International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 23 (2017) pp. 13663-13676

Optimized Mechanism for Minimizing Data Overhead in IoT Environments: Design, Simulation and Performance Evaluation

Omar Said
College of Computers and Information Technology, Taif University, Taif, KSA.
College of Science, Menoufia University, Shibin El Kom, Menoufia, Egypt.

ORCID: 0000-0002-8868-8000

Alaa Elnashar
College of Computers and Information Technology, Taif University, Taif, KSA.
College of Science, Minia University, Minia, Egypt.

Osama Elshakankiry
College of Computers and Information Technology, Taif University, Taif, KSA.
Faculty of Electronic Engineering, Menoufia University, Menouf, Menoufia, Egypt.

Abstract

The Internet of Things (IoT) communicates billions of active and/or passive physical objects and creates a mesh of networks producing a huge information bulk. This big size of data represents a challenge for the IoT environments due to the problems arise during information storage, transmission, management and transactions. So, minimization of information size without affecting the IoT system efficiency may overcomes such problems. In this paper, an optimized data collection mechanism for IoT environments is presented. The proposed mechanism provides some recommendations and determines restrictions during collection, transmission, and processing of data to minimize its size. The mechanism comprises three strategies, dynamic prioritization for collected data, coding optimization, and energy saving. These strategies are employed to achieve the desired goal. Finally, an IoT environment to simulate, test and evaluate the proposed mechanism is constructed using network simulator NS2. Accordingly, the simulation results are analyzed to ensure that the proposed mechanism is robust and efficient. The results prove that the proposed mechanism outperforms the traditional IoT in data reduction rate, energy consumption, end-to-end delay, and throughput.

Keywords: Internet of Things; IoT Simulation; Data Fitting, Data Prioritization; Data Minimization.

INTRODUCTION

The Internet of Things (IoT) is a new model that is rapidly grows in the recent wireless telecommunications topic. The basic concept of IoT is to communicate billions of uniquely addressable everyday life objects. These objects may be active or passive. The active objects should have processing units and memories such as PCs, mobiles, and sensors. The passive objects don't have any processing units such as doors, clothes, and things in shopping malls. In addition, these objects should be communicated together to reach the IoT applications goals. Since these objects are uniquely addressed by using internet strategies, the information about them can be transmitted using the same ideas or (adapted versions) of Internet protocols. Furthermore, the management of IoT system can be achieved without human intervention such as IBM's Smarter Planet for autonomic computing [1-7].

The Internet of Things growth requires technical innovation in many important fields. An effective and simple item identification system should be developed in order to communicate physical objects to large databases and network (Internet). In addition, innovation systems to determine how the data can be collected and processed should be presented (only Radio-Frequency Identification (RFID) is available). Also, to detect the changes in things' status, new technologies should be created (only achieved through sensors). Furthermore, information processing at the network edges can be achieved by adding embedded system within such things themselves. Moreover, the new trends such as nano technology increases the opportunities of communicated extra things in IoT environments leading to excessive collected data size.

The IoT enables advanced applications in many fields. The most common field, which applied the IoT technology, is marketing [8]. Also, the traffic field uses IoT applications to provide many advantages such as adaptive time for each traffic sign to be on or off and how traffic flow can be distributed normally within different roads [9]. Furthermore, many countries such as France applied the IoT technology on glass containers that prepared with ultrasonic sensors to send a descriptive data about itself periodically [9]. In addition, there are many fields, which are still under research and use multimedia applications, such as e-learning, healthcare, and smart cities [10-12].

As stated above, there is a mixture of heterogeneous communication technologies which are required to accommodate the variety of IoT. Accordingly, there are many challenges arise such as scalability, privacy, and addressing.

Scalability is the capability of IoT system to handle the increased amount of data. Privacy requires techniques and technologies to protect sensitive and private data, communications, and preferences. Addressing challenge lies in finding a protocol that should comprise billions of objects to cover not only the current IoT but also its future expansion [13-15].

Recently, the deployment of automatic communication of objects in our lives represents a risk. For example, inserted nano RFID tools in the personal things such as devices and clothes can be triggered to reply with their ID and other data. This enables a powerful mechanism that would cover most of our lives. Also, this mechanism must include functions that are designed to manage trust, privacy and security of all the sending or receiving data. These functions may be either built in one specific layer or distributed through multiple layers, from the object abstraction to the service composition, in a manner that does not affect system performance or introduce excessive overheads [16-19].

This paper is organized as follows: In Section 2 the paper problem is defined. In Section 3 the related works are discussed. In Section 4 the paper contribution is showed. In Section 5 the proposed mechanism is introduced. In Section 6 simulation and evaluation of the proposed mechanism are demonstrated. Finally, the conclusion is introduced in Section 7.

PROBLEM FORMULATION

From the above discussion, it is clear that IoT contains billions of nodes. Each node produces data in an automated manner. In addition, variety of nodes implies variety of collected data which requires many complex protocols, algorithms, and techniques to manipulate such collected data. Furthermore, the continuous increase in the number of devices that are connected to the Internet makes the size of the collected data also increases dramatically. The rapid increasing of data size badly affects scalability, transmission, and processing. So, the main challenge for this paper is to minimize the collected data size without affecting the efficiency of IoT applications.

RELATED WORK

Most of researches presented in prioritization, coding and energy saving concerns only with Wireless Sensor Network (WSN). These researches can be abstracted as follows:

D. Virmani, et al, proposed Priority-Energy Based Data Forwarding (PEDF) algorithm that tries to select the best QoS for high priority packets. It saves energy, minimizes delay and maximizes throughput. The main weak point of this algorithm is that it is oriented only with WSN. It is not suitable for IoT environments. In addition, it is not simulated or implemented. Also, the algorithm does not reflect the manner in which the claims described in [20] are achieved. M. Monowar, et al., proposed a Prioritized Heterogeneous Traffic-oriented Congestion Control Protocol (PHTCCP) to adjustment congestion and flow rate of packets for prioritized diverse traffics dynamically. This protocol is executed from hop-by-hop. Intra-queue and inter-queue priorities are used in the proposed protocol along with weighted fair queuing to ensure the feasibility of data transmission rates. The protocol results are not significant since simulation parameters values are not sufficient. It is applied only on WSNs and not suitable for IoT [21]. F. Ullah, et. al presented the challenges for Wireless Body Area Network (WBAN) in addition to a cross-layer design structure and a model for routing, MAC, and PHY in WBAN. The model can select the best routing path for immigrant information based on many parameters such as low temperature and minimum hop-counts. This proposed model is designed only for WBAN. The proposed routing model of WBAN is not tested [22]. Furthermore, strengths and weaknesses of existing cross-layer protocols in the WBAN domain have also been introduced. C. Cantillas, et. al, proposed a technique to prioritize the WSN data. This technique uses three levels for prioritization, high, medium, and low based on the queuing theory with static levels of priority. It provides fast data transmission. The simulation environment parameters are not sufficient to indicate accurate results [23]. R. Karthikeyan, et al., proposed the same idea of C. Cantillas, et. al, but with different simulation environment [24]. Several studies [25-27] focus on dynamic prioritization for WSN. M. kumar, et al., [25] presented a data priority mechanism in heterogeneous WSNs in which Quality of Service (QoS) requirements of different wireless nodes are considered. The mechanism distributes the bandwidth among the transmitted data according to their QoS demands. The prioritization of data in this mechanism is based on the QoS. In addition, the prioritization of data is not achieved during collection phase. Also, the optimal bandwidth distribution algorithm may be executed in a specific network type, but with diversity of networks such as IoT environment may be failed. Furthermore, the user opinion (desire) is neglected in the classification procedure for this mechanism. Moreover, this mechanism is tested for only WSN but the IoT comprises more and more collected data. The simulation of this mechanism is not sufficient due to lake in its parameter values. M. Kumar, et al., proposed a technique called Dynamic Data Prioritization (DDP) to provide priority based planning to enhance the delivery of packets through wireless links. This technique guarantees QoS and improved transmission rate in wireless device networks [26]. C. Liang, et al, presented a framework called RushNet. This framework prioritizes two traffics in multi-hop sensor networks such as low-priority traffic that has large size without delay sensitivity and high-priority traffic for low data size with latency sensitivity. The latency reduction, which is achieved by RushNet framework, is minor relative to the complexity that is added by the same framework [28].

Network coding can be defined as the mathematical operations which are used to mix packets by relay nodes to reduce the number of transmitted packets within the network. The broadcast nature of wireless networks and link diversity enrich research topic in network coding. Unfortunately, most of existing network coding techniques are designed only for WSN. The network coding techniques are classified into two classes, inter-session and intra-session. The inter-session techniques mix the packets that are received from different sessions in case of bottleneck status [29-38]. The intra-session techniques mix the packets from the same sessions and address the packet loss problem [39-51]. The inter-session and intra-session techniques should be adapted to accommodate with IoT environments. The IoT environment sessions are dynamic, scalable, heterogeneous, and their natures change suddenly. So, the wireless network techniques should consider these properties in their design.

PAPER CONTRIBUTION

The IoT environment transmits a massive size of data. This massive data may require high storage capacity, powerful processing unites, and high bandwidth to be able to transmit these types of data (i.e. QoS). The IoT environment comprises heterogeneous objects which have different processing capabilities, storage capacities, and transmission channels. This heterogeneity represents the main challenge against handling of IoT massive data. To solve this challenge, the IoT massive data should be minimized. This paper uses data fitting, data compression, and data prioritization to minimize the data which should be transmitted and handled through the IoT.

PROPOSED MECHANISM

The proposed mechanism comprises three strategies: data fitting, data compression, and data prioritization. The minimization processes contain two concepts. The former is to minimize the size of data which will be transmitted through the IoT system without packets' neglecting (data fitting and compression). The latter is to decrease the size of data by neglecting part of it (prioritization). In the following subsections, applying of these three strategies on the IoT data will be discussed, see Fig. 1.

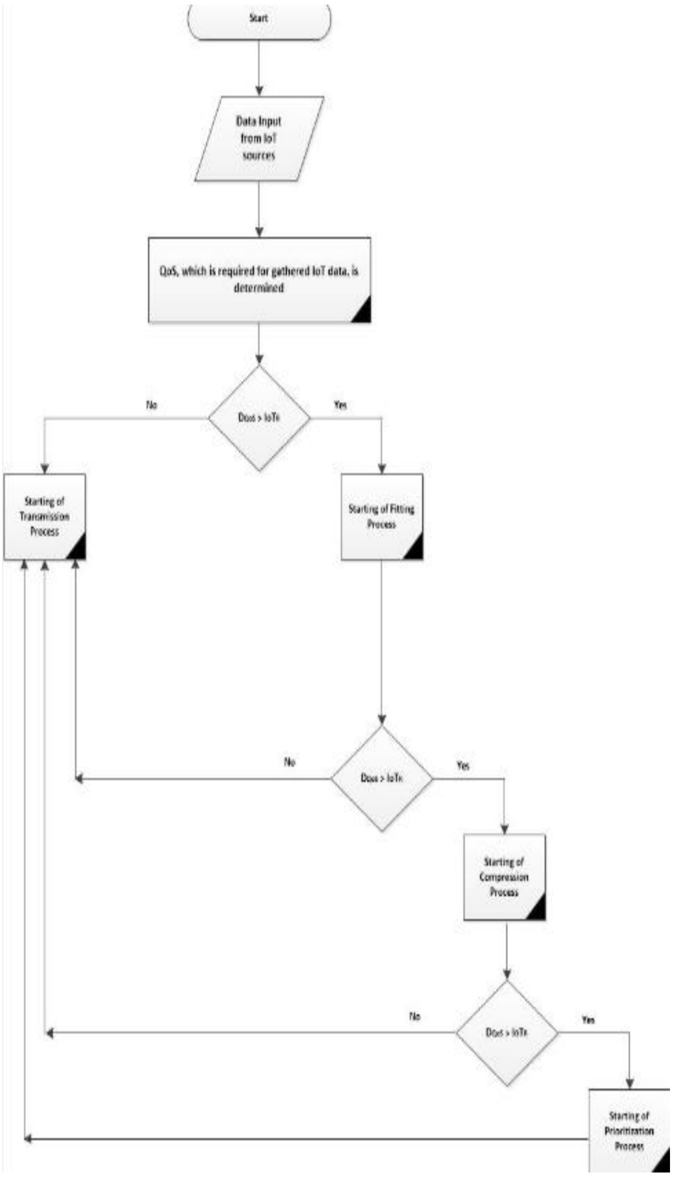

Figure 1: Flowchart of the proposed strategies

Data fitting

The first method in our Data Reduction Model (DRM) is data fitting. The IoT environment contains different networks such as WSN, RFID, Mobile and others. WSN is considered as the core of IoT [18]. So, it can be taken as an example to clarify how the data fitting mechanism can be applied on the IoT system. It is well known that data is gathered from many sensors in the WSNs. The sensors are organized in pairs. The data that is gathered by each sensor is also organized in a pair (x, y). Then, the data fitting mechanism can be applied on the output data pairs. The data fitting mechanism has many methods. In this paper, the least squares method is selected to decrease the data size. There are two types of the least squares method, linear and nonlinear [52, 53]. In the linear method, the data gathered by sensors should have a regular interval and this is considered as a special case of IoT. As in most of the cases the sensors gather different and random data, the nonlinear least square method should be applied in the IoT system. Dealing with random data fitting may consume processing time and increase the complexity of the proposed mechanism based on the size of data that is transmitted through IoT systems. So, for simplicity, the linear least squares method will be applied.

To apply the linear least squares, data should be given as follow: $\{(S_{p1_i}, S_{p2_i}), (S_{p1_i}, S_{p2_i}), \ldots \ldots \ldots, (S_{p1_n}, S_{p2_n})\}$, where i = 1 to n and n is the number of values which is gathered by sensors (sensors pairs). Equation 1 determines the relation between 'a' and n-vector b.

$$a \approx f(b) \tag{1}$$

Where 'b' is an independent variable, 'a' is a response variable, and 'f': $R^n \rightarrow R$ determines the relationship between 'a' and 'b'. Every so often, 'b' is a feature vector, and 'a' is predictable data. So, 'f_{app}' is required. 'f_{app}' is the approximation of the function 'f'. 'f_{app}' is based on the observed data which is gathered by sensors in the WSN. Equation 2 determines the relation between 'f_{app}' and 'f'.

$$f_{app} = \theta_1 f_1(b) + \theta_2 f_2(b) + \theta_3 f_3(b) + \cdots \theta_m f_m(b) \tag{2}$$

Where m is the number of sensors in the IoT system, $f_1(b), f_2(b), \ldots$ are functions for $(S_{p1_i}, S_{p2_i}, \ldots)$, respectively, and $\theta_1, \theta_2, \theta_3, \theta_m$ are the parameters which will be chosen based on the relation between the gathered data and the output data. Equation 3 determines the approximation 'a_{app}'.

$$a_{app} = f_{app}(b_i) \tag{3}$$

The main target is determined in equation 4.

$$a_{app_i} \approx a_i \tag{4}$$

The prediction error is determined by equation 5. Hence, θ should be chosen to minimize the prediction error.

$$r_i = a_{app_i} - a_i \tag{5}$$

Compression

The second method in the proposed DRM is compression. Compression of data is used to minimize the size of files which will be transmitted from one thing to another in the IoT system. There are many types of compression techniques. Selection process of a special compression type should be based on the implementation of all of these types. Till now, there are no results for applying different compression types on the IoT environment. Hence, one compression type is selected as a trial and the results are shown in section 6. Distributed Source Coding (DSC) [54] exploits the correlation which may be appeared in the IoT system without the need for communication between things.

As the IoT system comprises passive and active things, hence applying the DSC by all of the IoT objects is extremely difficult (if not impossible). So, DSC will be implemented by a limited number of objects in addition to the sink points which will gather information from the sensors, mobiles, etc. As mentioned before, WSN is considered as a core of IoT system. Hence, the information which will be transmitted by sensors represents most of the IoT system load. Compression of correlated sources will be achieved using DSC. The compression using DSC occurs in distributed manner. For example, suppose that there are many sensors in the IoT. These sensors are distributed over a specific area. The mission is to compress the data for each sensor, collect the compressed data, and achieve the decompression at sink point, see Fig. 2.

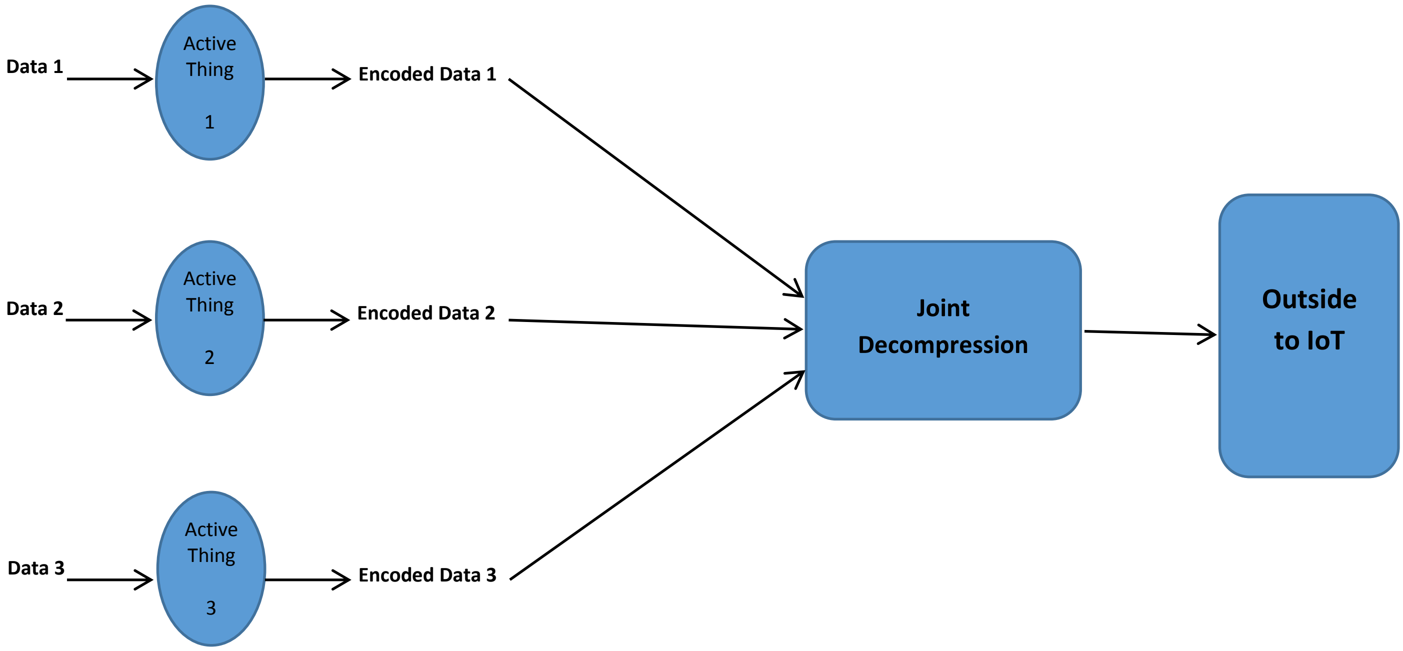

Figure 2: General view of the compression process

The things in the IoT system can be organized in groups (clusters). So, DSC can be applied at each cluster. Each cluster has many nodes and one (or more) header node. The header node is used to gather the information that is transmitted by other nodes in the cluster. This information may be gathered by sensors, transmitted by RFID, or exchanged using mobile networks. The header nodes in the IoT system will implement the DSC technique to compress information, see Fig. 3. So, selection process of header node in each cluster, which found at [54], should consider the implementation of DSC technique. There are two categories of DSC, lossy and lossless.

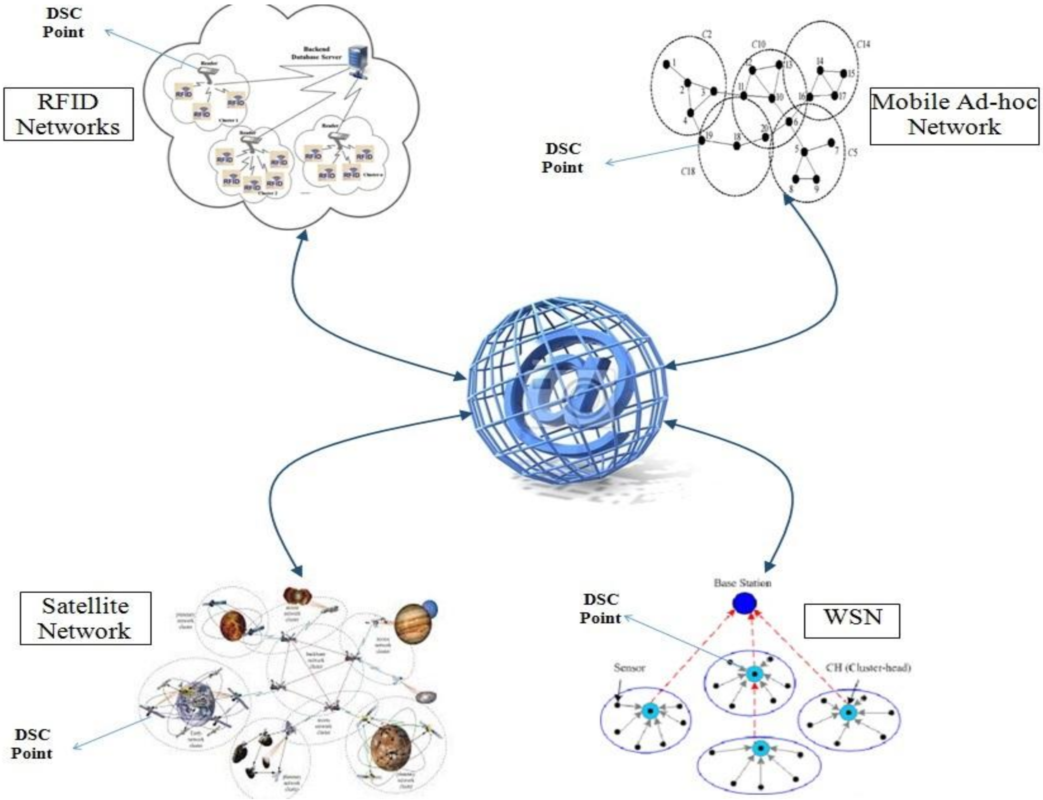

Figure 3: DSC Point for each network in the IoT environment

In the proposed DRM, lossy DSC is implemented as it deals with different source rates such as in the IoT system. In addition, lossless DSC will produce additional overhead. Therefore, in the proposed DRM, DSC with Raptor code is used to compress the IoT transmitted packets. Raptor is considered as a new version of fountain codes. Raptor code is based on Luby Transform (LT). Infinite number of encoded symbols can be generated using LT code with parameters (k, C, Ω(x)) under degree distribution (Ω(x)), where k is the number of bits of code-word C. Tanner graph [8] is used to create an encoded symbol (XORed symbols) using random selected degree d that is extracted by Ω(x). In this case, the source nodes are called variable nodes and the encoded nodes are called check nodes. Hence, the communication between variable nodes and check nodes should be achieved to extract the encoded symbols. At least cklog(k) connections should be found to ensure that the decoding process will be accomplished successfully. Raptor code is a mixture between LT and Low-Density Parity Check (LDPC). It ensures that the complexity of encoding and decoding processes linearly varies with the variable k. hence, LT will be used to compress a constant ratio symbols while LDPC is used to compress other symbols. Raptor code rateless specs are assured by continuously sampling using the degree distribution, Ω(x). One from DSC using raptor code is matching of any desire rate even in case of runtime dynamic change. In addition, it has a low complexity which makes it suitable to be applied in the IoT environments due to the limited resource of IoT objects. The decoding process should guarantee a correction of errors which may be resulted due to encoding process. These errors may be correlated. Encoding and decoding process are found at [54].

Prioritization

After the data is abstracted using linear least squares technique and compressed using DSC (Raptor code), the QoS of IoT system should be determined. In case of a healthy IoT system, the data can be transmitted without problems. But, in case of an unhealthy IoT system, the data should be more reduced. To achieve this mission, the proposed DRM should comprise technique other than data abstraction and compression. The proposed model should prioritize the IoT data before its processing. In this model, queuing system is used to organize the data into multi-levels such that the most important data will be transmitted first. The proposed DRM uses are four prioritization levels to categorize the data into four classes; class1, class2, class 3, and class 4. The data prioritization may be implemented at the object or at the sink node. For example, each sensor may send a part from its observed data after its system is adapted to accomplish this function. On the other hand, if the system between each sensor and its sink node has a good QoS, the sensor can send its data without prioritization and the sink node may apply the prioritization idea, see Fig. 4. Each class will be represented by a queue. The IoT data are classified depends on its importance, which is determined by the user. Each block of categorized data is pushed in one queue. For integration purpose, the IoT active things should be equipped with memory system support the queuing theory.

The queuing model uses linear data model which is defined in equation 6.

$$P_{lm} = \lambda_l + \mu_m \tag{6}$$

Where λ is the birth rate, μ is the death rate, and P_{lm} represents the transition probability from state *l* to state *m.* As each class is represented by one queue, there are four queues in the prioritization model, Q1, Q2, Q3, and Q4. At any time after the QoS for targeted part of IoT system is tested, any packet can be en-queued into one of these four queues using data classifier as shown in Fig. 5. There are eight steps to accomplish this target.

Step 1 The incoming packets are classified by classifier according to specific criteria that is defined by server and may be changed within interval time. The classifier selection has four initial probabilities, P_1, P_2, P_3, and P_4 where $\sum_{i=1}^{4} p_i = 1$.

Step 2 If the transmission process of a specific packet cannot be completed due to low QoS such as bandwidth, congestion, or memory, the next packet in the same queue should be serviced by providing the QoS.

Step 3 The suspended packet should be transformed to the next queue which has a lower priority.

Step 4 Within time interval, the Q3 is empty and the processing of packets is running in Q1 provided that the required QoS is found.

Step 5 The handling of packets in Q1 proceeds until it became empty. Then, the handling process transfers to service the packets in Q2.

Step 6 If one packet in the Q2 has incomplete service it should be transmitted to Q3.

Step 7 If one packet in the Q3 has incomplete service it should be transmitted to Q4.

Step 8 In case of new packet come to Q1 and the service process in Q4, the controller is immediately jumped to Q1 to start servicing of packet in Q1. After that, the controller returns the service to Q4.

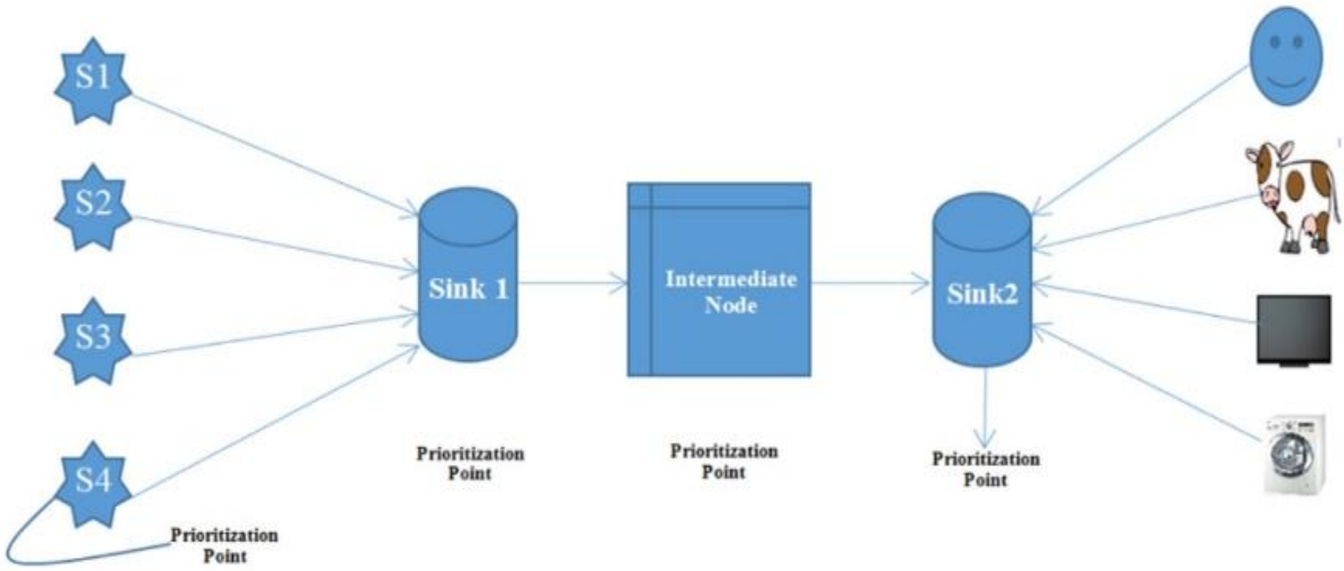

Figure 4: Apply the prioritization process

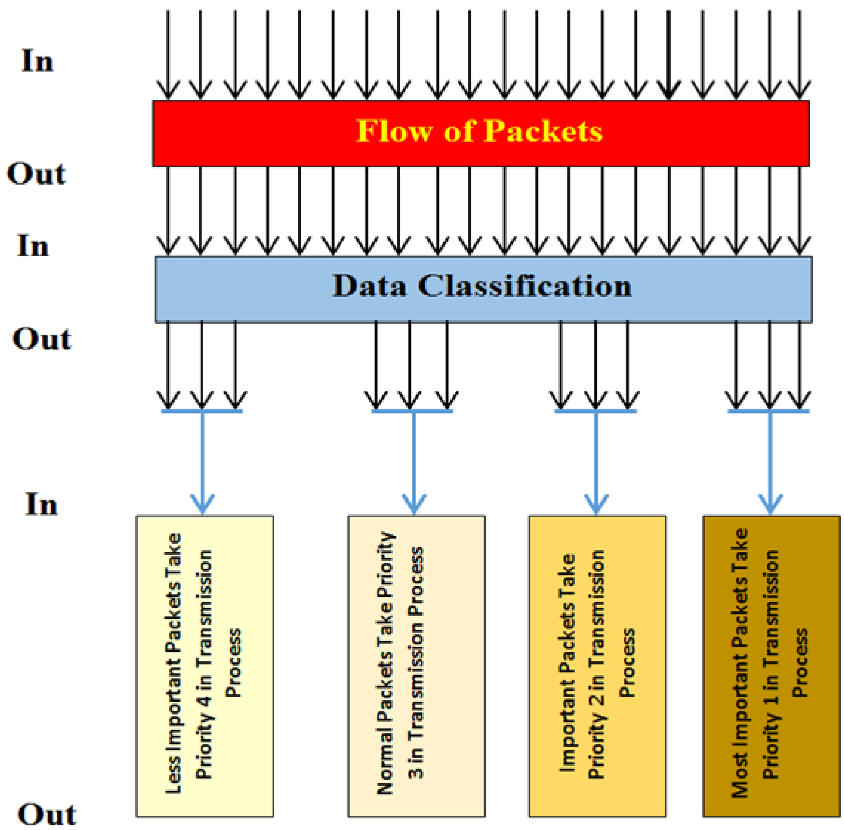

Figure 5: En-queued process after data classification.

Transitions of state

The proposed queuing system can be more described by clarification of Markov chain model [55]. The queuing system contains four queues; Q1, Q2, Q3, and Q4 which could be defined as state 1, state 2, state 3, and state 4 respectively. The time used by the proposed queuing system to service one packet equals to τ (where τ = 1, 2, 3, 4 …). There are two transitions states, service or waiting. Fig. 6 shows transitions among different states.

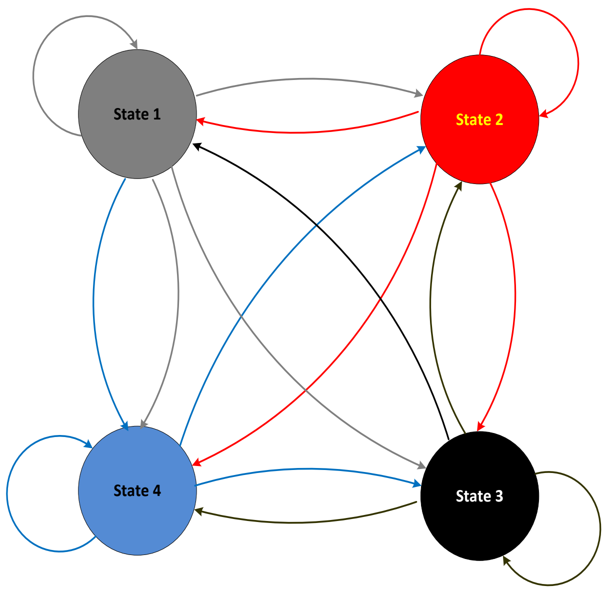

Figure 6: Four states transition diagram

Markov chain is defined as { $Y^{(\tau)} + \tau \geq 1$ } where the selected queue can be determined by $Y^{(\tau)}$ at τ^{th} time interval. The state space of Y is {Q_1, Q_2, Q_3, Q_4}. The probabilities of queues initial selection are determined by equations 7 to 10 where $\sum_{i=1}^{4} Pr_i = 1$

$$Pr_1 = P\,[Y^{(0)} = Q_1] \quad (7)$$

$$Pr_2 = P\,[Y^{(0)} = Q_2] \quad (8)$$

$$Pr_3 = P\,[Y^{(0)} = Q_3] \quad (9)$$

$$Pr_4 = P\,[Y^{(0)} = Q_4] \quad (10)$$

The probabilities of transaction for the proposed queuing system over four states are determined by $S_{ij}(i, j = 1, 2, 3, 4)$. Furthermore, $Y^{(\tau)}$, which represents the state transition matrix, can be expressed using table 1.

Table 1: State transition matrix

	Q_1	Q_2	Q_3	Q_4
Q_1	S_{11}	S_{12}	S_{13}	S_{14}
Q_2	S_{21}	S_{22}	S_{23}	S_{24}
Q_3	S_{31}	S_{32}	S_{33}	S_{34}
Q_4	S_{41}	S_{42}	S_{43}	S_{44}

The row dependent model $P_{ij} = \lambda_i + i(\mu_j)$ is applied to handle all of transitions.

SIMULATION AND EVALUATION

Simulation environment

To test the proposed DRM, a simulation environment is constructed. The network simulation package, NS2, is used to build the simulation environment [56]. The simulation of IoT comprises different networks such as WSN, RFID, and Mobile Ad-Hoc networks. These networks are communicated using the Internet. These networks are selected because they have heterogeneous nodes which represent one of the main spec of IoT environment. Furthermore, passive things are represented in RFID networks. Moreover, huge number of nodes should be found in the proposed simulation environment to simulate the rapid extensibility of IoT systems. In addition, using the Internet as a communication tool between the simulated networks shows the core concept of IoT. The simulation program moves from the simple state to the complex state passing through the middle state. Simply, the parameters, which determine the IoT state, are network size, protocols, things, overlapped areas, number of users, etc. the autonomies spec of IoT is simulated using stored file. This file comprises a group of commands which is run automatically depending on predetermined events without human intervention. There is a mapping between the events and commands.

For the simulation of WSN, it comprises number of sensors which increases over a time. These sensors are randomly distributed in an area that has a dynamic shape change. There are more than one sink nodes which are located at the area center. Horizontal and vertical coordinates are used to determine the locations of network sensors. The two main factors, which represent the most important factors in the simulation of WSN, are energy and distance between sensors. As regards the RFID network simulation, it contains many components such as tags, applications, sites, and readers. These components determine the scalability of RFID network. The nodes of RFID can be found at different geographical areas. In addition, it represents the passive things in the IoT such that each node may communicate with other nodes using stored information that are propagated in the space and can be used by especial nodes using TCP/IP over wireless access points. The management of RFID network is distributed among many network controllers. In addition, the frequency that is used in this simulation equals a range between 860-960MHz. Huge collected data and dynamic positions of RFID nodes are considered the most important factors in the simulation of RFID [58]. Regarding the simulation of mobile ad hoc networks, the area of this network is square. The mobile ad hoc network nodes are classified into two categories, client and servers. The source of data (anycast requests) is considered as a client. Also, the client acts as an intermediate node that may be used to transmit the collected data. The node that replies to the client's request is considered as a server. The requests and replies are determined randomly.

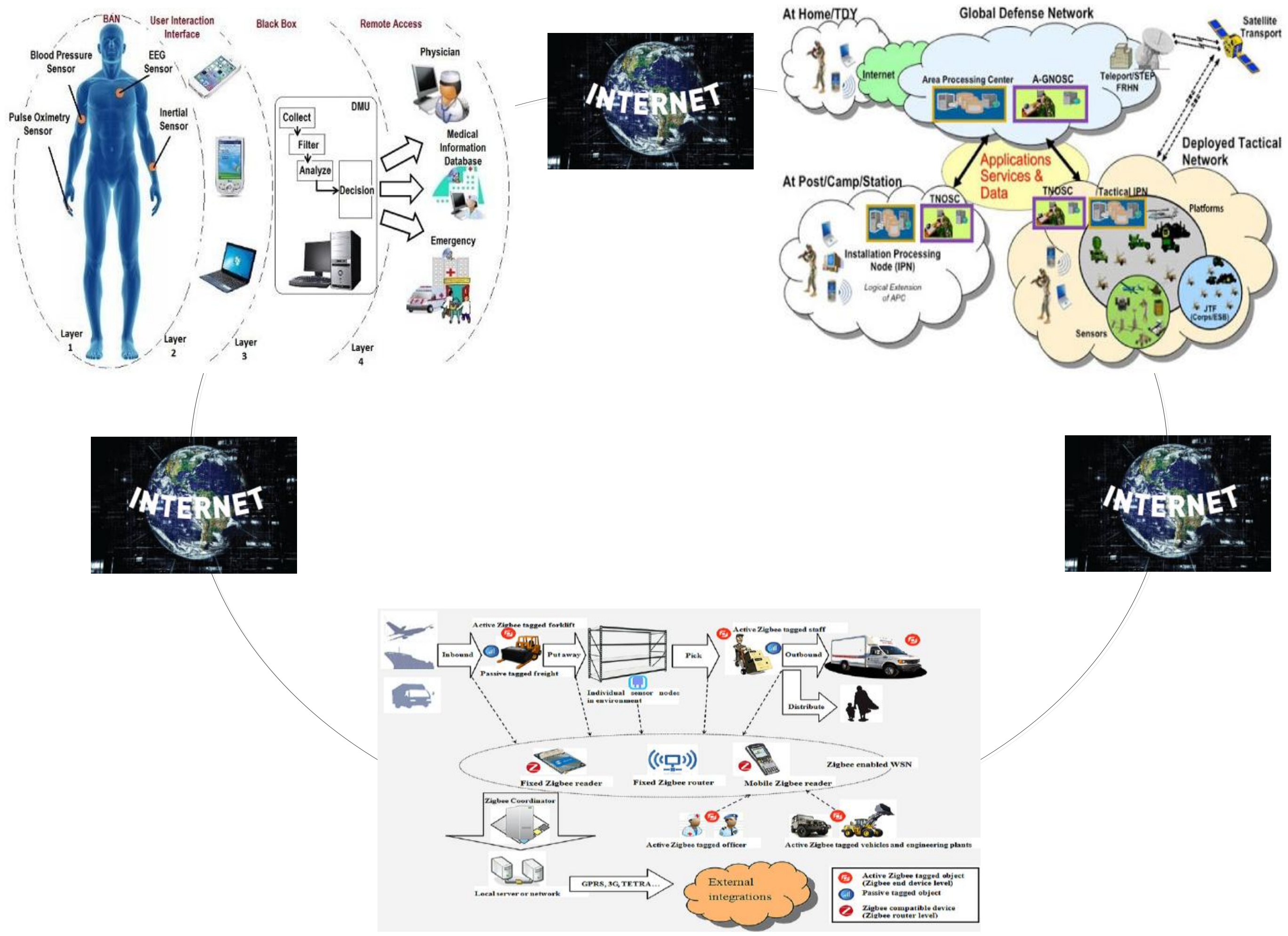

Figure 7: The general view of simulation model

In the mobile ad hoc networks simulation, the nodes initial positions are determined using uniform probability distribution. Furthermore, many multicast requests can be received by any server and replies may be initiated depending on the decision of acceptance or rejection of client request. Mobile node speed in mobile ad hoc network and their movements is determined using the random direction mobility model that has been applied in [58]. Distance between nodes is considered as a most important factor for mobile ad hoc network simulation model [58].The Internet is simulated as a collection of nodes, each with its own parameters, protocols, and statistics, see Fig 7. The routing mechanism which is used in this simulation is found at [59]. The description of IoT system before implementation of the proposed data reduction is found at [59]. The simulation experiment is run 20 times and the average of results is taken. The simulation time is 100 minutes.

Results and discussion

The performance metrics used to measure the efficiency of the proposed DRM are data reduction rate, energy consumption, end-to-end delay, and throughput. In this section, each performance metric results are shown and discussed.

Data reduction performance metric represents the most important metric. This is because it measures the effect of applying the proposed DRM on the IoT environment directly. This performance metric is evaluated by determining the number of packets which are transmitted through the IoT environment in the traditional mode; i.e. when the IoT don't apply the proposed DRM, and in the data reduction mode; i.e. when the data reduction is applied. Fig. 8 and table 2 show the results of the DRM effect. The X axis represents the number of transmitted packets. The Y axis represents the simulation time points in minutes. It is clear that the number of transmitted packets is decreased in case of data reduction mode when it is compared to that in the IoT tradition mode. This is due to the strategies which are implemented during the IoT simulated session. These strategies are least mean squares, compression and prioritization. Each strategy has a relation with the IoT status such as normal, congestion or starvation. Also, the packets are sent in IoT traditional mode but it is not guaranteed that these packets are reached corrupted to their destination. The corruption may be occurred due to congestion or network starvation. The hesitations in the two plots appear because the number of nodes in each packet, the time of sending, the size of network, and the destinations are determined randomly and changed periodically. For all of the simulation time points, the number of transmitted packets is decreased because the simulation supposed that the IoT has huge number of transmitted packets which may cause congestion and the DRM strategies are implemented.

The energy consumption performance metric is also considered as an important metric because when the number of transmitted packets increases the energy that is consumed for each node will be also increased. In addition, energy consumption is related to the life span of the IoT system. If the energy consumption rate is low, the life span of the nodes will be increased which affect the entire IoT system life span. Measurement of energy consumption could confirm the effect of applying the DRM on the IoT by clarifying the status of each node's energy and/or the average of energy consumption for different networks in the IoT environment. Fig. 9 and table 3 show the energy consumption rate in Jules. The X axis represents the average of energy consumptions. The Y axis represents the simulation time in minutes. It is notable that over the simulation time, the rate of energy consumption increases because the number of transmitted packets increases. The energy consumption rate is reduced when applying the proposed DRM. Also, in case of traditional IoT system, the increase of energy consumption rate over the simulation time has high rates. But, in case of applying the proposed DRM, the increase in the energy consumption rate is smooth over the simulation time.

The end-to-end delay performance metric clarifies the status of the entire IoT system. The end-to-end delay determines if the sent packets are received to the destination without delay (or within accepted delay), i.e. the IoT environment is healthy and the QoS required for packet transmission is sufficient. Queuing delay, processing delay, transmission delay, and propagation delay are considered in the calculation process of the end-to-end delay. Fig. 10 and table 4 show the results of end-to-end delay. The x axis represents the average of end-to-end delay in the IoT system. The Y axis represents the simulation time in minutes. It is clear that the end-to-end delay for the proposed DRM is less than the end-to-end delay for the traditional mode. This is due to the decrease in the number of transmitted packets through the IoT simulated environment. Decreasing the transmitted packets makes the IoT devices servicing each packet without delay. In addition, Fig. 10 shows the decrease of the end-to-end delay over the simulation time. The hesitations in the two plots reflect the random status of IoT environment components and parameters which may be changed over the time.

The throughput performance metric determines the number of packets which are transmitted and reached to their destinations without problems. So, the throughput reflects the general efficiency of the IoT system. Fig. 11 and table 5 show the throughput results. The X axis represents the number of packets that are transmitted correctly. The Y axis represents the simulation time in minutes. It is notably that the throughput for the proposed RDM is more than that for the traditional mode. This indicates to the high efficiency of the IoT system that is gained by decreasing the overload that may be caused by the transmitted packets. The number of transmitted packets is decreased in case of the traditional mode because the lost packets increase due to the IoT starvation that causes a decrease in the QoS.

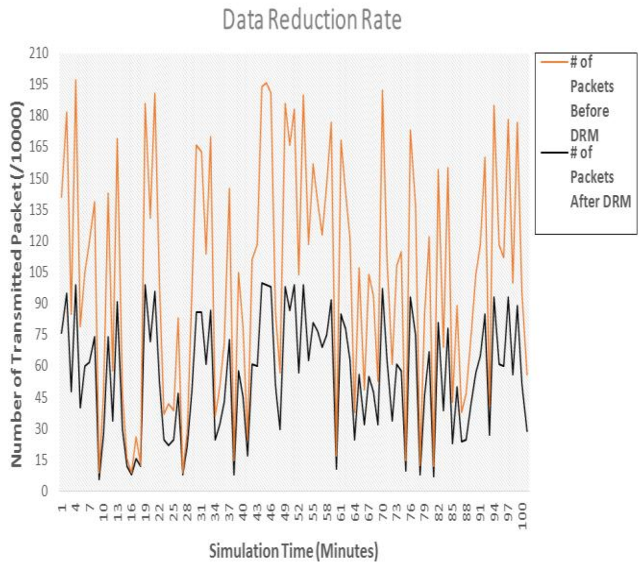

Figure 8: Data Reduction Rate.

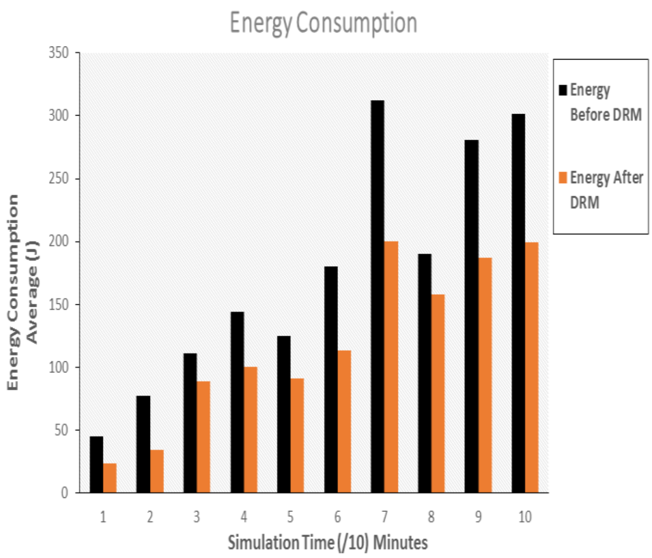

Figure 9: Energy Consumption.

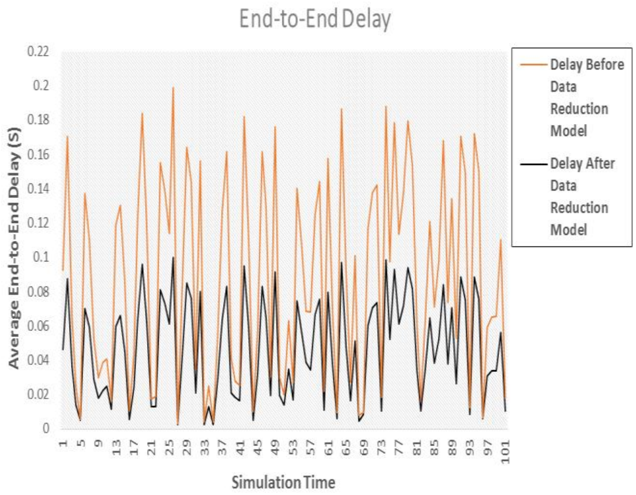

Figure 10: End-to-End Delay.

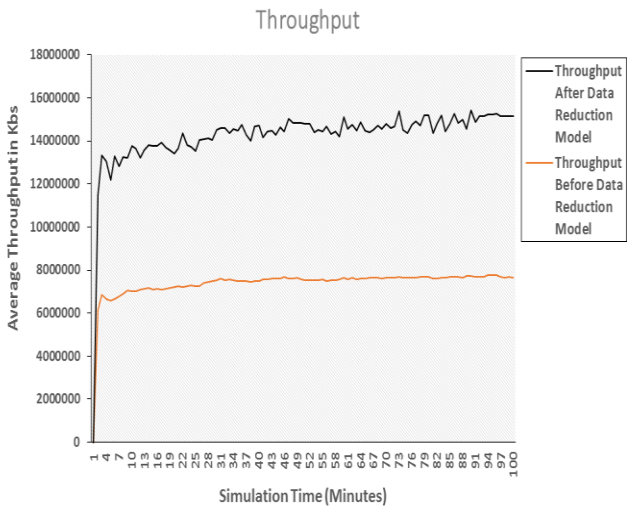

Figure 11: Throughput.

CONCLUSION

In this paper, an optimized data collection mechanism for IoT environments is presented. This mechanism is formalized in a model that is called Data Reduction Model. The proposed mechanism provides some recommendations and determines restrictions during collection, transmission, and processing of data to minimize its size. The proposed mechanism comprises three targets, dynamic prioritization for collected data, coding optimization, and energy saving. These strategies are employed to achieve the desired goal. IoT environment to simulate, test and evaluate the proposed mechanism is constructed using network simulator NS2. Accordingly, the simulation results ensured that the proposed method is robust and efficient. The simulation results proved that the implementation of DRM decreases the number of transmitted packets through IoT environment with 30.247% ↓. In addition, DRM decreases the energy consumption rate with 24.961% ↓. Furthermore, the end-to-end delay is decreased with 29.117% ↓, and the throughput is increased with 21.714% ↑.

ACKNOWLEDGEMENT

We are very grateful to Taif University for supporting this research.

REFERENCES

[1] F. Ducharme, “The Smarter Planet: Built on Informatics and Cybernetics”, Journal of Systemics, Cybernetics and Informatics, Volume 12, No. 5, 2014.

[2] T. Qu, et al., “Internet-of-Things-based just-in-time milk-run logistics routing system,” 2015 IEEE 12th International Conference on Networking, Sensing and Control (ICNSC), Taipei, Taiwan, pp. 258 – 263, 9-11 April 2015.

[3] A. Alvi, A. Shah, and W. Mahmood, “Energy efficient green routing protocol for Internet of Multimedia Things,” IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Singapore, pp. 1 – 6, 7-9 April 2015

[4] Z. Zhou, et al., “E-CARP: An Energy Efficient Routing Protocol for UWSNs in the Internet of Underwater Things,” IEEE Sensors Journal, Vol. PP, Issue: 99, 2015.

[5] X. Hua-Mei and Y. Kun, “Routing Protocols Analysis for Internet of Things,” 2nd International Conference on Information Science and Control Engineering (ICISCE), Shanghai, China, pp. 447 – 450, 24-26 April 2015.

[6] O. Said and M. Masud, “Towards Internet of Things: Survey and Future Vision,” International Journal of Computer Networks (IJCN), Volume 5, Issue 1, pp. 1-17, 2013.

[7] P. Diogo, P. Reis, and N. Lopes, “Internet of Things: A system's architecture proposal,” 9th Iberian Conference on Information Systems and Technologies (CISTI), Barcelona, Spain, pp.1 – 6, 2014.

[8] Y. Bo and H. Guangwen, “Application of RFID and Internet of Things in Monitoring and Anti-counterfeiting for Products,” International Seminar on Business and Information, Wuhan, Hubei, China, pp. 392- 395, 2008.

[9] J. Jiong, J. Gubbi, S. Marusic, and M. Palaniswami, “An Information Framework for Creating a Smart City Through Internet of Things,” IEEE Internet of

Things Journal, Volume 1, Issue 2, pp. 112 – 121, 2014.

[10] H. Yinghui and L. Guanyu, "A Semantic Analysis for Internet of Things," IEEE International Conference on Intelligent Computation Technology and Automation, Shenzhen, China, pp. 336- 339, 2010.

[11] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS '00), January 2000. DOI:10.1109/HICSS.2000.926982

[12] S. Lindsey and C. Raghavendra, "PEGASIS: Power-Efficient Gathering in Sensor Information Systems", IEEE Aerospace Conference Proceedings, 2002, Vol. 3, 9-16 pp. 1125-1130.

[13] J. Kim and J. Lee, "OpenIoT: An open service framework for the Internet of Things," IEEE World Forum on Internet of Things (WF-IoT), pp. 89 – 93, Seoul, South Korea, 6-8 March 2014.

[14] E. Zouganeli and I. EinarSvinnset, "Connected Objects and the Internet of Things – a Paradigm Shift," International Conference on Photonics in Switching, Pisa, Italy, pp. 1-4, 2009.

[15] X. Gordillo, G. Romero, E. Abasolo and A. Carrera, "Test bed for evaluating reference models of Internet of Things (IoT)," IEEE Colombian Conference on Communications and Computing (COLCOM), 4-6 June 2014, Bogota, Columbia, pp. 1 – 6, 2014

[16] Z. Tongzhu, W. Xueping, C. Jiangwei, L. Xianghai and C. Pengfei, "Automotive recycling information management based on the internet of things and RFID technology," IEEE International Conference on Advanced Management Science (ICAMS), Changchun, China, pp. 620 – 622, 2010.

[17] D. Muriel and F. Juan, "Expanding the learning environment: combining physicality and virtuality The Internet of Things for eLearning," IEEE International Conference on Advanced Learning Technologies (ICALT), Sousse, Tunisia, pp. 730-731, 2010.

[18] A. Luigi, I. Antonio, and M. Giacomo, "The Internet of Things: A survey," Science Direct journal of Computer Networks, Volume 54, Pages: 2787–2805, 2010.

[19] J. Jiong, J. Gubbi, S. Marusic, and M. Palaniswami, "An Information Framework for Creating a Smart City Through Internet of Things," IEEE Internet of Things Journal, Volume1, Issue 2, pp. 112 – 121, 2014.

[20] D. Virmani, et al., "Priority Based Energy-Efficient Data Forwarding Algorithm in Wireless Sensor Networks," International Journal of Advances in Computer Networks and its Security (IJCNS), Vol. 1, No 1, 2011.

[21] M. Monowar, et al., "Prioritized Heterogeneous Traffic-Oriented Congestion Control Protocol for WSNs," The International Arab Journal of Information Technology, Vol. 9, No. 1, January 2012.

[22] F. Ullah, "Patient Data Prioritization in the Cross-Layer Designs of Wireless Body Area Network", Journal of Computer Networks and Communications, Volume 2015, Article ID 516838, 21 pages, 2015.

[23] C. Cantillas, et al., "Queuing Schemes for Wireless Sensor Nodes Transmitting Prioritized Data", Presented at the DLSU Research Congress 2015, De La Salle University, Manila, Philippines, March 2-4, 2015.

[24] R. Karthikeyan, et al., "Priority based Packet Scheduling Approach for Wireless Sensor Networks", International Journal of Innovative Research in Science, Engineering and Technology , Volume 3, Special Issue 1, February 2014.

[25] M. Kumar, et al., "Dynamic Data Prioritization for Quality-of-Service Differentiation in Heterogeneous Wireless Sensor Networks", International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 11, November 2014.

[26] M. Kumar and V. Reddy, "Dynamic Data Prioritization Technique for Improving Quality of Service in Wireless Sensor Network," International Journal of Engineering Research & Technology, Vol. 3, No 11, November 2014.

[27] M. Nabi, et al., "Dynamic Data Prioritization for Quality-of-Service Differentiation in Heterogeneous Wireless Sensor Networks", 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), Salt Lake City, UT, USA, pp: 296 – 304, 27-30 June 2011.

[28] C. Liang, et al., "RushNet: Practical Traffic Prioritization for Saturated Wireless Sensor Networks", ACM SenSys (Conference on Embedded Networked Sensor Systems), pp: 105-118, New York, NY, USA, 2014.

[29] L. Li, R. Ramjee, M. Buddhikot, and S. Miller, "Network coding-based broadcast in mobile ad-hoc networks," in IEEE INFOCOM, pp. 1739–1747, May 2007.

[30] W. Lou and J. Wu, “On reducing broadcast redundancy in ad hoc wireless networks,” IEEE Transactions on Mobile Computing, vol. 1, no. 2, pp. 111– 122, Apr-Jun 2002.

[31] S. Yang and J. Wu, “Efficient broadcasting using network coding and directional antennas in MANETs,” IEEE Transactions on Parallel and Distributed Systems, vol. 21, no. 2, pp. 148–161, Feb 2010.

[32] S. Yang, J. Wu, and F. Dai, “Efficient backbone construction methods in manets using directional antennas,” in ICDCS, 2007.

[33] Z. Dong, C. Zhan, and Y. Xu, “Delay aware broadcast scheduling in wireless networks using network coding,” in IEEE NSWCTC, 2010, pp. 214–217.

[34] C. Zhan and Y. Xu, “Broadcast scheduling based on network coding in time critical wireless networks,” in IEEE International Symposium on Network Coding, June 2010.

[35] P. Ostovari, J.Wu, and A. Khreishah, “Deadline-aware broadcasting in wireless networks with local network coding,” in IEEE ICNC, Jan 2012.

[36] P. Ostovari, A. Khreishahand, and J. Wu, “Deadline-aware broadcasting in wireless networks with network coding,” in IEEE GLOBECOM, Dec 2012.

[37] C. Fragouli, J. Widmer, and L. Boudec, “A network coding approach to energy efficient broadcasting: From theory to practice,” in IEEE INFOCOM, 2006, pp. 1–11.

[38] C. Fragouli, et al., “Efficient broadcasting using network coding,” IEEE/ACM Transactions on Networking, vol. 16, no. 2, pp. 450–463, Apr 2008.

[39] D. Nguyen, T. Tran, T. Nguyen, and B. Bose, “Wireless broadcast using network coding,” IEEE Transactions on Vehicular Technology, vol. 58, no. 2, pp. 914–925, 2009.

[40] L. Lu, M. Xiao, M. Skoglund, L. Rasmussen, G. Wu, and S. Li, “Efficient network coding for wireless broadcasting,” in IEEE WCNC, 2010, pp. 1–6.

[41] W. Fang, F. Liu, Z. Liu, L. Shu, and S. Nishio, “Reliable broadcast transmission in wireless networks based on network coding,” in IEEE INFOCOM Workshops (INFOCOM WKSHPS), 2011, pp. 555–559.

[42] D. Welsh and M. Powell, “An upper bound for the chromatic number of a graph and its application to timetabling problems,” The Computer Journal, vol. 10, no. 1, pp. 85–86, 1967.

[43] D. Koutsonikolas, Y. Hu, C. Wang, M. Comer, and A. Mohamed, “Efficient online WiFi delivery of layered-coding media using inter-layer network coding,” in ICDCS, 2011, pp. 237–247.

[44] U. Horn, K. Stuhlm¨uller, M. Link, and B. Girod, “Robust internet video transmission based on scalable coding and unequal error protection,” Signal Processing: Image Communication, vol. 15, no. 1, pp. 77–94, 1999.

[45] D. Wu, Y. Hou, and Y. Zhang, “Scalable video coding and transport over broadband wireless networks,” Proceedings of the IEEE, vol. 89, no. 1, pp. 6–20, 2001.

[46] A. Majumda, D. Sachs, I. Kozintsev, K. Ramchandran, and M. Yeung, “Multicast and unicast real-time video streaming over wireless lans,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 6, pp. 524–534, 2002.

[47] D. Koutsonikolas, Y. Hu, C. Wang, M. Comer, and A. Mohamed, “On the performance of network coding in multi-resolution wireless video streaming,” in NetCod, 2010, pp. 1–6.

[48] M. Halloush and H. Radha, “Practical network coding for scalable video in error prone networks,” in PCS, 2009, pp. 1–4.

[49] L. Lu, M. Xiao, and L. Rasmussen, “Relay-aided broadcasting with instantaneously decodable binary network codes,” in ICCCN, 2011, pp. 1–5.

[50] H. Shwe and F. Adachi, “Power efficient adaptive network coding in wireless sensor networks,” in IEEE ICC, 2011, pp. 1–5.

[51] R. Chandanala and R. Stoleru, “Network coding in duty-cycled sensor networks,” in INSS, 2010, pp. 203–210.

[52] F. Reichenbach1, et al., “A Distributed Linear Least Squares Method for Precise Localization with Low Complexity in Wireless Sensor Networks,” Chapter in Distributed Computing in Sensor Systems, pp.514-528, 2006.

[53] M. Newville, et al., “Non-Linear Least-Squares Minimization and Curve-Fitting for Python Release 0.9.6,” book, 2017.

[54] S. Wang, et al., “Distributed Source Coding”, John Wiley & Sons Ltd, 2017.

[55] X. Liu , C. Liu, W. Liu, and X. Zeng, “Wireless channel modeling and performance analysis based on Markov chain,” 29th IEEE Control And Decision Conference (CCDC), Chinese, pp: 2256 – 2260, 28-30 May 2017.

[56] The Network Simulator – ns-2. 2008, Available from: http://www.isi.edu/nsnam/ns/

[57] A. Lewis, et al., "Using ant colony optimization to construct meander-line RFID antennas," Springer chapter, Biologically-Inspired Optimization Methods 2009; 210: 189–217.

[58] G. Di Caro, "Analysis of simulation environments for mobile ad hoc networks," Technical Report, No. IDSIA-24-03, Available from: http://people.idsia.ch/~gianni/Papers/IDSIA-24-03.pdf

[59] O. Said, "Analysis, design and simulation of Internet of Things routing algorithm based on ant colony optimization," International Journal of Communication Systems, Wiley, 2016, DOI: 10.1002/dac.3174

List of Abbreviations

Abbreviation	**Description**
IoT	Internet of Things
RFID	Radio-Frequency Identification
WSN	Wireless Sensor Network
PEDF	Priority-Energy Based Data Forwarding
PHTCCP	Prioritized Heterogeneous Traffic-oriented Congestion Control Protocol
WBAN	Wireless Body Area Network
QoS	Quality of Service
DDP	Dynamic Data Prioritization
DRM	Data Reduction Model
DSC	Distributed Source Coding
LT	Luby Transform
LDPC	Low-Density Parity Check