Provable, Reliable And Secure Data Aggregation Through Integrated Distributed Mechanism In Iot Based WSN Environment

Mr. Bharat Kumara[1] , Dr. S Anantha padmanabhan[2]

[1]Research scholar, dept of ECE Gopalan College of engineering and Management Bangalore .

[2]Prof & head Dept of ECE, Gopalan College of engineering and management.

Abstract

Internet if Things aka IoT have made human life easy and comfortable since it possesses application in almost every aspect of human life. Moreover, few major application includes agriculture, healthcare, military services and so on. Moreover, WSN (Wireless Sensor Network) is an integral part of IoT based application which sense the data and transmit to the base station or sink node. However, since WSN possesses a restricted environment and also generates huge amount of data and further causes the data redundancy. Although data redundancy is efficiently solved through the various data aggregation mechanism, security remains a primary concern for adoptability in the real time environment. Hence, in this research work we design and develop Integrated Mechanism; the main aim of this mechanism is to provide provable, reliable and secure data aggregation. Integrated Distributed Mechanism aka IDM is integration of three modules in a distributed manner. First module of IDM includes the secure and efficient data aggregation; second module includes designing of misclassified data aggregation and third module includes identification and discarding of dishonest data packets and further nodes. IDM is evaluated considering the various parameter such as non-functioning nodes, energy utilization, packet identification and packet misclassification; further comparative analysis proves that IDM marginally outperforms the existing model

Keywords: Security, Data Aggregation, WSN, Integrated distributed Model, secure data aggregation

1 Introduction

IoT (Internet of Things) has aided numerous advantage in sensor that has led the enormous growth in sensor based area such as healthcare, agriculture, management and many more sectors; Moreover, WSN has been integral part of IoT; moreover IoT provides the intelligent services to these sensor nodes. There are different types of sensor based on the application; also different condition such as suburban, rural, non-rural, underwater areas can be exploited through the WSN through data collection. Furthermore, WSN possesses various advantage of being light edge device, low cost and ease of deployment which has shown in development of smart health care monitoring device, smart wearable devices, smart vehicles, intelligent buildings, smart cities. However, these application faces several challenges which is current trend of research; moreover, research problem is discussed later in same section.

Figure 1 shows the typical data transmission in designated WSN environment; it comprises CH (Cluster Head), BS (Base Station aka sink) and sensor node. Data is sensed through sensor node and transmitted to the CH (Cluster Head); further these data are aggregated at cluster head and sent to sink. Moreover, the amount of data sensed is large and complex in nature that makes the storage and processing. Another major issue is data redundancy which can be solve through data aggregation [5]-[7]. Normal data aggregation includes MIN, MAX and COUNT .

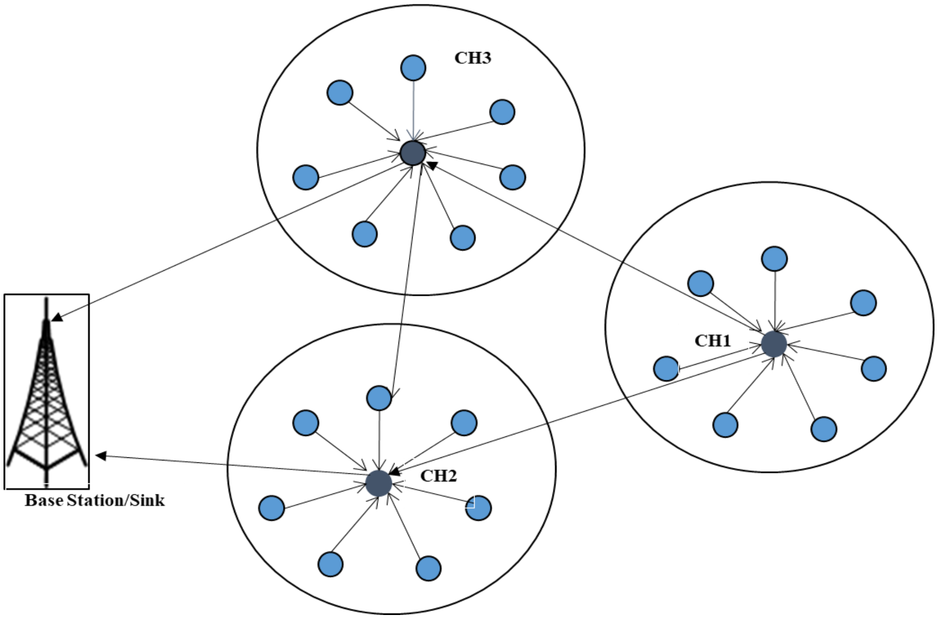

Figure 1 Data Transmission in WSN environment

The method used for the reduction of the energy intake and to removal of the redundant data is data aggregation technique. In the process of data aggregation, the sensor nodes are arranged as a tree, root being the base station. The transitional sensor nodes aggregate the data arriving from the leaf nodes and then pass on the aggregated output to the root, base station. Yet, sometimes this process gives some issues in few applications, namely, remote healthcare observing systems. The sensors nodes are frequently positioned in an unfriendly environment which has minimal bandwidth and uncertain communication channel [8]. This might allow for hostile data alterations and forgery of the data, emerging in the infringement of the privacy of the user. For instance, an intruder might counterfeit a duplicate alarming reading and share it the whole network to devalue the performance of the network. Furthermore, privacy also should be taken care of in remote healthcare observing systems, as the data which we get from the system is only pertinent to the respective patient who is observed. E.g. motion sensors can show some nature of the patient such as walking, having a meal or sleeping etc. Revelation of such health related data lead to repercussion as it breaches the privacy of the patient. So, efficient way of aggregating various data and to store the patient data safely is a vital work in a limited resource environment. Due to the location, transmission media, reduced physical protection the WSNs can be risked to multiple attacks [9]. There are different types of attacks that can influence the network during the process of data aggregation, some of them are as follows:

A. Negation of the service: This attack is often called as the Jamming attack as this sends the radio signals for blocking the frequencies which are used by the WSN. Most of the part of the network can be affected if the magnitude of the opponent increases. This attack can cause the aggregator node during the aggregation process to restrict the data from moving to the upper levels. Node compromise attack is another name of this attack.
B. Supervision attack: during this particular attack, all the data stored in the node is removed from the node by the opponent. If the node is seized with this particular attack, whole protected data will be removed from the node during the data aggregation process.
C. Sybil attack: during this type of strike, the intruder has the ability to create multiple recognitions inside the network. The data aggregation is influenced because of this attack in different ways, namely, 1) to begin with, the intruder will cause more than one recognition for producing extra votes for the aggregator election scenario and selects a hostile node as the aggregator. If the opponent is allowed to create multiple entries with non-identical readings, the outcome received after aggregation would be unscrupulous. 2) The opponent can initiate this Sybil attack and create n or multiple recognitions. This forces the base station to receive the outcome of aggregation.
D. Attack of selective forwarding: the refusal of any compromised node to transmit the obtained message as the opponent will the control over that compromised node and it might command the node not to forward the message or refuse the message. The aggregation outcomes are influenced by these types of attacks.
E. Attack with replay: the intruder would track the network traffic without even the network awareness and would replay the same in the later interval to confuse the aggregator, which in turn affects the outcome from the aggregator.
F. Secret attack: the introduction of wrong data in the network without disclosure of its existence is this attack. As the introduced wrong data is also included in the aggregation process, the outcomes of the aggregation would alter.

This research is carried out in four section, first section starts with background of WSN and IoT along with their integrated application. Further, problem of WSN related to data redundancy and secure data aggregation is discussed; first section ends with motivation and contribution of research. Second section performs discuss various related work aim to secure data aggregation; also their shortcomings are highlighted. IDM (Integrated distributed Mechanism) is designed and developed in third section along with three different modules; performance evaluation of proposed IDM is carried out in fourth section of this research along with comparative graphs and analysis.

1.1 Motivation and contribution of our research work

WSN have attracted research and academia industry in past decades and application of same has been in every area of human life; however due to dense deployment of sensor nodes there is chance of data redundancy that further causes high energy consumption. This is solved through the data aggregation; Data aggregation possesses dynamic characteristics and has wide range of application that provides the flexibility in designing the energy efficient model. Thus motivated by the application, low cost deployment along with security concern, we design and develop secure and efficient model of data aggregation that provides the data security. Further contribution of research work is given as:

1. This research work design and develops Integrated Distributed mechanism aka IDM; IDM is integration of three different mechanisms to provide provable, reliable and secure data aggregation.
2. First Module of IDM focus on the efficient data aggregation that can utilize energy in efficient manner, second module performs the misclassified data packets and third module is designed for identification of dishonest data packets and further discarding the data packets.
3. IDM is evaluated with wide variety of parameter i.e. non-function nodes, energy utilization, correct and misclassified packet identification of honest and dishonest data packets; also comparison has been carried with the existing model and comparative analysis is carried out on different malicious packets.

2 Related Work

The constructive method of combining the redundant data to high standard data is Data aggregation. This method also preserves energy and reduces the bandwidth usage which in turn increases the system's total lifecycle. On the other hand, privacy protecting data aggregation has become a research interest as it confirms the privacy of the vital data during the process of aggregation. The WSN data aggregation methods are analyzed by the researchers; [10] The edge computing endued IoT with data aggregation has pulled enormous observation with multiple works proposed on them. MDC route, which determines the convex hulls in the route design of the data aggregation. In [11], the MDC routes are shortened with the relay devices and the same is delivered in the paper. In [12] the location of the hyperedges in the hypergraph are determined using the Delaunay triangulation. CISIL is designed in way to determine the MDC route. In [13], MDC load balance optimization and greedy expansion are used by LEEF. In [14], by the usage of the star topology, the latency is systematically decreased and the improvement in the load balance can be observed. In [15], MDC routes are split into triangles to achieve improvement in the load balance. In [16], the construction of the MDC routes are done depending on the convex hulls and condensed more by making use of the center of mass. On the contrary, energy cost is reviewed by the impact of the realistic environment by the reviewer [17][18].

In [19], MDC route is located by utilizing the terrain impact for the reduction of the energy cost. In [20], in the realistic environment the relay devices and the MDC's are employed. In [21], in data aggregation both the obstacle avoidance and the collision avoidance are visualized. In [22], the issue of the privacy protecting data aggregation is examined in circumstances of cyber-physical model. In [23], Radial Bias Function Neural Network (RBFNN) prognosticates the speed of the data aggregation process. This aids to model the energy saving MDC routes. In [24], the modelling of an event distribution depending on the data aggregation model with the 3D environment review is done based on Deep Q-Network (DQN). In [25], the energy conscious data analysis and the data aggregation make use of the reinforcement learning method. In [26], Deep Reinforcement learning (DRL) dependent method is used to unite MDC route design and blockchain. Although, the above mentioned method will reduce either energy cost or the ratio of aggregation. In [27], The Blockchain is taken into account as the new privacy preserving tool. Hence, this is deployed to get the privacy protecting data aggregation. In [28], in the user privacy preserving data aggregation the blockchain's anonymous behavior is made use of. In [29], a structure which is not centralized depending on the blockchain is delivered, called as CrowdBC. The registration of the end-users are done without the actual identity and the vital data is reserved in a shared storage. In [30], for the electrical information accumulation, privacy and protection concerns are contemplated by deploying the blockchain and the edge computing mechanisms in smart grid. In [31], blockchain dependent shared cloud framework is modelled with the fog nodes, which are entitled software explained networking. Nevertheless, these mechanisms will not reach the security expectations depending on the data aggregation because of the conventional block header model and the block generation procedures. In [32]A secured and energy conscious data aggregation execution offloading for fog aided IoT networks. Especially, development of this 3 layer protected, fog-aided design is done to respond to a security threats coming and start the aggregation process like cipher text can be done. In the meantime, to optimize the total energy intake of the execution methods, a momentum descent dependent energy saving offloading algorithm is modelled. This can accomplish the minimal value with increased convergence speed. Ultimately, the protection and computation based examinations disclose the modelled data aggregation method is a efficient data processing mechanism and it accomplishes notable lead in the energy intake process. Initially, a group-dependent key establishment method was delineated. Meters were classified to groups, meters in that category develop the keys to encrypt the information, and the issue of the failure of the meter is attenuated. Other groups will not be influenced if there are damaged meters in one category. Furthermore, by permitting the meters for updation of their keys, multiple process like dynamic join, leave and replacement of the meter techniques are delineated. Moreover, apart from the above mechanism few mechanism that were managed to provides the secure data aggregation are discussed in [32]-[34]; in [32], the designed mechanism are able to verify itself and reliability is proved, however due to its complexity and limited scope of adoption makes it restricted. Similarly [33] used the compressive sensing based data aggregation security approach, where the much focus was given to compressive sensing and security lacks behind. Other work as [34] is designed for IoT based application for solving the complexity of the above model; although approach seems to be adoptable but it is restricted and lacks the dynamic adoption.

Above secure data aggregation in WSN lacks uniformity and cannot be adopted for dynamic purpose; also few methods ignored energy and lifetime constraint. Another major question raises is even if there is good data aggregation in above related work, protected data aggregation can be executed. Hence, considering the above shortcoming, this research work develops IDM (Integrated Distributed Mechanism) to provide integrated mechanism in distributed way.

3 Proposed Mechanism

IoT based WSN has enormous application where sensors and IoT based devices interact in data transmission without any involvement of human; however due to the vulnerability of domain and the sensors, secure data transmission has to be reviewed and there is a need of model where the data can be transmitted as well as the dishonest aka malicious data packets along with their node can be identified. Hence in this section of the research we design and develop Integrated Mechanism for reliable and verifiable secure data aggregation; figure 2 shows the IDM workflow which comprises five distinctive blocks. At first the research preliminaries are initialized along with network design such as node placement, cluster Head selection. Second block refers to designing the mathematical model for efficient and secure data aggregation through our designed formulation to utilize energy. Third block computes the misclassified data aggregation i.e. how many honest data packets are identified as malicious (also referred as dishonest data packets) and vice-versa. In fourth block, we design the particular condition that distinguish the malicious node and honest node, also energy utilization Is carried out. Further based on the condition designed, fifth blocks discard the malicious node.

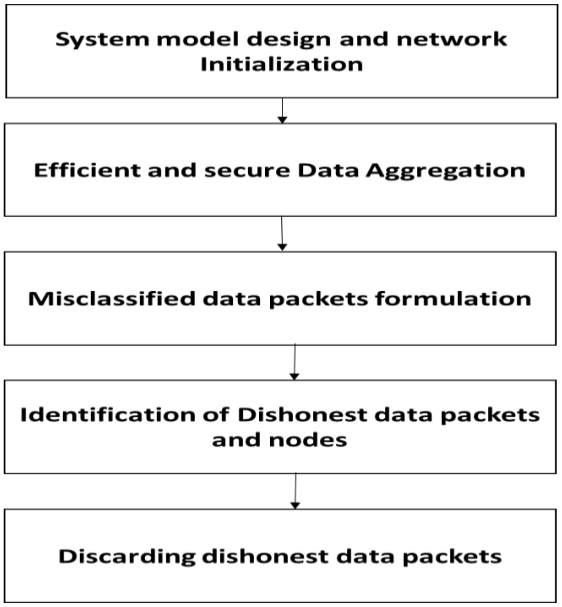
System model design and network Initialization
Efficient and secure Data Aggregation
Misclassified data packets formulation
Identification of Dishonest data packets and nodes
Discarding dishonest data packets

Figure 2 Integrated Distributed workflow

3.1 System model

Let's consider a model which comprises a number of users i.e. $A = \{a_1, a_2, \ldots, a_r\}$ and cluster head ; all R have data as $B = \{b_1, b_2, \ldots, b_r\}$ to Cluster Head. In here, each node are rated as malicious node, normal node or the honest node with respect to $b_m$defined reputation; master computes the average as given as:

$$C = R^{-1} \sum_{m=1}^{R} b_m \quad (1)$$

Further, we consider the untrusted DIC (Data Information Center) and two type of nodes are considered named honest node and malicious nodes and considers two distinctive threats, first threat is aggregated data quality and second one is comprising the privacy.

$$C^o = \sum_{m=1}^{\ddot{q}} a_m^o y_m^o \quad (2)$$

3.2 Optimal and secure data aggregation

Let's consider the aggregated data as z_k and additional noise Θ_k for achieving the optimal security for the data; further we formulate the final data which is the added noise and the sensed data from wsn and given as:

$$z_{\ddot{k}} = z_k + \Theta_k \quad (3)$$

In above equation, additional noise is nearly approximate to the $P(P, I^2)$; moreover considering the approximation above equation can be formulated as:

$$C^{o} = \sum_{m=1}^{\ddot{r}} a_m^{o} y_m^{o} \tag{4}$$

P' indicates the authentic data among all the whole data; further integration of additional noise is carried out through using random generation function. Moreover the designed random function is given as $O(.)\colon U$ to the original sensed data and further it can be given as $z_{\ddot{k}} = O(z_k)= z_k + \Theta_k$. Once we design the manipulated which is combination of original sensed data and additional data, it becomes highly improbable for identifying the original data. However, an underlying problem was found i.e. tradeoff between the privacy and accuracy. In order to get rid of this, we initialize Integrated Mechanism parameter denoted as ζ. Integrated approach parameter is computed through given equation.

$$\zeta = [C - \ddot{C}] \tag{5}$$

Through the parameter definition and analysis, it is observed that nominal value of parameter possesses better data aggregation accuracy. Moreover with the given number of participants as b_m with additional parameterΞ_k and weights a_m, we are able to compute the parameter that can be proved. In order to prove, let's consider the original sensed data (without additional information) which is given as:

$$C = \sum_{m=1}^{\ddot{r}} a_m b_m \tag{6}$$

Further additional data can be given as:

$$C' = C + \sum_{m=1}^{\ddot{r}} a_m \Theta_k \tag{7}$$

Further, it is observed that Θ_k is equal to $2I_m^2$; thus considering it the equation can be modified as:

$$X(\Theta_k) = 2\sum_{k=1}^{\ddot{r}} (a_k^{o})^2 \, (I_m^2) \tag{8}$$

Further, considering the probability distribution for for all μ is greater than zero, parameter probability can be given as:

$$\text{probable}[\ddot{C} - C] \geq \frac{2}{\zeta^2} \sum_{j=1}^{R'} (a_m^{o})^2 \, (I_m^2) \tag{9}$$

Thus, we substitute $\text{probable}[\ddot{C} - C] < \mu$ is equal to r in above equation and we get

$$\frac{2}{\zeta^2} \sum_{m=1}^{P'} (a_m^{o})^2 \, (I_m^2) = 1 - r \tag{10}$$

Further the optimal security accuracy parameter is given as:

$$\zeta = (\sqrt{2})\ (1-\mathrm{r})^{-1/2}\left(\sum_{\mathrm{j}=1}^{\mathrm{P}'}(\mathrm{a}_{\mathrm{m}}^{\mathrm{o}})^{2}\ (\mathrm{I}_{\mathrm{k}}^{2})\right)^{-1/2} \tag{11}$$

Moreover, since $\mathrm{I}_{\mathrm{k}}^{2} = \frac{\beta^{2}}{2\Xi_{\mathrm{k}}}$, replacing this we can get back the parameter initialized and given as:

$$\zeta = \beta\ (1-\mathrm{r})^{-1/2}\left(\sum_{\mathrm{k}=1}^{\mathrm{P}'}\ (\mathrm{a}_{\mathrm{m}}^{\mathrm{o}})^{2}\frac{1}{2\Xi_{\mathrm{k}}}\right)^{-1/2} \tag{12}$$

Furthermore, we tend to minimize the initialized parameter, as discussed earlier that nominal value of parameter gets better accuracy model; hence this can be formulated as optimization problem and it is defined through below equation

$$\min_{\Xi_{\mathrm{k}},\ \tau_{\mathrm{k}}}\sum_{\mathrm{k}=1}^{\mathrm{P}'}\frac{(\mathrm{a}_{\mathrm{m}}^{\mathrm{o}})^{2}}{\Xi_{\mathrm{k}}} \tag{13}$$

Such that $\tau_{\mathrm{k}} - \mathrm{T}_{\mathrm{k}}(\Xi_{\mathrm{k}})^{2}$ is greater than or equal to zero

$$\sum_{\mathrm{k}=1}^{\mathrm{P}'}\ \tau_{\mathrm{k}} \leq \mathrm{D}$$

pa_{i} is greater than or equal to zero; also Ξ_{k} is greater than zero

Further the above optimization problem is formulated through below equation.

$$\Xi_{\mathrm{k}}' = \left(\frac{(\mathrm{a}_{\mathrm{m}}^{\mathrm{o}})^{2}\ \ (\mathrm{T}_{\mathrm{k}})^{2/3}}{\sum_{\mathrm{l}=1}^{\mathrm{P}'}(\mathrm{a}_{\mathrm{m}}^{\mathrm{o}-1})^{2}\ \ (\mathrm{T}_{\mathrm{k}})^{1/3}}\right)^{-1/2}\mathrm{D} \tag{14}$$

Also considering the earlier equation, we modify the equation and nominal value of parameter is formulated as:

$$\tau_{\mathrm{k}}^{\wedge} = \left(\frac{(\mathrm{a}_{\mathrm{m}}^{\mathrm{o}})^{2}\ \ (\mathrm{T}_{\mathrm{k}})^{2/3}}{\sum_{\mathrm{l}=1}^{\mathrm{P}'}(\mathrm{a}_{\mathrm{m}}^{\mathrm{o}-1})^{2}\ \ (\mathrm{T}_{\mathrm{k}})^{1/3}}\right)^{-1/2}\mathrm{C} \tag{15}$$

Moreover through the earlier assumption in same section, we observed that $\sum_{\mathrm{i}=1}^{\mathrm{O}'}\ \mathrm{pa}_{\mathrm{i}} \leq \mathrm{B}$, hence

$$\sum_{\mathrm{k}=1}^{\mathrm{P}'}\mathrm{T}_{\mathrm{k}}\ \Xi_{\mathrm{k}}^{2}\ = \mathrm{D} \tag{16}$$

Moreover, we optimize the designed problem through taking the function i.e. $\mathrm{L}{:}\ \mathrm{R}^{\mathrm{N}} \times \mathrm{R} \rightarrow \mathrm{R}$ and can be given as:

$$N(\Xi_k, \Psi) = \sum_{K=1}^{P'} (a_m^o)^2 \frac{1}{\Xi_k} + \sum_{i=1}^{O'} \Psi \left(T_k\, \Xi_k^2 - D\right)^2 \tag{17}$$

In the above equation, Ψ indicates the multiplier; further we take the derivatives of N considering the nominal Ξ_k and given as:

$$\frac{\partial N}{\partial \Xi_k} = \frac{(a_m^o)^2}{\Xi_k^2} + 2\ \Psi\, T_k \Xi_k = 0 \tag{18}$$

Above equation can be formulized through given below equation:

$$\Xi_k = \left(\frac{(a_m^o)^2}{2\ \Psi\, T_k \Xi_k}\right)^{1/3} \tag{19}$$

Further, we substitute the 15 into 12

$$\left(\frac{1}{2\,\Psi}\right)^{1/3} = \frac{D}{\sum_{l=1}^{P'} (a_m^o)^2\ (T_k)^{1/3}} \tag{20}$$

3.3 Misclassified data packet computation

Any data aggregation can be reliable and it can be validated one and only if it has the quality of the data aggregation; hence we introduce the mechanism for detecting the dishonest node. This can be carried out considering the below process.

Let's consider the J_0 as any data aggregation parameter for the efficiency where as J_1 indicated the non-efficient and thus we can formulated the packet misclassified i.e. number of nodes that are honest but judged as the honest.

$$R_h = R(J_1 | J_0) \tag*{[21]}$$

Moreover the misclassification rate can be denoted as P_m where the dishonest packet are considered as the honest one; this can be formulated as:

$$R_m = R(J_0 | J_1) \tag{22}$$

Accordingly, we design the test static and formulated as:

$$N = \|\ z_k^m - \hat{z}_k^m\ \|^2 \tag{23}$$

The above equation provides the deviation among the defined two term in above equation; let's consider the data $z_k = (z_k^1, z_k^2, \ldots\ldots, z_k^p)$. Further, we design the test for misclassified and classified packets.

$$N \lesseqgtr_{J_1}^{J_0} (\vartheta) \tag{24}$$

Thus if the data is absolute then the participants are updates else it is discarded which can be represented through below equation

$$z_k^m \leftarrow z_k^m \tag{25}$$

Else

$$z_k^m \leftarrow z_k^{m-1}$$

Further, let's consider an energy parameter where the energy constraint of absolute packets is denoted through $\mathbb{I}_1$ and dishonest packets is denoted through $\mathbb{I}_0$ where $\mathbb{I}_1 > \mathbb{I}_0 > 0$; further let's consider the attack probability parameter which can be denoted as p and probable risk are denoted as $R(\vartheta, R)$ thus any attack risk are computed through below equation

$$T(\vartheta, \mathbf{r}) = \left(\mathbb{I}_1\left(1 - R_h(\vartheta)\right) - ER_h(\vartheta)\right)\left(1 - \sum_{k=1}^{P_o} r_k\right) + \left(\mathbb{I}_0\left(1 - R_o(\vartheta)\right) - ER_h(\vartheta)\right) \\ - ER_h(\vartheta)) \sum_{k=1}^{P_o} r_k \tag{26}$$

Further considering the cluster utility as $vh_e(\vartheta, R)$, it can be observe that $vh_e(\vartheta, R) = R(\vartheta, R)$

3.4 Malicious (dishonest node) detection and discard

WSN is considered to be most vulnerable network, hence detection solely is not going to solve problem, it is also required to be discarded; moreover, identification and discarding of these nodes are carried out through sensor updation in mth time and parameter is presented through t_k^m; initial parameter is denoted through a_m^o. Furthermore, deviation among the packets and its reference $func_{Mx} = {}_m^{Mx}\{func_m\}$ can be given through $func^{\cdot\cdot}$. Packet discarding can be carried out through below condition.

const1: if $func_m$ is either equal or less than $func_{Mx}$ then updation increases and value of $func_m$ decreases; thus node reliability increases if node provides

Const2: if $func_m$ is greater than $func_{Mx}$, then the reliability is reduced and increase in value of $func_m - func_{Mx}$; this suggest that if one of the node in WSN sends the dishonest packet then parameter value reaches to null and hence it is discarded; updation function is given as:

$$w_m^o = w_m^{o-1} + 2/(\varkappa(func_k - func_{\ddot{M}x}) + 1) \times \left(1 - u_k^{o-1}\right) \times \mathfrak{e}\{-\mathcal{F}func_m\} \\ + 2/(\varkappa(func_m - func_{\ddot{M}x}) - 1) \times (1 - \mathfrak{e}\{-\mathfrak{I}(func_m - func_{Mx})\}) \tag{27}$$

In equation 27, $\mathfrak{I}$ and $\mathcal{F}$ are real numbers used for scaling the w_m^o; moreover these two parameters are evaluated through monitoring and updating the node value. Further, we use the honest data for weight computation; thus let's consider the $\mathbb{B}' = \{\mathbb{b}_1, \mathbb{b}_2, \ldots\ldots, \mathbb{b}_r\}$ which denotes the honest data verified through the above condition; thus average aggregation weights are given as;

$$a_m^o = \beth_m^o \left(\sum_{m=1}^{R\prime} \beth_m^o\right)^{-1} \tag{28}$$

The above equation provides the guarantee for the honest data as honest data is considered for computing the average aggregation

4 Performance Evaluation

WSN has been adopted widely and generates huge data and causes data redundancy which can be solved through data aggregation, however providing secure data aggregation is major task; this research work develops IDM for provable, secure and reliable data aggregation. This section of research evaluates IDM (Integrated Distributed Mechanism) on windows10 platform with using visual studio 2017 using sensoria simulator [35]; system architecture includes 8GB CUDA enabled RAM and 2TB of hard disk; In order to design prototype, C# is used as a programming language.

Moreover, IDM model considers 100 sensor node for the simulation and three distinctive type of dishonest node i.e. 10, 20 and 30 nodes are induced to prove secure and reliability of the model. Performance evaluation is carried out in different phase, at first network energy utilization is evaluated to show the efficient data aggregation; also number of non-function nodes parameter is considered for evaluation of same. Further, to prove the security and reliability comparison with existing model is carried out . Moreover, in order to prove the model efficiency and security analysis different malicious nodes are induced and further the model efficiency is observed in terms of Energy consumption and network failed nodes. Furthermore, comparative analysis is carried out with the existing model [36] in terms of malicious packet identification and throughput.

4.1 Energy utilization over simulation time

Energy utilization is one the important parameter for evaluation of any WSN model; even if the model is secure, the model should be efficient i.e. energy utilization has to be improvised. Below figure i.e. figure 3 shows the energy utilization of IDM model through varying the different number of malicious sensor nodes; in below figure x-axis presents the simulation time and y axis presents the energy utilization. Moreover, graph shows that as the number of rounds increases, more energy is required.

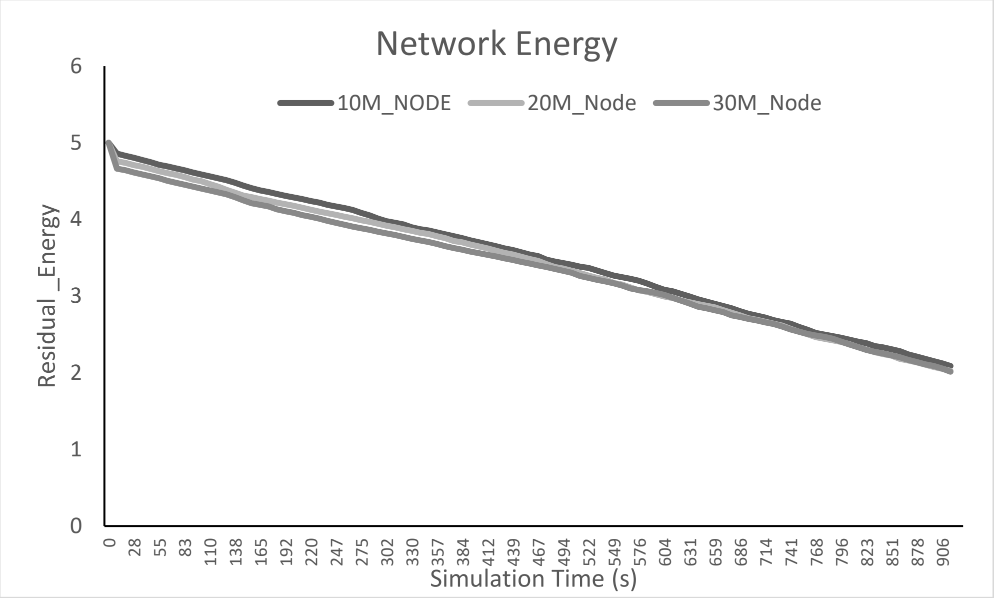

Figure 3 Energy utilization consideirng the different malicious nodes.

4.2 Non-Function over number of rounds

In WSN, number of nodes functioning plays critical role as more number of nodes functioning makes the efficient data transmission; below figure shows the dead nodes. In general nodes fail due to various reason due to architectural design or physical and environmental issue; in below figure i.e. figure 4, x-axis presents number of round and y-axis presents number of dead nodes. Moreover, through the graph it was observe that as number of rounds increases there is increase in dead number of nodes; also it was observing that there is sharp increase with increase in induction of malicious nodes.

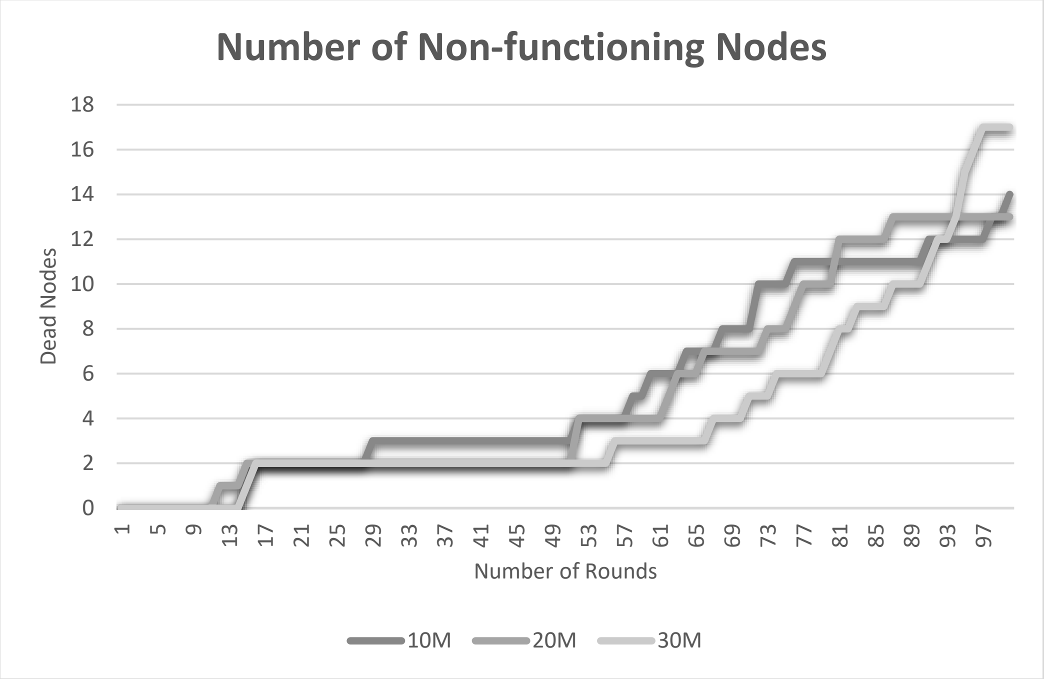

Figure 4 Non-functioning nodes

4.3 Throughput comparison with various dishonest node induction

In general throughput is defined as the rate at which work is getting done; the more throughput shows the better efficiency of model; below graph shows throughput performance comparison of existing and IDM model; Moreover, for 10 malicious nodes, existing model observes the throughput of 0.434 and IDM model observes the throughput of 0.672. Further, in case of 20 malicious node throughput observed by existing model is 0.099 whereas IDM model observes the throughput of 0.171. Similarly, in case of 30 malicious nodes, existing model achieves throughput of 0.1288 whereas IDM model achieves throughput of 0.2604

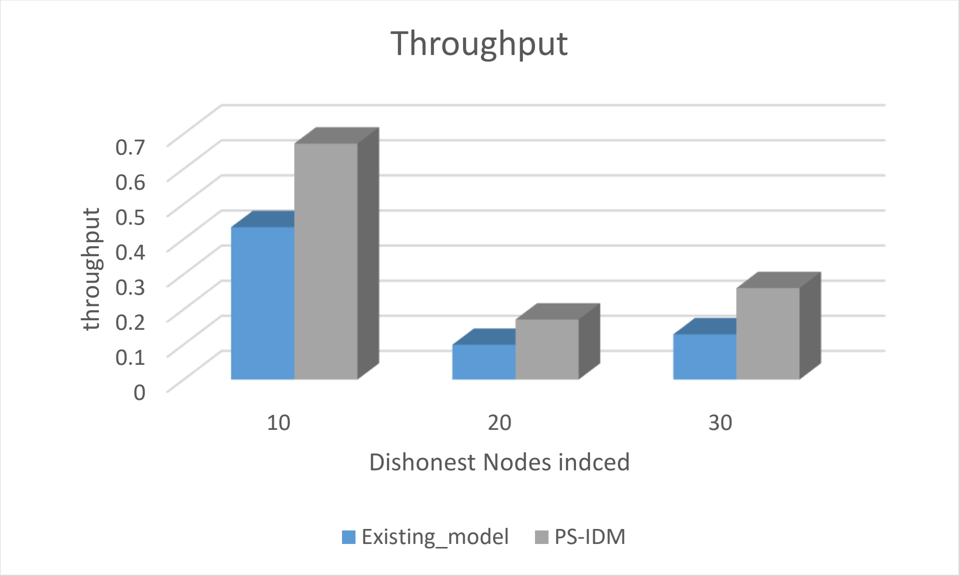

Figure 5 throughput comparison

4.4 Correct Packet classification comparison with various dishonest node induction

Identification of malicious nodes parameter express the security concern of data aggregation mechanism; the more number of packet identification shows the efficiency of model. In below figure i.e. figure 5 comparison graph is plotted between the existing and IDM model where x-axis shows the various number of nodes induced and y-axis shows packets that are malicious. Furthermore, in case of 10 malicious nodes packets identified by the existing model is 62 whereas packets identified by the IDM model is 96. In case of 20 malicious node induction, existing model identifies the 11 malicious packets whereas IDM model identifies 19 malicious packet. At last, 30 nodes are induced as the malicious node and 46 malicious packets are detected by the existing model and 93 is detected by the IDM model.

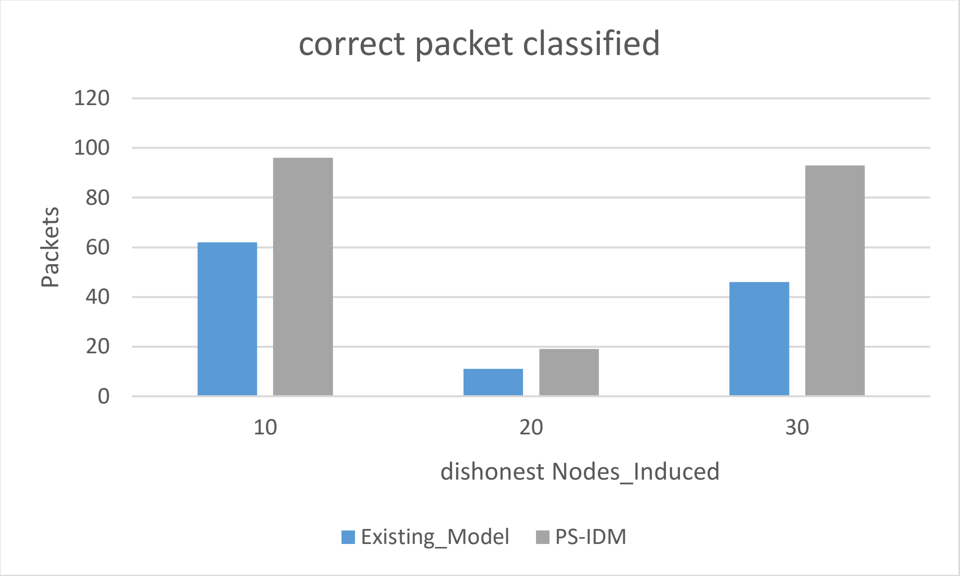

Figure 6 correctly classified malicious data packets

4.5 Packet misclassified comparison with various dishonest node induction

Packet misclassification is one of the important parameter from security concern as it evaluates the wrongly identified packets i.e. packets might be hones but detected as the malicious packet. Hence the less number of misclassified packets shows better and efficient model. Figure 7 shows the comparison of misclassified data packets. Further when 10 malicious nodes, existing model misclassifies 38 packets whereas IDM model misclassifies only 4. Similarly, for 20 malicious nodes, existing model misclassifies 10 packets whereas IDM mechanism misclassifies 2 packets. At last, in case of 30 malicious nodes, existing model observes 54 misclassified packets whereas IDM model observes only 7.

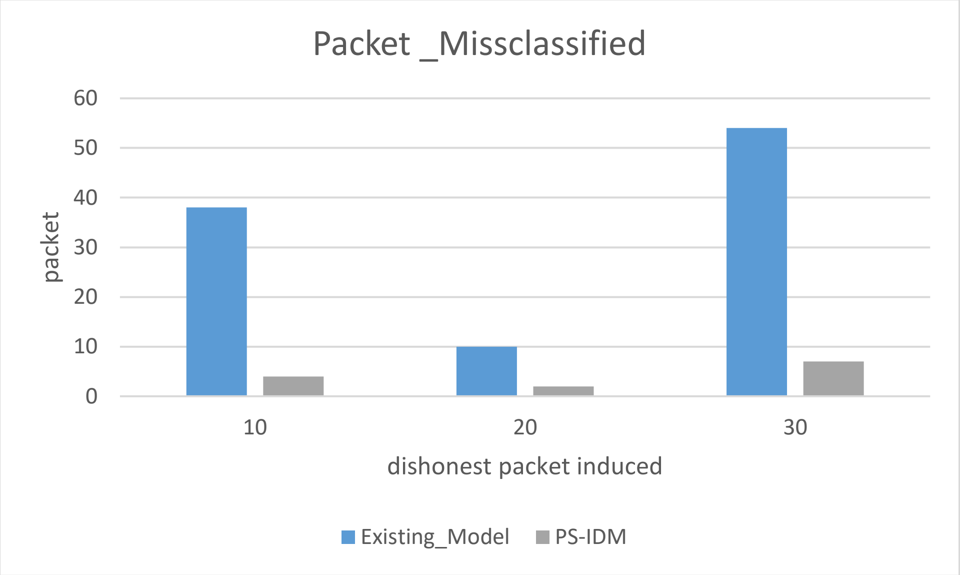

Figure 7 Missclassified data packets comparison

4.6 Comparative analysis

In this sub-section of the research improvisation of IDM model over the existing model is computed considering the above evaluation parameter; at first throughput comparison is carried out where IDM model simply outperforms the existing model. Further, malicious packet identification parameter is considered for evaluation and observe that IDM model is improvised by 54.83% for 10 malicious nodes, 72.72 % for 20 malicious nodes and more than double of improvisation for 30 malicious nodes. Similarly, packet misclassified is another parameter considered for evaluation where 89.47%, 80% and 87% of improvisation is observed for 10, 20 and 30 malicious nodes respectively.

Conclusion

In WSN based IoT application, Data aggregation is one of the fundamental operation; Data aggregation not only avoids the data redundancy but also enhances the network lifetime and subsequently energy. Data aggregation has been proven to boon for the various application, however security is considered as the major concern due to high vulnerability of the network. This research work designs and develop IDM (Integrated Distributed Mechanism); IDM comprises three integrated modules; first module is designed for efficient and secure data aggregation module which focuses not only secure but consider energy utilization as well. Second module Computed the misclassified data packets computation and third module identifies the dishonest data packets, nodes and discards through designed constraint. IDM is evaluated considering the energy utilization, non-functioning nodes which proves the model efficiency; further correct and misclassified packet are computed and compared to prove the model reliability in comparison with existing model. IDM model observes improvisation of 54.83% for 10 malicious nodes, 72.72 % of 20 malicious nodes and more than double of improvisation for 30 malicious nodes. packet misclassified is another parameter considered for evaluation where 89.47%, 80% and 87% of improvisation is observed for 10, 20 and 30 malicious nodes respectively.

Integrated Distributed mechanism is provable, secure and reliable mechanism and outperforms the existing model in comparative analysis; however, there are still many challenge in terms of security and efficiency which can be looked into future work.

Reference

1. K. Shafique, B. A. Khawaja, F. Sabir, S. Qazi and M. Mustaqim, "Internet of Things (IoT) for Next-Generation Smart Systems: A Review of Current Challenges, Future Trends and Prospects for Emerging 5G-IoT Scenarios," in IEEE Access, vol. 8, pp. 23022-23040, 2020, doi: 10.1109/ACCESS.2020.2970118.
2. A. Shahraki, A. Taherkordi, Ø. Haugen and F. Eliassen, "A Survey and Future Directions on Clustering: From WSNs to IoT and Modern Networking Paradigms," in IEEE Transactions on Network and Service Management, doi: 10.1109/TNSM.2020.3035315.
3. M. T. Lazarescu, "Design of a WSN Platform for Long-Term Environmental Monitoring for IoT Applications," in IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 3, no. 1, pp. 45-54, March 2013, doi: 10.1109/JETCAS.2013.2243032.
4. V. Gupta and S. De, "Energy-efficient Edge Computing Framework for Decentralized Sensing in WSN-assisted IoT," in IEEE Transactions on Wireless Communications, doi: 10.1109/TWC.2021.3062568.
5. L. M. Dang, M. Piran, D. Han, K. Min, and H. Moon, "A survey on internet of things and cloud computing for healthcare," Electronics, vol. 8, no. 7, p. 768, 2019.View at: Publisher Site | Google Scholar.
6. H. Zhang, Y. Hu, R. Wang, Z. Li, P. Zhang and R. Xu, "Energy Efficient Frame Aggregation Scheme in IoT over Fiber-Wireless Networks," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2021.3051098.
7. K. T. Kadhim, A. M. Alsahlany, S. M. Wadi, and H. T. Kadhum, "An overview of patient's health status monitoring system based on internet of things (IoT)," Wireless Personal Communications, vol. 114, no. 3, pp. 2235–2262, 2020.
8. A. Mohammadali and M. S. Haghighi, "A Privacy-Preserving Homomorphic Scheme with Multiple Dimensions and Fault Tolerance for Metering Data Aggregation in Smart Grid," in IEEE Transactions on Smart Grid, doi: 10.1109/TSG.2021.3049222.
9. A. Fu, X. Zhang, N. Xiong, Y. Gao, H. Wang and J. Zhang, "VFL: A Verifiable Federated Learning with Privacy-Preserving for Big Data in Industrial IoT," in IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2020.3036166.
10. A. Abbas, and M. Younis, "Establishing connectivity among disjoint terminals using a mix of stationary and mobile relays", Computer Communications, vol. 36, no. 13, pp. 1411-1421, 2013.
11. Y. K. Joshi, and M. Younis, "Restoring connectivity in a resource constrained WSN", Journal of Network and Computer Applications, vol. 66, pp. 151-165, 2016.

12. W. Lalouani, M. Younis, and N. Badache, “Interconnecting isolated network segments through intermittent links”, Journal of Network and Computer Applications, vol. 108, pp. 53-63, 2018.
13. S. Lee, M. Younis, B. Anglin, and M. Lee, “LEEF: Latency and energy efficient federation of disjoint wireless data collection position segments”, Ad Hoc Networks, vol. 71, pp. 88-103, 2018.
14. J. L. V. M. Stanislaus, and M. Younis, “Mobile relays based federation of multiple wireless sensor network segments with reduced-latency”, In: Proceedings of IEEE International Conference on Communications, pp. 6407-6411, 2013.
15. J. L. V. M. Stanislaus, and M. Younis, “Delay-Conscious Federation of Multiple Wireless Sensor Network Segments Using Mobile Relays”, In: Proceedings of IEEE Vehicular Technology Conference, pp. 1-5, 2012.
16. [F. Senel, and M. Younis, “Optimized interconnection of disjoint wireless sensor network segments using K mobile data collectors”, In: Proceedings of IEEE International Conference on Communications, pp. 492-496, 2012.
17. L. Goratti, T. Baykas, T. Rasheed, and S. Kato, “NACRP: A connectivity protocol for star topology wireless data collection position networks”, IEEE Wireless Communications Letters, vol. 5, no. 2, pp. 120-123, 2016.
18. Z. Xu, L. Chen, C. Chen, and X. Guan, “Joint clustering and routing design for reliable and efficient data collection in large-scale wireless data collection position networks”, IEEE Internet of Things Journal, vol. 3, no. 4, pp. 520-532, 2016.
19. I. F. Senturk, K. Akkaya, and S. Jananse fat, “Towards realistic connectivity restoration in partitioned mobile data collection position networks”, International Journal of Communication Systems, vol. 29, no. 2, pp. 230- 250, 2016.
20. X. Wang, L. Xu, S. Zhou, and W. Wu, “Hybrid Recovery Strategy Based on Random Terrain in Wireless Sensor Networks”, Scientific Programming, vol.2017, Article ID 5807289, 2017.
21. Z. Mi, Y. Yang, and J. Y. Yang, “Restoring connectivity of mobile robotic sensor networks while avoiding obstacles”, IEEE Sensors Journal, vol. 15, no. 8, pp. 4640-4650, 2015.
22. J. Yu, K. Wang, D. Zeng, C. Zhu, and S. Guo, “Privacy-preserving data aggregation computing in cyber-physical social systems”, ACM Transactions on Cyber Physical Systems, vol.3, no.1, Article 8, 2018.
23. J. Wang, H. Zhang, Z. Ruan, T. Wang, and X. D Wang, “A Machine Learning Based Connectivity Restoration Strategy for Industrial IoTs”, IEEE Access, vol. 8, pp. 71136-71145, 2020.
24. K. Toyoshima, T. Oda, M. Hirota, K. Katayama, L.Barolli, “A DQN Based Mobile Actor Node Control in WSAN: Simulation Results of Different Distributions of Events Considering Three-Dimensional Environment”, In: International Conference on Emerging Internetworking, Data and Web Technologies. Springer, Cham, pp. 197-209, 2020.
25. C. Xu, K. Wang, P. Li, R. Xia, S. Guo, and M. Guo, “Renewable energy aware big data analytics in geo-distributed data centers with reinforcement learning”, IEEE Transactions on Network Science and Engineering, vol.7, no.1, pp.205-215, 2020.

26. C. H. Liu, Q. Lin, and S. Wen, “Blockchain-enabled data collection and sharing for industrial IoT with deep reinforcement learning”, IEEE Transactions on Industrial Informatics, vol. 15, no. 6, pp. 3516-3526, 2018.
27. M. Du, K. Wang, Y. Liu, K. Qian, Y. Sun, W. Xu, and S. Guo, “Spacechain: a three-dimensional blockchain architecture for IoT security”, IEEE Wireless Communications, vol.27, no.3, pp.38-45, 2020.
28. M. Yang, T. Zhu, K. Liang, W. Zhou, and R. H. Deng, “A blockchain based location privacy preserving crowdsensing system”, Future Generation Computer Systems, vol. 94, pp. 408-418, 2019.
29. M. Li, J. Weng, A. Yang, W. Lu, Y. Zhang, L. Hou, J. Liu, Y. Xiang and R. H. Deng, “Crowd BC: A blockchain-based decentralized framework for crowdsourcing”, IEEE Transactions on Parallel and Distributed Systems, vol. 30, no. 6, pp. 1251-1266, 2018.
30. S. Chen, Z. You and X. Ruan, "Privacy and Energy Co-Aware Data Aggregation Computation Offloading for Fog-Assisted IoT Networks," in IEEE Access, vol. 8, pp. 72424-72434, 2020, doi: 10.1109/ACCESS.2020.2987749.
31. Y. Chen, J. -F. Martínez, L. López, H. Yu and Z. Yang, "A dynamic membership group-based multiple-data aggregation scheme for smart grid," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2021.3063412.
32. X. Yan et al., "Verifiable, Reliable, and Privacy-preserving Data Aggregation in Fog-assisted Mobile Crowdsensing," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2021.3068490.
33. M. Zhang, H. Zhang, D. Yuan and M. Zhang, "Learning-based Sparse Data Reconstruction for Compressed Data Aggregation in IoT Networks," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2021.3059735.
34. B. Yin and X. Wei, "Communication-Efficient Data Aggregation Tree Construction for Complex Queries in IoT Applications," in IEEE Internet of Things Journal, vol. 6, no. 2, pp. 3352-3363, April 2019, doi: 10.1109/JIOT.2018.2882820.
35. J. N. Al-Karaki and G. A. Al-Mashaqbeh, "SENSORIA: A New Simulation Platform for Wireless Sensor Networks," 2007 International Conference on Sensor Technologies and Applications (SENSORCOMM 2007), 2007, pp. 424-429, doi: 10.1109/SENSORCOMM.2007.4394958.
36. A. B. Alexandru and G. J. Pappas, "Private Weighted Sum Aggregation," in IEEE Transactions on Control of Network Systems, doi: 10.1109/TCNS.2021.3094788.