Performance of individual and group nodes is calculated. Different number of clusters are used for calculation purpose ranging from small-scale network (0-500 nodes), medium-scale network (500-3000 nodes) and large-scale network (3000-5000 nodes). Average cluster stability of 61% is observed by the authors.

Various multi-layered outlier detection models for resource constraint hierarchical MANET is proposed by researchers in [51]-[59]. Lighter statistical techniques are applied to various layers due to constraint of available resources. In [51], authors have proposed a multi-layered outlier detection algorithm using hierarchical similarity metric with hierarchical categorized data. Two QoS parameters, APDR &AT, have been taken for performance analysis and it is observed that with the proposed approach APDR improvement is 9.1% to 22.1% while AT improvement of 0.61% to 104.1% for a network having nodes from 100-3000, i.e. small and medium scale networks. In [52], an anomaly detection approach is proposed using cross layer for resource constraint devices where two layers, namely MAC and routing, are used for outlier detection. Packet drop count is used at MAC layer while missed IP DSN is used at routing layer. It is shown that the proposed multi-layered approach can detect and isolate black hole attacks from the network. Similar to [52], in [53] a similar anomaly detection approach is proposed using cross layer for resource constraint devices where two layers, namely MAC and routing, are used for outlier detection. At the first level a decision tree classification is used for generating instances and at the next level accumulated measure of fluctuation of the received classified instances are used. Similar to the works proposed by the researchers in [52]-[53], works in [54]-[59] also uses a two level outlier detection schemes using MAC and routing layers.

III. Proposed Approach

This section explains the proposed multi-region and multi-layered architecture in detail as shown in fig. 1. In this architecture, outlier detection process is divided into three regions: primary, secondary and tertiary. Each region executes same set of experimentations with different data and algorithmic approaches. Data classification and selection of algorithms is based on prior experimentation, evaluation and analysis. In detail, proposed Architecture has following components:

Fig. 2 shows all possible experimentations in using the proposed architecture with scope of availability of data. Three invigilator regions are divided among three ranks rank 1, rank 2 and rank 3. A priority level is associates with each region using ranks. Three priority levels are: low, medium and high. All processes defined with high priority region are mandatory to execute, At least one process should be executed in medium priority region and execution of processes is lowest priority region are not mandatory. Experiment 1 shows that all data should be availability to top management in an organization which are given a high priority status and filtration is required thereafter. This scenario is preferred if data contains important information. For example, personal identities, passwords, financial reports etc. Experimentation 2 made all data to lower management people and filtered data is available to top management. This scenario is practiced in regular experimentations. Top management has to put minimum efforts in deciding that whether organization resources should be used in tackling specific attacks or not. Experimentation 3 shows a scenario where lower management is given highest priority and they have to tackle all attacks which are possible through available resources. Thereafter, top management will decide whether new resources need to be brought in for unknown attacks or not. Experimentation 4 shows a scenario where lower management is given higher priority with minimum data. Important data elements are either filtered or encrypted before sending it to rank 3 processes. In this case, data patterns from unknown data are made available for attack scanning. However, this work uses experiment 1 for analysis. Comparative analysis of all possible scenarios will be drawn in future.

- **Primary Automated Invigilator Region:** This region is constructed to group resources, experimentation conducted and their evaluations, and organization strategies that can be managed by lifetime, owner and organization specific criteria.
- **Secondary Automated Invigilator Region:** This region is constructed to group resources, experimentation conducted and their evaluation and department specific strategies that can be managed by middle management of an organization using either general or specific criteria.
- **Tertiary Automated Invigilator Region:** This region is constructed to group data, resources and regular experiments that can be managed by non-technical or clerical staff using general trained procedures.
- **Monitoring System:** Fig. 3 shows proposed initial monitoring system. In this system, various sensors are deployed for collecting raw information. Functionalities of various sensors deployed in collecting raw data are explained as follows:
  - *Alarm information sensor:* This collects information about unknown single or group of sources which are continuously sending or receiving data as shown in fig. 4.
  - A controller unit in this sensor collect information from single or group of hosts. This information is stored in data flow table which is used to analyze the context of data communication. Data controlling unit sends the received

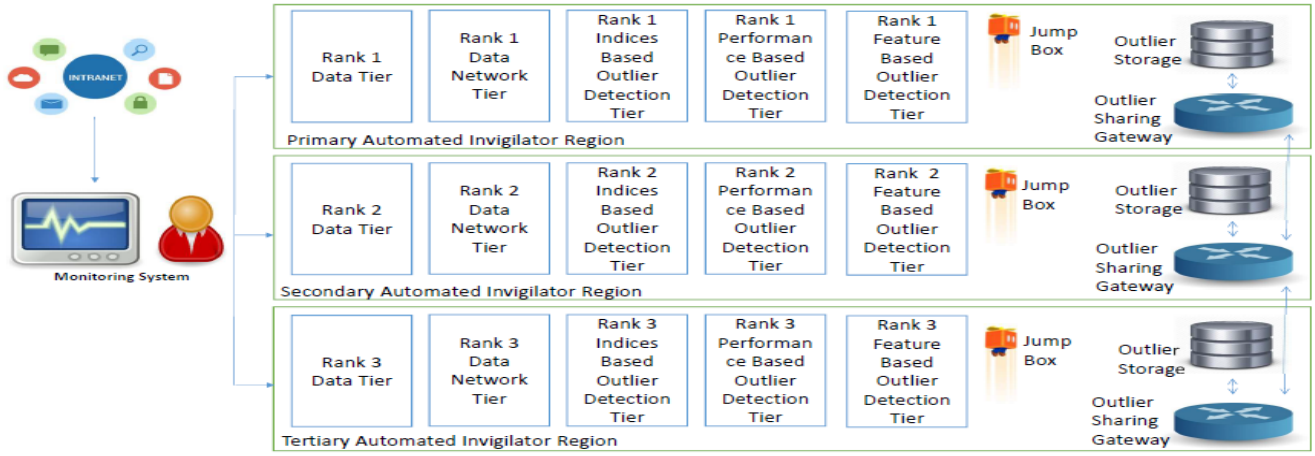
INTRANET
Monitoring System
Rank 1 Data Tier
Rank 1 Data Network Tier
Rank 1 Indices Based Outlier Detection Tier
Rank 1 Performance Based Outlier Detection Tier
Rank 1 Feature Based Outlier Detection Tier
Jump Box
Outlier Storage
Outlier Sharing Gateway
Primary Automated Invigilator Region
Rank 2 Data Tier
Rank 2 Data Network Tier
Rank 2 Indices Based Outlier Detection Tier
Rank 2 Performance Based Outlier Detection Tier
Rank 2 Feature Based Outlier Detection Tier
Jump Box
Outlier Storage
Outlier Sharing Gateway
Secondary Automated Invigilator Region
Rank 3 Data Tier
Rank 3 Data Network Tier
Rank 3 Indices Based Outlier Detection Tier
Rank 3 Performance Based Outlier Detection Tier
Rank 3 Feature Based Outlier Detection Tier
Jump Box
Outlier Storage
Outlier Sharing Gateway
Tertiary Automated Invigilator Region

Figure 1: Proposed Ranking Based Multi-Layered Outlier Detection Architecture

Experimentation 1
Rank 1: Highest Priority (All Data)
Rank 2: Medium Priority (Lesser Data)
Rank 3: Lowest Priority (Least Data)

Experimentation 2
Rank 1: Highest Priority (Least Data)
Rank 2: Medium Priority (Lesser Data)
Rank 3: Lowest Priority (All Data)

Experimentation 3
Rank 3: Highest Priority (All Data)
Rank 2: Medium Priority (Lesser Data)
Rank 1: Lowest Priority (Least Data)

Experimentation 4
Rank 3: Highest Priority (Least Data)
Rank 2: Medium Priority (Lesser Data)
Rank 1: Lowest Priority (All Data)

Figure 2: Proposed Experimentation over Multi-Layered Outlier Detection Architecture

data for information extraction using packet parser. Extracted information is scanned through signature and verification processes. If signatures are not verified then a alarm is generated for monitoring system. During packet parsing process, if additional information is required for estimating and testing the flow of data using rule based system then a request is passed to controller unit for collecting the required data. If a flow is found using new data then necessary information is stored in data flow table.

- *Regular Inventory Information Sensor:* This sensor collects basic information about network and communications over such networks. Information includes unattended local and remote system's records, server connected and clientless system's records, operating system and installed software in the network, software packages and installation tracking records, used and unused hardware resources in the network, network assets specifications, software compliance records etc.

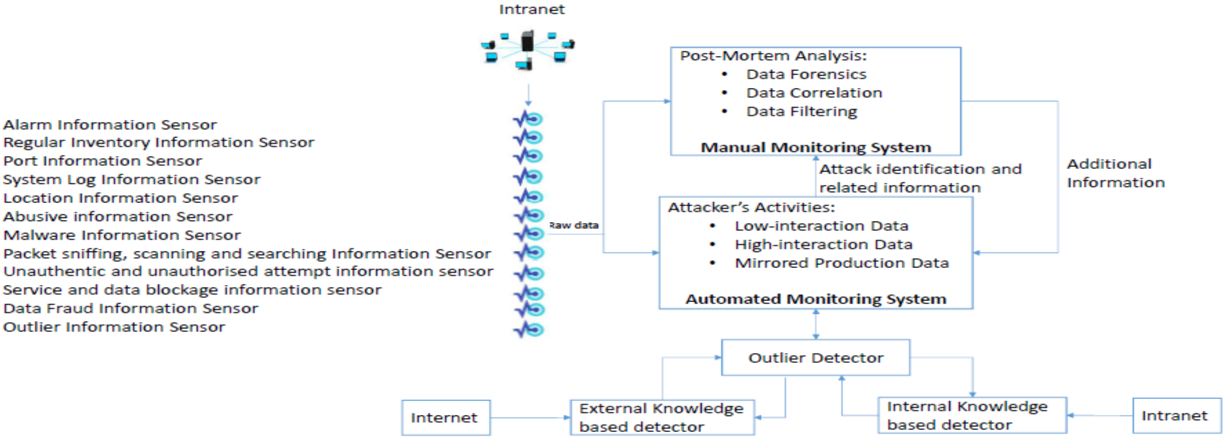
Intranet
Alarm Information Sensor
Regular Inventory Information Sensor
Port Information Sensor
System Log Information Sensor
Location Information Sensor
Abusive information Sensor
Malware Information Sensor
Packet sniffing, scanning and searching Information Sensor
Unauthentic and unauthorised attempt information sensor
Service and data blockage information sensor
Data Fraud Information Sensor
Outlier Information Sensor
Raw data
Post-Mortem Analysis:
Data Forensics
Data Correlation
Data Filtering
Manual Monitoring System
Attack identification and related information
Additional Information
Attacker's Activities:
Low-interaction Data
High-interaction Data
Mirrored Production Data
Automated Monitoring System
Outlier Detector
Internet
External Knowledge based detector
Internal Knowledge based detector
Intranet

Figure 3: Proposed Initial Intrusion Monitoring System

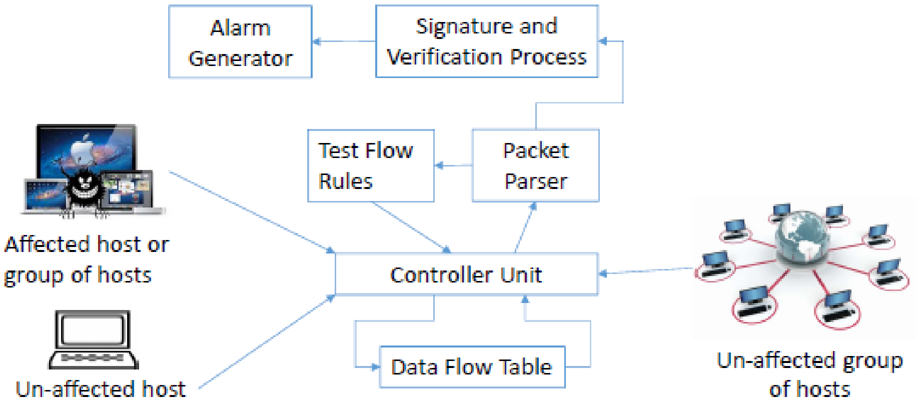
Alarm Generator
Signature and Verification Process
Test Flow Rules
Packet Parser
Affected host or group of hosts
Controller Unit
Data Flow Table
Un-affected host
Un-affected group of hosts

Figure 4: Proposed Alarm Information Sensor System

- *Port Information Sensor*: This sensor collects the information about those hosts which sends client requests to single or range of servers. These requests find active ports and exploits the vulnerability in those services which are running on ports. Major goal of this sensor is to scan the requests and keep track of those requests which are receptive or useful to specific needs of request senders.
- *System Log Information Sensor:* This sensor generates alerts to those system logs which are suspicious. For example, information related to login and session activity includes login frequency at single or multiple points in fixed premises, session time spent by each activity and change in output of any software or website. Similarly, resource utilisation includes password failure attempts in login activity, commands, and procedures executed in resource utilisation and their operating frequency. Further, frequency of system and user file read, write, edit, create and delete operations are scanned using this sensor.
- *Location Information Sensor:* This sensor is used to collect locations of packets transmitted over a network. Location includes source, intermediate and destination addresses, ports used by packets over every node used for its transmission, addresses of hosts and intermediate devices where packets are stored etc.
- *Abusive information Sensor*: This information sensor collects data like spamming emails, individual harassments, pornography contents or violence through words. Sensor is enabled with content and context based scanning methodology. Contents based methodology searches words, lines and paragraphs used to discredit or discriminate someone. Context based methodology applies rule based mechanism to correlate data with events. All unsolicited events are reported using this sensor.
- *Malware Information Sensor*: This information sensor collects data about those software which are intentionally applied or installed in the network with an intention of

damaging the resources. For example, worm, trojan, virus, shellshock, botnet, miscellaneous attacks etc.

- *Packet sniffing, scanning and searching Information Sensor:* This information sensor collect data that is relevant for identifying unscrupulous activities. For example, activities involved in unnecessary scanning of network devices used for finding weak network points to exploit vulnerabilities. Similarly, man-made tricks and threats are tried to be identified using outlier detection processes.
- *Unauthentic and unauthorised attempt information sensor*: This information sensor measures the frequencies used in various activities like: login to the system, access to resources, packet/message signature modification etc. Attempts greater than certain threshold are considered in unauthentic and unauthorised activities.
- *Service and data blockage information sensor:* This information sensor collects the data related to services running in the network, whether any service is interrupted or not and what are the network resources used by these services.
- *Data Fraud Information Sensor:* This information sensor identify falsification, fabrication, spoofing and scams in resource related data.
- *Outlier Information Sensor:* This information sensor evaluates the outliers in data collected using various sensors like: service blockage, unauthorised access, unauthentic executions, location information etc. This is a threshold based content sensor for outlier prediction.

- **Data Tier:** In this tier, data sensed by sensors is collected for feature based analysis. Different data features are made available to different regions. In experimentation, multiple scenarios are generated for outlier identification. In one scenario, whole collect data is made available to primary region and least data is available to tertiary region. Whereas, second scenario provides more data to tertiary region and filtered data is passed to primary region through second region. Purpose of this data passing and filtration is to analysis and identify the best scenario suitable for outlier identification. Information passed through different regions contains different level details. For example, packet level details including MAC, IP and TCP header details are considered as low level details. Filtered data including host IP addresses, user with host having specific address information, type of possible attack etc. are considered as high level details. Low level details are passed to one region whereas high level details are passed to another region.

- **Data Network Tier:** In this tier, a data network is constructed from collected information. The process of constructing an initial data network constitutes clusters. These clusters are formulated using interconnected data contents. A process of selecting trusted cluster-head is executed in data network. Cluster head select process is based on trust management. Trust management goes through trust generation, trust propagation and trust accumulation phases. Trust generation phase identifies the number of interconnected other data units and their duration. Generated trust value is passed to all connected hosts in trust propagation phase. This increases the importance of those resources which are used for data transmission. Passed trust value through multiple channels is aggregated in trust accumulation phase. Accumulated trust value is used for outlier detection in subsequent phases. The complete process of constructing hierarchical network through data network based trust management process is explained in pseudocode 1.

Pseudocode 1: Hierarchical data network construction algorithm using interconnected clusters

Goal: To measure the data trust value and interconnect data nodes in hierarchical network.

```
1. Data collected through sensors is connected using context based process.
2. Implicit data information provides linkage between different data units. Connected data units are put in one cluster.
3. Number of interconnection and their duration are used for computing trust value. Higher trust value means either data unit has used large number of resources or same resources are used multiple times for longer duration.
4. A complete interconnected data unit based network is formulated using trust values.
5. Each data unit in connected network is picked one by one. Features of data unit are extracted for analysis. This process of analysis is explained in detail as follows:
    a. Extracted features of most trusted data units (higher in the hierarchical interconnected network) are compared with features of least trusted data units (bottom most in the hierarchical interconnected network). This comparison formulates a dissimilarity matric. Dissimilarity score in dissimilarity matrix indicated usage of similar data elements in data units at different time.
    b. Process of dissimilarity matrix calculation is executed cluster-wise in complete interconnected network.
    c. A comparative analysis of dissimilarity matrix is executed at cluster-level. Those clusters having a large difference (greater than a threshold) in comparison with other higher trust clusters are either dissolved or consider as outliers.
    d. After comparison of clusters, a comparative analysis of low and high trusted data unit is performed. A similar process of dissolving or outlier consideration is executed inside the cluster as performed among clusters at network level.
    e. Step 5a to 5d are repeated for every data unit and cluster present in the network.
6. All clusters, consisting of multiple data units, are further divided into sub-clusters named as states or counties. States are bigger in volume as compared to counties.
7. A process of measuring the stability of clusters is executed using dissimilarity matrix calculation at state, county and cluster level. Similarity of patterns is observed within the states, state to county, within county, county to cluster,
```

cluster to network, state to cluster, state to network, county to cluster, county to network and cluster to network.

8. Any dissimilar pattern with dissimilarity score greater than zero exhibits re-clustering or alternate scenario processing. In this computation, a minimal 1% (picked randomly) marginal error is acceptable and considered in comparison. Re-clustering process includes re-computation of trust score in trust management using alternative approach.
9. An alternative approach consider importance of resources used in data unit analysis. Trust score is directly proportionate to resource importance. Resource importance based trust management approach is applied for generating new dissimilarity matrix. If dissimilarity score computed from new dissimilarity matrix is greater than zero for internal and external comparisons of state, county, cluster and network then jump box is called for skipping this phase calculations and start computing outliers using other processes.
10. Else if dissimilarity score computed from new dissimilarity matrix is lesser than zero or zero then networks consisting of clusters, states and counties are considered as stable for next phase outlier detection process.

In this tier, trust score based initial outlier detection process is executed as well. Trust management computes trust score using number of interconnected data units with every other data unit and duration of its connectivity. This trust score computation in trust management is helpful in initial dissimilarity matrix. If repeated dissimilarity score calculation falls greater than zero then trust score is computed from importance of resources and their duration using feedback, performance and energy scores. Feedback mechanism computes number of positive responses in transmitting connected data units. Performance is measured using various QoS parameters: throughput, delay, jitter, priority, protection, resilience and residual error rate. Energy score is computed from sender, receiver and intermediate node energy levels. Three factors (feedback, performance and energy level) used for trust score computation are rated on a scale from 1 to 10. An aggregated average of three factors gives final trust score value. Further, multiple cycles are used to compute the trust score because of variations in dissimilarity score. Each of these cycle alternatively computes feedback value either from neighboring connected resources or destination of data unit. Performance is initially measured from every transaction made by target node. Thereafter, size of network considered in performance measurement is reduced from n-hop connectivity to single hop (neighboring nodes). Similarly, energy level is measure initially for all transactions at source, destination and intermediate nodes. Thereafter, it is reduced to energy consumption for transactions made to importance resources. With every new value of trust score, trust management cycle is processed through dissimilarity score matrix calculation. Every new dissimilarity score matrix calculation is compared with other dissimilarity score matrix at state, county, cluster and network level. All data units or resources showing dissimilarity score greater than zero are considered as outliers and processed for further analysis in next tier.

- **Indices Based Outlier Detection Tier:** Indices based outliers are identified from clusters formed in previous tier. State, county and cluster's stability is measured sing cluster validation method. Cluster validity measures stability and it can be classified as internal, external and relative methods of measurement. In internal indices measurements, data units of state is compared to other data units of same state, data units of county is compared with other data unit of same county and data units of cluster is compared with other data unit of same cluster. These calculations are performed for dissimilarity score calculation. In external indices, data units of state is compared with data unit of other state in other county or cluster. Similarly, data units of county or cluster is compared with data unit of other county or cluster. Relative cluster validation method rate the clusters based on trust score and importance of resource before internal or external comparisons. In order to measure the stability of clusters, dissimilarity score uses compactness, separation/connectedness and connectivity as parameters. Various internal indices used in this work include are: silhouette index (SI) [60], Dunn index (DI) [61], Davies-Bouldin index (DBI) [62], Calinski-Harabasz measure (CHI) [63], Density-Based Cluster Validation (DBI)[64] etc. Various external indices used in this work include are: F-measure Indices (FI), Mutual Infor-mation (NMI) measure Indices(NMII), Purity Indices (PI), Entropy Indices (EI), Rand Indices (RI) and Jaccard Indices (JI). Pseudocode 2 discusses various steps followed in cluster stability calculations using internal, external and relative indices [65]-[68].

Pseudocode 2: Hybrid outlier detection in hierarchical network using internal, external and relative indices.

```
1. Constructed hierarchical network consists of clusters and
   dissimilarity matrix used to measure distance metric.
2. In this tier, small states, counties or clusters are consider as
   outliers and analyzed in next tier.
3. Each of the remaining state, county or cluster provides
   internal, external ad relative indices value.
4. Each of the indices value is analyzed for state, county and
   cluster.
5. **If** index value of internal, external or relative indices is
   beyond acceptable limits **then**
6.     A new list of indices confirming clusters stability is
       formalized. Each entry in this list confirms the stability
       of cluster.
7.     Internal indices entry is acceptable if indices
       comparison within same state, county or cluster is
       within acceptable range. Similarly, external and
       relative indices values are computed using comparison
       with clusters in idle network using same specification.
8. **end if**
9. In order to accept the complete experimentation, at least
   50% (selected randomly) of the indices (internal, external
   or relative) should validate the stability of network.
10. **If** number of total indices are lesser than 50% of total
    indices used for computation **then**
```

```
11.     State, county and clusters in network are not considered
        to be stable and a process of re-computing the trust and
        dissimilarity is executed again.
12. end if
```

- ***Performance Based Outlier Detection Tier:*** In this work, performance is measured using QoS parameters. These QoS parameters are further used for outlier detection. Various QoS parameters used in this process are: throughput, delay, jitter, priority, protection, resilience and residual error rate. These parameters are explained as follows[69]:

  - *Throughput*: It is defined as total number of packets successfully received at destination per unit of simulation time. Throughput is measured at different locations like: within state, within county including state, within county excluding state, within cluster including state and county, within cluster including county but excluding state, within cluster including state but excluding county, and within cluster excluding state and county. Purpose of measuring throughput for different scenarios is to identify outliers at all possible locations using this QoS parameter.
  - *Delay:* It is the end-to-end delay which includes processing, propagation and transmission delays. These delays are computed for every packet transmitted from source to destination. If source and destination lies within state then delay is country for a state and if either of source or destination lies within county then delay is counted for a county. Similarly, if source and destination lies with cluster or network then delay is counted for them only. In outlier detection process, nodes having delay greater than average value of structure (state, county, cluster or network), where source and destination are present, are processed for outlier detection using other QoS parameters or next tier outlier detection process.
  - *Jitter:* It is the delay variation in packet transmission or receiving. Like is any network, minimum jitter is good for better performance, minimum value of jitter in state, county, cluster or network is helpful in considering a packet consisting of data unit or resource as inlier. Higher jitter value ensures presence of outliers. Initially, every structure having jitter higher than average value of jitter for a network is consider as outlier and processed further either to other performance parameters or next tier outlier detection process. In continuation, every resource in every possible structure is scanned. If any resource within its structure is having jitter value higher than average jitter value of its structure then resource is considered as outlier else inlier in jitter based outlier detection process.
  - *Priority:* It is measured locally as importance of a data unit or resource. Increase in number of connected data unit with target data unit increases the importance or priority of data units in evaluation. Similarly, increase in number of connections with resource increases its priority. As discussed in trust management, number of connections increases the trust score. Thus, priority is directly computed from trust value. Resources avoiding priority channels are consider as outlier in priority based outlier detection process and sent further to outlier detection through other performance based parameters or next tier outlier detection process.
  - *Protection:* It is measured for a resource locally. It is defined as the capability of a resource to discard data packets received from unidentified or unauthentic resource. A resource in the network is unidentified or unauthentic if trust score is below average value of the structure. Like in other QoS parameter based outlier detection process, if protection for a resource is under acceptable limits then resources are considered as outliers and sent to other QoS parameter based outlier detection process or next tier outlier detection process.
  - *Residual Error Rate:* It is defined as the total number of incomplete, lost or duplicate data units present inside a specific structure. All data structures having residual error rate higher than network residual error rate are put in scrutiny. Residual error rate is also measured for resources. For a resource, it is the total number of mishandled data packets received from authentic resource which includes packet discarded, lost, misrouted or modified.
  - *Packet delivery rate:* It is measured for a resource only. It is defined as the total number of packets where the target resource is helpful for its transmission to its destination. All resources should have packet delivery rate higher than average value of packet delivery rate of its structure i.e. resources present within state should have packet delivery rate higher than average value of its state. Similarly, packet delivery rate for resources present inside county or cluster should have packet delivery rate greater than average value of county or cluster respectively. If any of these criteria is not satisfied then resources are scrutinize through other outlier detection processes.
  - *Routing Overhead:* This QoS parameter measures number of control packets present in the network. If a data unit is connected with every other data unit through control packet data unit then it is considered as outlier. Further, inside any structure those data units which are having number of control packet interconnections greater than number of data routing interconnections are put under outlier detection based scrutiny process. Routing overhead for a resource is number of control packet sent to other resources. Like routing overhead computation for data unit, those resources are put under scrutiny which are having number of sent control packets greater than number of data unit packets.
  - *Packet Dropped Ratio:* It is measured for every possible structure. Packet dropped ratio of a structure is number of packets dropped by resources within structure to number of packets received for delivery. If packet dropped ratio of s structure is lesser than average value of its network then those structures are put under scrutiny.
  - *Connection Establishment per Time:* It is defined as total number of control data connection convertible to data connections per unit time. For a resource, it is its capability to accept control packets followed by data delivery per unit time. Higher connection establishment per time is better for inlier consideration. If number of connection establishment for a resource or data unit is lower than

average value of its structure then that resource or data unit is sent for outlier detection process.

- *Concurrent Connection Establishment per Time:* It is defined differently for data unit and resource. For data unit, number of interconnection established at same time indicates it's important at a specific time period. Similarly, number of sessions established by a resource to handle multiple interconnection indicates importance of its services. Thus, if concurrent connection establishment for data unit or resource is lower than average value of its structure then that particular data unit or resource scrutinize under next QoS parameter or next tier outlier detection process.
- *Transactions per Time:* It is measured as number of time a particular data unit or resource is helpful in completing a transaction. A transaction can happen in any of three forms: sending, receiving or forwarding data or control packet. If transaction per time for a data unit or resource is lower than average value of network then that particular data unit or resource is scrutinize under next tier outlier detection process.

Overall process of QoS parameters based outlier detection is explained using pseudocode 3.

Pseudocode 3: Individual QoS parameter scanning and performance evaluation based outlier detection process

```
Each data unit and resource in the network is scanned for QoS based evaluation one by one.
Interconnection of data unit or resource is identified in following order: state, county, cluster and network.
List of inliers and outliers are stored at different places.
Performance of each QoS parameter is evaluated in an order specified above.
**If** any QoS parameter based performance is not met as per pre-defined procedure **then**
    Targeted data unit or resource is stored at outlier location.
**else**
    Targeted data unit or resource is stored at inlier location.
**end if**
**If** number of QoS parameter predicting a data unit or resource as outlier is larger than number of QoS parameter predicting it as inlier **then**
    Targeted data unit or resource is scrutinize for next tier outlier detection process.
**end if**
```

- ***Feature Based Outlier Detection Tier:*** In feature based outlier detection process, features of data unit is extracted including action performed over packet, time of action, starting location of packet transfer, type of protocol used for packet transmission, protocol layer, flag bits indicating control packet, data packet, packet segment in sequence etc., size of packet, size of data transfer, flags indicating source, intermediate or destination information etc. Interconnection of data units is performed and verified before processing it to outlier detection. Verification includes analysis of interconnection of features. For example, an interconnection should involve at least a source address, destination address, control packet and setting of flag bits. Apart from this, it may contain an intermediate node, data packet, subsequent data packet segmentation, miscellaneous flag bit settings etc. Feature based outlier detection tier filtered the data at multiple layers because a data unit may be involved with different protocol performing functionality connected with single process. In this process of data analysis, data is pre-processed, parsed and labeled. Thereafter, data is filtered for different layers. Each of these layers execute outlier detection process with different mechanism. This process of outlier detection at each layer is explained as follows:

- ***Initial Phase Outlier Detection (Layer-1):*** In this phase of outlier detection, analysis is performed using density based clustering and machine learning based outlier detection. This process involves data pre-processing, learning/training, evaluation and prediction. In pre-processing, data is prepared for comparative analysis. This comparative analysis involves state interests including the purpose of interactions, amount and type of data exchanged, and duration of interactions for similar type of data exchange. In an experiment, these features are compared with other structures having similar targets. All those data elements having similar nature are put in training dataset and other elements are put in testing dataset for machine learning based experimentation. Thus, pre-processing phase prepares training and testing dataset for experimentation. Learning phase uses training and testing datasets for context based outlier detection with simulation optimization. Trained data is used directly in comparison process with new data. This mechanisms reduces efforts involved in outlier prediction.
- ***Medium Phase Outlier Detection (Layer-2):*** In this phase, outliers are identified from interconnections among data units with ethical processes. Single data unit may be interconnected with one or more other data units. This interconnection establishes relationship fruitful for network. Thus, each of these relation is helpful in predicting inliers. For example, if there are three data units (DU1, DU2 and DU3) and DU2 is predecessor of DU1, and DU2 is predecessor of DU3 then all data units are claiming to be interconnected. If all data units are connected with a process then there is no chance of presence of outlier else if either of them is not connected then there are chances that this particular data unit is outlier. Similarly, it could be possible that there are two or more predecessor for one data unit or they are indirect predecessor. In all cases, criteria for outlier consideration is that the data units are not interconnected through a process. Subsequently, these interconnections and relationship with processes construct graphs. Each of these graph has either a path or a circuit i.e. a sequence of data units can be generated. Previously, trust values of data units are also computed. If sequence is drawn as per trust values then data units are consider as inliers else outliers. In conclusion, trust value based sequence

verification is validation of sequence generated using graph.

- ***Advanced Phase Outlier Detection (Layer-3):*** In this phase of outlier detection, authenticity of messages and sources is checked for detection. Since complete data is available by the time this phase starts its analysis thus, signatures can be verified through rule based analysis. Identity of source, destination and intermediate devices confirms their authenticity. Whereas, signature verification process verifies message authenticity. A rule based process backtracks if any message or resource identity is found to be suspicious. Rule based analysis possesses verification process using trust path and resource verification. Output of these processes is stored in databases for making a decision to redirect the data unit or resource for an alternative action.

- ***Jump Box:*** The proposed architecture is flexible to be implemented using one outlier detection process executed in one invigilator region or multiple outlier detection processes executed in one or multiple invigilator regions. This flexibility is possible through jump box. This box collect required information from one region and instruct the other region to start its observation with limited or extended data. In conclusion, by passing certain processes may give various advantages like: speedup the outlier detection process, execute outlier detection processes as per availability of hardware or other resources, and relevance of processes.

- ***Outlier Storage System**:* In this storage system, information about outliers is stored. This information contains outcomes of each tier and overall detection of outliers from each invigilator region. This individual and aggregated decisions are stored to provide a provision of accepting outcomes from one or multiple processes. This system will give advantage to those networks where there is scarcity of resources.

- ***Outlier Sharing Gateway System:*** In this gateway system, outcomes and information of outliers across each region is synchronized. This synchronization extends the work to analyze those outliers which are not common across regions. Further, jump box's outlier information is collected as per convenience and it may be incomplete. Thus, it may require to collect information from reliable resources. Gateway system is designed to reliably sharing such information. This system uses time based synchronization.

IV. Results and Analysis

In order to validate the proposed approach, this section uses internal, external and QoS based indices in simulation. A comparative analysis of these indices is performed for identifying the index having best value for a network. Simulation, results and analysis is discussed in detail as follows:

- **Simulation Setup**

In order to analyze the performance, different network scenarios are generated starting from small scale network (5 nodes) to large scale network (6000 nodes.). Table 1 shows the details of simulator parameter used in simulation analysis.

Table 1: Simulation Parameters

Simulator Parameters	Value
Minimum and Maximum number of	5 to 6000
Type of channel used for communication	WirelessChannel
Radio Propagation Model	Ray Tracing
Physical network interface used	WirelessPhy
MAC Type	802.11
Queue maintained at each node	Priority Queue
Signal receiving and transmitting	OmniAntenna
Size of Queue configured at each node	80
Distance on X-dimension	2000 meters
Distance on Y-dimension	2000 meters
Mobility Model	Random WayPoint
Packet rate	10 packets/second
Size of packet transmission (in bits)	1024 bits
Simulator	ns-3[70]
Total simulation duration	2000sec
Minimum and Maximum velocity set for transmission	0.1 m/s to 10 m/s

- **Comparative Performance Evaluation of Indices**

This section explains the performance of indices used in analysis. Performance indicates strength, stability and goodness of structures. As each structure consists of a group of data units or resources, a process indicating its stability is helpful in proper execution of outlier detection process. Further, stability process itself indicates the presence of inliers and outliers. A detailed experimentation study of indices in proposed structures is explained as follows:

- **Comparative Performance Evaluation of Internal Indices**
  Internal indices compared the features of data units present in structures with data units of other structures within same network. In this work, following indices are used for analysis: DI, RMSSDI, RSI, CHI, DBI and SI. Fig. 5 shows performance analysis of DI, RMSSDI, RSI and SI indices for a network consisting of a minimum of 5 and maximum of 6000 nodes. Eleven scenarios are generated to analyze the performance. These scenario consists of networks of 5, 10, 50, 100, 500, 1000, 2000, 3000, 4000, 5000 and 6000 nodes. Results shows that optimal indices value for 5, 10, 50, 100, 500, 1000, 2000, 3000, 4000, 5000 and 6000 nodes are observed during 0 sec. to 200 sec. (slot 1), 0 sec. to 200 sec. (slot 1), 400 sec. to 600 sec. (slot 3), 800 sec. to 1000 sec. (slot 5), 1200 sec. to 1400 sec. (slot 7), 1400 sec. to 1600 sec. (slot 8), 1400 sec. to 1600 sec. (slot 8), 1400 sec. to 1600 sec. (slot 8), 1400 sec. to 1600 sec. (slot 8), 1400 sec. to 1600 sec. (slot 8) and 1400 sec. to 1600 sec. (slot 8) with 2, 4, 5, 17, 26, 34, 40, 40, 51, 52 and 54 clusters respectively. In order to execute internal indices based outlier detection process using pseudocode 3, threshold values for indices are computed and these values for DI, RMSSDI, RSI and SI are 0.381, 0.214, 0.057 and 0.365

respectively.

In individual computations, results show that SI index value decreases for small scale network (5 to 50 nodes) and medium scale network (500 to 3000 nodes) but increases for large scale network (3000 to 6000 nodes). DI and RMSSDI indices values shows decrease for small scale network (5 to 500 nodes) and medium to large scale network (500 to 6000 nodes). However, an increase is observed for small to medium scale network (50 to 100 nodes). In RSI index, a minimum variation is observed as compared to other indices. A decrease for small scale network (5 to 50 nodes) and medium to large scale network (3000 to 6000 nodes) is observed. An increase for small to medium scale network is observed (500 to 3000 nodes).

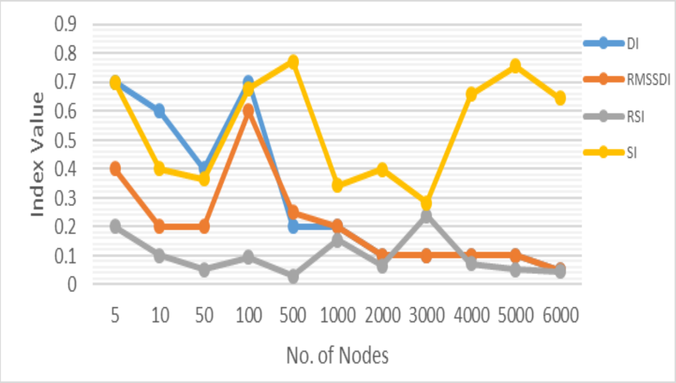

Figure 5: Comparative Performance Analysis of Internal Indices (DI, RMSSDI, RSI and SI)

Fig. 6 shows analysis of two more internal indices CHI and DBI. DBI has shown an increase in index value for small to large scale network (5 to 6000 nodes) whereas, CHI shows an increase for small to medium scale network (5 to 1000 nodes) and medium to large scale network (3000 to 6000 nodes). However, a decrease for medium scale network (1000 to 3000 nodes) is observed in this case. In order to execute internal indices based outlier detection process using pseudocode 3, threshold values for indices are computed and these values for CHI and DBI are 75 and 29.5 respectively.

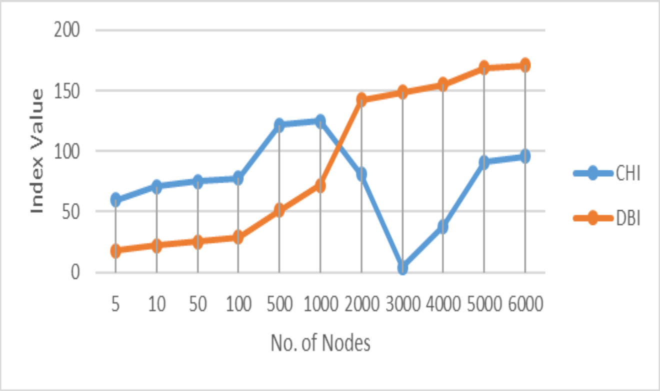

Figure 6: Comparative Performance Analysis of Internal Indices (CHI and DBI)

- **Comparative Performance Evaluation of External Indices**

In external indices, features of data units in one structure and network are compared with features of data unit in another structure and network. Predefined features are preferred for identified as compared to newly identified features [71]. In this work, performance of four external indices is evaluated for 5, 10, 50, 100, 500, 1000, 2000, 3000, 4000, 5000 and 6000 nodes datasets. Results shows that optimal indices value for 5, 10, 50, 100, 500, 1000, 2000, 3000, 4000, 5000 and 6000 nodes dataset are observed during 400 sec. to 600 sec. (slot 3), 400 sec. to 600 sec. (slot 3), 400 sec. to 600 sec. (slot 3), 400 sec. to 200 sec. (slot 3), 200 sec. to 400 sec. (slot 2), 1200 sec. to 1400 sec. (slot 7), 200 sec. to 400 sec. (slot 2), 1400 sec. to 1600 sec. (slot 8), 1600 sec. to 1800 sec. (slot 9), 400 sec. to 600 sec. (slot 3), and 600 sec. to 800 sec. (slot 8) with 4, 8, 18, 16, 5, 33, 12, 41, 26, 45 and 47 clusters respectively. Fig. 7 shows the comparative analysis of FI, NMII, PI and EI's threshold index variation. In conclusion, it is found that increase in index value for external indices lies between 0.65 and 1.

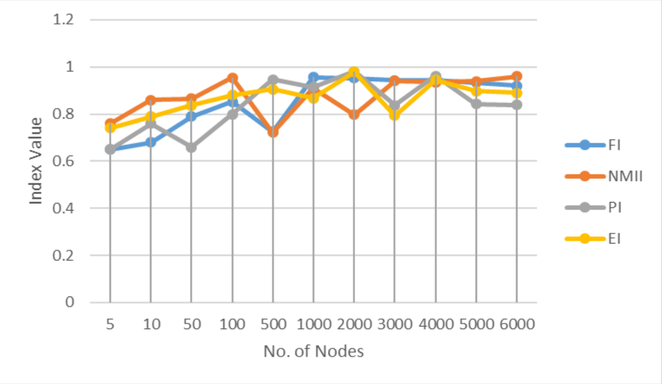

Figure 7: Comparative Performance Analysis of External Indices (FI, NMII, PI and EI)

- **Comparative Performance Evaluation of QoS Indices**

Fig. 8 and fig. 9 shows comparative analysis of two QoS indices: throughput and jitter. Fig. 8 shows the comparative analysis of throughput for 5 to 6000 nodes network. A comparative analysis of throughput with and without presence of outliers shows that the proposed approach is successful in improving the QoS. Maximum improvement is observed for 500 and 1000 nodes network. Least change is observed for 3000 and 5000 nodes network.

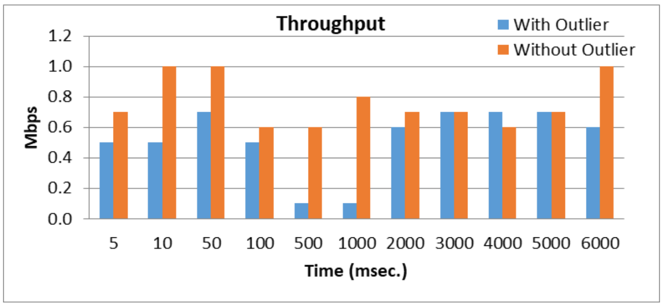

Figure 8: Comparative Performance Analysis of Throughput

Fig. 9 shows comparative analysis of jitter calculations for 5 to 6000 nodes networks with different packet transmission rates. Results show that processes without presence of outliers shows a minimum of 0.15% and maximum of 14.9% improvement in jitter values. Highest jitter is observed for 1000 nodes network with 1 pkt/sec. because of instability in clusters. Minimum jitter is observed for 500 nodes network with 1 pkt/sec. In conclusion, small to medium scale network (500 nodes) with 1 pkt/sec. gives best performance for proposed approach.

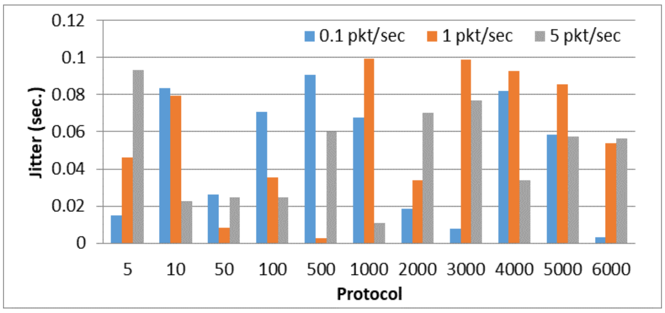

Figure 9: Comparative Performance Analysis of Jitter

V. Conclusion

Outlier detection through multi-layered multi-tiered approach is suitable for network consisting of known resources. It may contains resourceful or resource constraint devices. Purpose of proposed scheme is to integrate trust management with data units and resources and analyze the collected data for presence of outliers. In proposed model, multiple invigilator regions executes similar outlier detection processes over collect data and found that the proposed integrated system is efficient in terms of performance indices. These indices could be internal, external or QoS based. Various internal indices used for measuring the stability of structures are: DI, RMSSDI, RSI, SI, CHI and DBI. Various external indices used for measuring the stability of structures are: FI, NMII, PI and EI. Both internal and external indices confirms the formation of structure and outlier detection processes. Further, two QoS based indices (throughput and jitter) are used in this work. Simulation analysis of proposed scheme over 5 to 6000 nodes network shows that a minimum of 0.15% and maximum of 14.9% improvement is observed in jitter for network without outliers as compared to network with outliers. In throughput computations, it is observed that the proposed approach is successful in improving QoS.

Acknowledgment

The authors would like to thank SEED Division of University of Petroleum and Energy Studies. This work is part of research project (https://sites.google.com/view/adarshproject2/) sponsored by University of Petroleum and Energy Studies (UPES), Dehradun, India under Grant: UPES/R&D/180918/14.

References

[1] V. Barnett and T. Lewis, *Outliers in Statistical Data*, John Wiley Sons, New York, 1994.

[2] A. Kumar and A. Aggarwal. “An Efficient Outlier Detection Mechanism for RFID-Sensor Integrated MANET”. *18th International Conference Intelligent Systems Design and Applications (ISDA 2018)*, December 06-08, 2018, VIT University, Vellore, India.

[3] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogerakiand, and D. Gunopulos, “Online Outlier Detection in Sensor Data using Nonparametric Models”, In *Proceedings of the 32nd international conference on Very large data bases*, (VLDB 2006), pp. 187-198, September 2006.

[4] W. Wu, X. Cheng, M. Ding, K. Xing, F. Liu, and P. Deng. “Localized Outlying and Boundary Data Detection in Sensor Networks”, *IEEE Trans. Knowl. Data Eng.*, 19(8), pp. 1145-1157, 2007.

[5] L.A. Bettencourt, A. Hagberg, and L. Larkey. “Separating the Wheat from the Chaff: Practical Anomaly Detection Schemes in Ecological Applications of Distributed Sensor Networks”. *Proc. IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS 2007)*, pp. 223-239, 2007.

[6] Y. Hida, P. Huang, and R. Nishtala. “Aggregation Query under Uncertainty in Sensor Networks”, 2003. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.58.19&rep=rep1&type=pdf (Accessed: 09 April, 2019).

[7] M.C. Jun, H. Jeong, and C.C.J. Kuo, “Distributed Spatio-Temporal Outlier Detection in Sensor Networks”. *Proc. SPIE*, pp. 273-284, *2006*.

[8] B. Sheng, Q. Li, W. Mao, and W. Jin. “Outlier Detection in Sensor Networks”. *Proc. MobiHoc*, pp. 219-228, 2007.

[9] T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos. “Distributed Deviation Detection in Sensor Networks”, *ACM Special Interest Group on Management of Data*, 32(4), pp. 77-82, 2003.

[10] S. Papadimitriou, H. Kitagawa, P.B. Gibbons, and C. Faloutsos. “LOCI: Fast Outlier Detection using the Local Correlation Integral”. *International Conference on Data Engineering*, pp. 315-326, 2003.

[11] Y. Zhang, N. Meratnia and P. Havinga. "Outlier Detection Techniques for Wireless Sensor Networks: A Survey". *IEEE Communication Surveys & Tutorials*, 12(2), pp. 159-170, 2010.

[12] E. Knorr and R. Ng. “Algorithms for Mining Distance-Based Outliers in Large Data Sets”. *International Journal of Very Large Data Bases*, pp. 392-403, 1998.

[13] S. Ramaswamy, R. Rastogi, and K. Shim. “Efficient Algorithms for Mining Outliers from Large Data Sets”, *ACM Special Interest Group on Management of Data*, 29(2), pp. 427-438, 2000.

[14] J. Branch, B. Szymanski, C. Giannella, and R. Wolff. “In-Network Outlier Detection in Wireless Sensor Networks”. *Proc. IEEE ICDCS*, pp. 1-8, 2006.

[15] K. Zhang, S. Shi, H. Gao, and J. Li. “Unsupervised Outlier Detection in Sensor Networks using Aggregation Tree”. *Proc. ADMA*, pp. 158-169, 2007.

[16] Y. Zhuang and L. Chen. “In-Network Outlier Cleaning for Data Collection in Sensor Networks”. *Proc. VLDB*,

pp. 23-54, 2006.

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

[18] S. Rajasegarar, C. Leckie, M. Palaniswami, and J.C. Bezdek. “Distributed Anomaly Detection in Wireless Sensor Networks”. *Proc. IEEE ICCS*, pp. 1-5, 2006.

[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*, 54(4), pp. 570-588, April 2011.

[20] S. Rajasegarar, C. Leckie, M. Palaniswami, and J. C. Bezdek. “Quarter Sphere Based Distributed Anomaly Detection in Wireless Sensor Networks”. *Proc. IEEE International Conference on Communications*, pp. 3864-3869, 2007.

[21] D.J. Hill, B.S. Minsker, and E. Amir. “Real-Time Bayesian Anomaly Detection for Environmental Sensor Data”. *Proc. 32nd Congress of the International Association of Hydraulic Engineering and Research*, pp. 1-10, 2007.

[22] D. Janakiram, A. Mallikarjuna, V. Reddy, and P. Kumar. “Outlier Detection in Wireless Sensor Networks using Bayesian Belief Networks”, *Proc. IEEE Comsware*, pp. 1-6, 2006.

[23] E. Elnahrawy and B. Nath. “Context-Aware Sensors”. *Proc. EWSN*, pp. 77-93, 2004.

[24] V. Chatzigiannakis, S. Papavassiliou, M. Grammatikou, and B. Maglariset. “Hierarchical Anomaly Detection in Distributed Large-Scale Sensor Networks”. *Proc. ISCC*, pp. 1-6, 2006.

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

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

[27] 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, pp. 1779-1782, May 2012.

[28] V. Chandola, A. Banerjee and V. Kumar. "Anomaly Detection: A Survey”. *ACM computing surveys*, 41(3), pp. 1-72, 2009.

[29] X. Luo, M. Dong and Y. Huang. "On distributed fault tolerant detection in wireless sensor networks", *IEEE Transactions on computers*, 55(1), pp. 58-70, Jan. 2006.

[30] 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, pp. 65-72, Sep. 2006.

[31] 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, pp. 722-727, December 2015.

[32] B. Krishnamachari and S. Iyengar. "Distributed Bayesian algorithms for fault tolerant event region detection in wireless sensor networks", *IEEE Transactions on Computers*, 53(3), pp. 241-250, March 2004.

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

[34] 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, pp. 902-913, March 2005.

[35] 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, pp. 16-23, Oct. 2005.

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

[37] W. Hu, T. Tan, L.Wang, S. Maybank. "A survey on visual surveillance of object motion and behaviors", *IEEE Transaction* on Systems, Man, and Cybernetics, XXXIV (3), pp. 334-352, 2004.

[38] P. Gogoi, B. Borah, D. K. Bhattacharyya. "Anomaly Detection Analysis of Intrusion Data using Supervised and Unsupervised Approach", *Journal of Convergence Information Technology,* V (1), pp. 95-110, 2010.

[39] V. A. Traag, A. Browet, F. Calabrese, F. Morlot. "Social Event Detection in Massive Mobile Phone Data Using Probabilistic Location Interference". In *Proceedings SocialCom/PASSAT*, pp. 625-628, 2011.

[40] K. Fisler, S. Krishnamurthi, L. A. Meyerovich, M. C. Tschantz. "Verification and change-impact analysis of access control policies". In *Proceedings of 27th International Conference on Software Engineering,* pp. 196-205, 2005.

[41] D. Jackson. *Software Abstractions: Logic, Languages, and Analysis*, MIT Press, ISBN: 978-0-26210114-1, 2006.

[42] 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/alloyref.pdf. [Accessed: Jan. 1, 2018].

[43] D. Jackson. "Alloy: a lightweight object modelling notation", *ACM Trans. Soft. Eng. Methodol.,* XI (2), pp. 256-290, 2002.

[44] A. Kumar, A. Agarwal, Charu. "Efficient Hierarchical Threshold Symmetric Group Key Management Protocol for Mobile Ad Hoc Networks". In *Proceedings Inter. Conf. on Contemporary Computing (IC3-2012),* pp. 335-346, 2012.

[45] J. C. M. Teo, 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, pp. 114-121, 2005.

[46] H.A. Wen, C.L. Lin, T. Hwang. "Provably secure authenticated key exchange protocols for low power computing clients". *Journal of Computers and Security*, XXV (2), pp. 106–113, 2006.

[47] Y.M. Tseng. "Efficient authenticated key agreement protocols resistant to a denial of service attack", *International Journal of Network Management*, XV (3), pp. 193–202, 2005.

[48] A. Kumar, K. Gopal, A. Aggarwal. "Outlier Detection and Treatment for Lightweight Mobile Ad Hoc Networks". In Proceedings International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, pp.750-763, 2013.

[49] N. Chugh, A. Kumar, A. Aggarwal. "Availability aspects through optimization techniques based outlier detection mechanism in wireless and mobile networks", *Int. Journal of Computer Networks and Communications (IJCNC),* X (6), pp. 77-96, 2018.

[50] A. Kumar, A. Aggarwal. "An efficient simulated annealing based constrained optimization approach for outlier detection mechanism in RFID-sensor integrated MANET". In Proceedings Inter. Conf. on Intelligent Systems Design and Applications (ISDA), 2018.

[51] A. Kumar, A. Aggarwal, D. Yadav. "A Multi-layered Outlier Detection Model for Resource Constraint Hierarchical MANET". In *Proceedings 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON),* 2018.

[52] G. Usha, M. Rajesh Babu, S.S. Kumar. "Dynamic anomaly detection using cross layer security in MANET", *Journal of Comput. Electr. Eng.,* XXXXXIX, pp. 231–241, 2017.

[53] A. Amouri, S. Morgera, M. Bencherif, R. Manthena. "A Cross-Layer, Anomaly-Based IDS for WSN and MANET". *Journal of Sensors,* XVIII (2), p. 651, 2018.

[54] I. Butun, S. D. Morgera, R. Sankar. "A Survey of Intrusion Detection Systems in Wireless Sensor Networks", *IEEE Commun. Surv. Tutorials,* XVI (1), pp. 266–282, 2014.

[55] L. Nishani, M. Biba. "Machine learning for intrusion detection in MANET : a state-of-the-art survey", *Journal of Intell. Inf. Syst.,* XXXXVI (2), pp 391–407, 2016.

[56] A. Amouri, V. T. Alaparthy, S. D. Morgera. "Cross layer-based intrusion detection based on network behavior for IoT". In *Proceedings IEEE 19th Wireless and Microwave Technology Conference (WAMICON),* pp. 1–4, 2018.

[57] M. A. Hayes, M. A. Capretz. "Contextual anomaly detection framework for big sensor data", *Journal of Big Data,* II (2), pp. 1-22, 2015.

[58] R. Agrawal, T. Imieliński, A. Swami. "Mining association rules between sets of items in large databases". In *Proceedings of the International Conf. on Management of Data,* pp. 207-216, 1993.

[59] M. Hahsler, R. Karpienko. "Visualizing association rules in hierarchical groups", *Journal of Bus. Econ.,* XXXXXXXXXVII (3), pp. 317–335. 2017.

[60] P. J. Rousseeuw. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis", *Journal of Computational and Applied Mathematics, XX, pp. 53-65,* 1987.

[61] J. C. Dunn. "Well-separated clusters and optimal fuzzy partitions", *Journal of Cybern.,* IV (1), pp. 95–104, 1974.

[62] D. L. Davies, D. W. Bouldin. "A Cluster Separation Measure", *IEEE Trans. Pattern Anal. Mach. Intell.,* PAMI-1 (2), pp. 224-227, 1979.

[63] T. Caliñski and J. Harabasz, "A Dendrite Method Foe Cluster Analysis", *Journal of Commun. Stat.,* 1974.

[64] D. Moulavi, P. A. Jaskowiak, R. J. G. B. Campello, A. Zimek, J. Sander. "Density-Based Clustering Validation". In *Proceedings of the SIAM International Conference on Data Mining,* pp. 839-847, 2014.

[65] "Evaluation of clustering." [Online]. Available: https://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-clustering-1.html#fig:clustfg3. [Accessed: 05-Jul-2018].

[66] Y. Liu, Z. Li, H. Xiong, X. Gao, J. Wu, S. Wu. "Understanding and enhancement of internal clustering validation measures", *IEEE Trans. Cybern.,* XXXXIII (3), pp. 982-994, 2013.

[67] F. Kovács, C. Legány, and A. Babos. "Cluster Validity Measurement Techniques." In *Proceedings 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases,* pp. 388-394, 2006.

[68] S. Huang, Y. Cheng, D. Lang, R. Chi, G. Liu. "A formal algorithm for verifying the validity of clustering results based on model checking", *Journal of PLoS One,* IX (3), pp. 1-14, e90109, 2014.

[69] S. Gurung, S. Chauhan. "A dynamic threshold based approach for mitigating black-hole attack in MANET", *Journal of Wirel. Networks,* XXIV (8), pp. 2957-2971, 2018.

[70] "The Network Simulator - ns-2." [Online]. Available: https://www.isi.edu/nsnam/ns/. [Accessed: 05-Jul-2018].

[71] T. Van Craenendonck, K. Leuven, H. Blockeel. "Using Internal Validity Measures to Compare Clustering Algorithms". In *Proceedings of AutoMLWorkshop ICML,* pp. 1-8, 2015.

Author Biographies

Dr. 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-Doc from SRI, 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, Uttar Pardesh, 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 cybersecurity, cryptography, network security, and ad-hoc networks.

Dr. Alok Aggarwal received his bachelors' and masters' degrees in Computer Science & Engineering in 1995 and 2001 respectively and his Ph.D degree in Engineering from IIT Roorkee, 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.