Middle-East Journal of Scientific Research
24 (Special Issue on Innovations in Information, Embedded and Communication Systems): 158-165, 2016
ISSN 1990-9233; © IDOSI Publications, 2016
DOI: 10.5829/idosi.mejsr.2016.24.IIECS.23155

Addressing Data Aggregation Issues in Wireless Sensor Network

T. Aanandhadurgadevi and S. Pavalarajan

PSNA College of Engineering and Technology, Dindigul, India

Abstract: Data aggregation in Wireless Sensor Network (WSN) aims to combine the several packets from sensor node in an efficient way by using different mathematical functions. This process will reduce the number of redundant information in the network. In order to prolong the network lifetime there is a need to reduce the redundant information using data aggregation techniques. In this paper we discussed the data aggregation approaches and the various performance measures. Also we discussed the issues in data aggregation and methods to solve those issues and finally we compared some data aggregation methods.

Key words: Wireless Sensor Network · Data Aggregation · Redundancy

INTRODUCTION

WSN consist of a huge number of independent sensor nodes to support a wide mixture of applications including biomedical health monitoring natural environment monitoring, surveillance and target tracking [1]. In a major WSN, each sensor node executes periodically some operations such as computing, sensing and transmits data to the sink node through multiple paths.

These sensor nodes usually have inadequate energy supply, e.g., batteries, which make energy efficiency a serious issue in sensor network design and sharing. Other resource constraints in WSN are short communication range, poor bandwidth, limited processing, storage and the energy consumption. Energy consumption is mainly address in the following stages: sensing, data processing and data transmission.

Generally, sensing and data processing have smaller amount energy consumptions than that of data transmission. To decrease the transmission cost, data aggregation scheme is used to enhance the lifetime of network. Data aggregation [2, 3] is well-defined as the process of aggregating the data from multiple sensors by using some aggregation functions such as MAX, MIN and SUM. This comforts in minimizing redundancy as well as the number of transmissions.

The aggregation function is affected by several factors such as aggregation function, topology used, selection of aggregation point and number of participated nodes in the network.

Data Aggregation Basics

Synopsis of Data Aggregation: Data aggregation is defined as the process of aggregating the data from various sensors to remove redundant broadcast and provide collective information to the base station [4, 5]. Data aggregation typically involves the combination of data from countless sensors at in-between nodes and transmission of the aggregated data to the base station, as shown in Fig. 1.

Data aggregation attempts to collect the most important data from the nodes and make it accessible to the base station in an energy efficient method with minimum data latency. Data latency is fundamental in many applications such as environmental monitoring, where the freshness of data is also a most important thing.

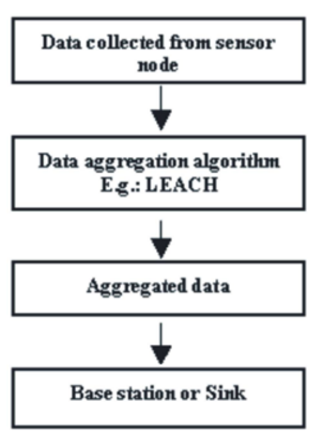

Fig. 1: Architecture of data aggregation.

Corresponding Author: T. Aanandhadurgadevi, PSNA College of Engineering and Technology, Dindigul, India.

Aggregation Function: One of the most vital functionality that aggregation techniques should provide is the ability to combine data coming from different sensor nodes.

There are several types of aggregation functions and most of them closely related to specific sensor application. Aggregation function classified into two types.

Duplicate sensitive and duplicate insensitive: Aggregation function could be average and minimum. If we use average function, it takes as a duplicate sensitive and minimum function is taken as duplicate insensitive function in WSN. If we used duplicate function, then the end result depends on the number of the times and same value has been considered. Or else aggregation function is assumed to be duplicate insensitive.

Lossy and Lossless: Data packet can be get-together with the help of lossy aggregation or by lossless aggregation. Lossy aggregation approach does not follow an ideal reconstruction but lossless aggregation ensures a complete reconstruct of all individual sensor data at base station.

Data Aggregation Based on Network Architecture: The design of data aggregation protocol for WSN needs extra care in terms of energy consumption. Usually, the energy dissipation of the node mostly depends upon the architecture of the WSN. Based on network architecture, broadly, a WSN is classified into flat network and hierarchical network and it is shown in Table 1.

Flat Networks: In flat networks, each sensor node plays the similar task and is furnished with more or less the same battery power. In such networks, data aggregation is done by data centric routing where the sink usually transmits a query message to the sensors, e.g., via flooding and nodes which have data matching the query send response messages back to the sink.

Hierarchical Networks: Flat network can result in too much communication and computation load at the sink node resultant in quicker depletion of its battery power. The expiry of the sink node breaks down the functionality of the network. Hence, in view of scalability and energy utilization, several hierarchical data aggregation approaches have been proposed. Hierarchical data aggregation involves data combination at special nodes, which reduces the number of messages transmitted to the base station. This improves the energy efficiency of the network.

Performance Metrics: The design of WSN is affected by many factors such as scalability, production cost, fault tolerance, operating environment, network topology, hardware constraints, transmission media and power consumption. To validate the performances, simulation parameters used are described below.

Network Lifetime: Network lifetime is well-defined as the number of data aggregation rounds till α % of nodes dies where α is specified by the scheme designer. For instance, in applications where the time that all nodes function collectively is crucial, lifetime of the network is measures with first sensor node die.

Packet Delivery Ratio (PDR): PDR is the ratio of the number of packets received by the sink to the total number of packets produced by the entire sensor node.

Latency: Latency is defined as the delay involved in data aggregation, routing and data transmission. It can be measured as the time delay between the data packets received at the base station and the data generated at the source nodes.

Energy Efficiency: The functionality of the WSN should be extended as long as feasible. In an ideal data aggregation system, each sensor should have expended the equal amount of energy in every data gathering round.

Table 1: Data Aggregation in Hierarchical Networks VS Flat Networks.

S.No	Hierarchical networks	Flat networks
1	Data aggregation performed by cluster heads or a sensing leader node.	Data aggregation is performed by different nodes along the multi-hop path.
2	Overhead involved in cluster or chain formation throughout the network.	Data aggregation routes are formed only in regions that have data for transmission.
3	Even if one cluster head fails, the network may still be operational.	The failure of sink node may result in the breakdown of entire network.
4	Lower latency is involved since sensor nodes perform short range transmissions to the cluster head.	Higher latency is involved in data transmission to the sink via a multi-hop path.
5	Routing structure is simple but not necessarily optimal.	Optimal routing can be guaranteed with additional overhead.
6	Node heterogeneity can be exploited by assigning high energy nodes as cluster heads.	Does not utilize node heterogeneity for improving energy efficiency.

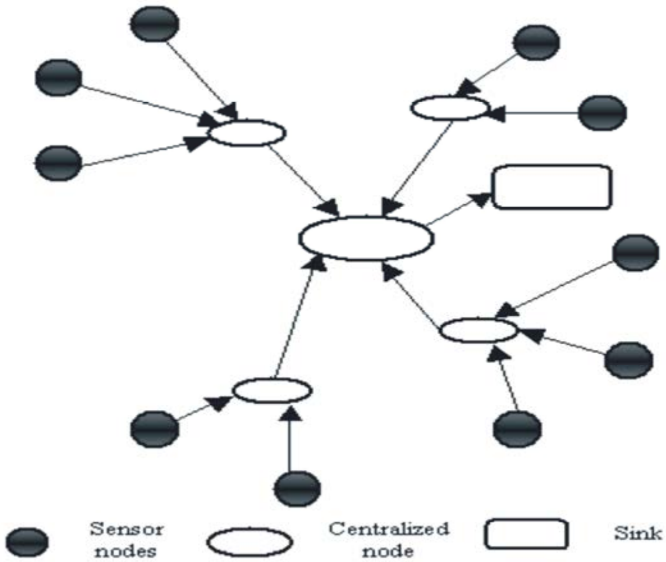

Fig. 2: Centralized based data aggregation

A data aggregation method is energy efficient if it maximizes the functionality of the sensor network. If we guess that all sensor nodes are uniformly essential, we should decrease the energy use of each sensor.

Bandwidth, Capacity and Throughput: These point out the capacity of data which can be sent over a link within a given time. It can be measured with the number of bits received at the sink.

Data Aggregation Approaches: Data aggregation is performed based on data aggregation approach. There are different approaches for data aggregation, which is described as follows.

Centralized Approach: In this approach each sensor sends data to central node through the shortest path. All the sensor nodes simply send the data packets to a node, which is have highest energy among all other nodes. This node is called as centralized node. This node performs aggregation of data coming from all sensors and sends the aggregated packets to sink, as shown in Fig. 2.

Tree-Based Approach: In a tree based approach, sensor nodes are organized into a tree for aggregating data packets. The type of tree is minimum spanning tree, sink node consider as root and source nodes consider as a leaves. Sensed information flows from leave node to sink and the aggregation done by intermediate node. This approach is suitable for application which involved in-network data aggregation, as shown in Fig. 3. Drawback of this approach, data packets loss at any level of tree the data will be dropped not only for single level but for whole related as well.

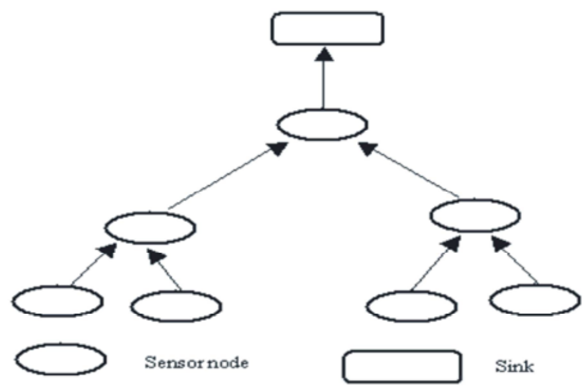

Fig. 3: Tree base data aggregation

Cluster Based Approach: In energy constrained WSN of large size, it is in efficient for sensors to transmit the data directly to base station. To overcome this problem cluster based network is used to forward the packets in energy efficient way. In this approach network is divided into a number of cluster. For each cluster one cluster head (CH) is chosen for aggregating the data and send the aggregated packets to base station, as shown in Fig. 4. This results in significant energy savings for the energy constrained sensors. The cluster head communicate with the base station directly or multi hopping through other cluster heads.

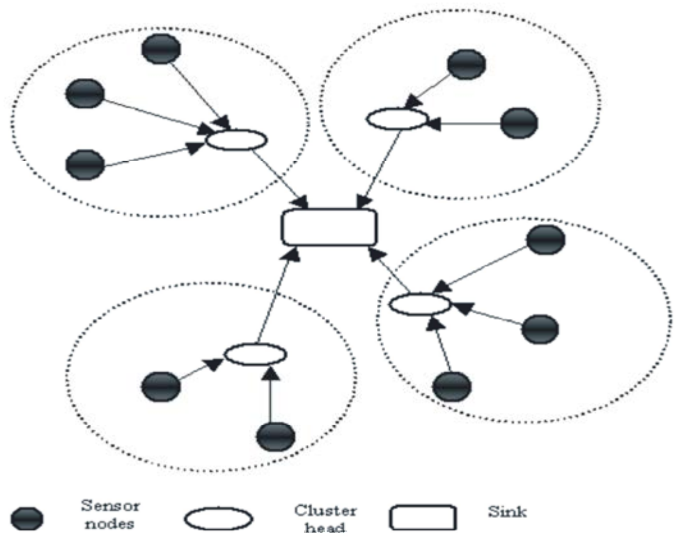

Fig. 4: Cluster based data aggregation

Multi-path Approach: In this approach, entire nodes can send data packets to its possibly multiple neighbors. Hence data packet flow from sensor node to the sink node along multiple path, lot of intermediate nodes between sensor nodes to base station so aggregation done in every intermediate node. Using this approach we will make the network robust but some additional overhead is involved. The example of this system is similar to ring topology, where sensor network is separated in to concentric circle with defining levels according to hop distance from sink.

Issues in Data Aggregation: Data aggregation techniques address issues like redundancy, delay, accuracy, traffic load and energy. Performance of the data aggregation technique gets affected by these issues. In our paper, we discuss open issues and methods to solve these issues.

Redundancy Elimination: In WSN, sensors are overlap with each other in sensing region. Due to this deployment, identical objects are generated by sensor nodes and forward it to sink. This redundant transmission of data to the sink increases the energy consumption. So to reduce the energy consumption and prolong lifetime of the network there is need of redundancy elimination.

Intra cluster and inter cluster aggregation is used to reduce the redundancy. In this approach data aggregation is done at CH using aggregation functions and forwards the aggregated data to sink. During intra cluster aggregation redundancy inside the cluster is eliminated at CH. After intra cluster aggregation, each CH considered as individual node for performing inter cluster aggregation.

In the following paragraph, we review the redundancy removal approaches.

TTCDA [6] - The authors proposed Two Tier Cluster based Data Aggregation (TTCDA) algorithm for randomly distributed sensor nodes in the network to reduce computation and communication cost. The TTCDA effectively considered the packets and data aggregation using additive and divisible aggregation functions at CH and base station. The aggregation functions are applied according to temporal spatial and correlation of packets and data aggregated by each sensor node. It also reduces transmission of redundant data by improving energy and bandwidth utilization.

The TTCDA algorithm works with three phases for aggregating the packets. Cluster formation, intra cluster aggregation and inter cluster aggregation. The performance of TTCDA is measured by aggregating packets generated with variable rate (PGR) and real data in the packet.

BECPA [7] - The author projected a novel Bandwidth Efficient based Packet Aggregation Algorithm (BECPA) to decrease the number of transmission of packets from sensor node to sink. The proposed model uses the heterogeneous nodes in terms of energy. It focuses on the symmetric aggregation function which is completely compressible. The aggregation function used considers the correlation of number of packets generated at an uneven rate and data within the packet by each node in the cluster.

The performance of BECPA is calculated by aggregating the uneven rate of packets produced with consideration of network lifetime, energy consumptions and bandwidth utilization.

BECDA [8] - The author proposed Bandwidth Efficient Cluster-based Data Aggregation (BECDA) algorithm presents the solution for the efficient data gathering with in-network aggregation. It considers the network with heterogeneous sensor nodes in terms of energy and mobile sink to aggregate the data packets. BECDA considers the WSN model that refers to cluster-based aggregation, since clustering enhance the scalability by stabilizing the network topology.

The best approach is achieved by intra and inter-cluster aggregation on the randomly scattered nodes with the uneven data production rate.

The proposed algorithm uses the correlation of data within the packet for applying the aggregation function on the data aggregated by sensor nodes.

Delay: Delay is defined as time taken by recipient to collect a packet is higher than the time taken by the sender to forward a packet. Data aggregation function decrease the number of packet transmission but increase delay because data from closer source node may have to be held back at an in-between node in order to be aggregated with data approaching from sources that are farther away. The latency due to aggregation will be relative to the number of hops between the sink and the farthest source.

Tree based methods used to decrease the delay. In tree based system when source node wants to send the data to target, data may flow from various paths consumes high bandwidth and requires more delay to reach target. Minimum spanning tree provide a path between source nodes to target which has minimum weight. This path will provide fast delivery of data to destination which reduces delay.

Following method decrease the transmission delay of the sensor nodes.

Fast Aggregation Scheduling [9] - Energy efficiency is a significant issue for data collection in WSNs. Data aggregation is a fundamental operation aiming to save energy by dropping the number of packet transmissions through the network. Besides the energy efficiency, the necessity of real-time communication is becoming more and more vital in emerging applications. Here, outdated information would be unrelated and even lead to harmful effects on the system.

Thus, it is important to provide a guarantee on the delivery time and consider the total latency involved in aggregating data.

Minimizing latency is of major problem for efficient data aggregation in WSNs. In this paper, to minimize time latency, authors focus on aggregation scheduling problem and proposed an efficient distributed algorithm that generates a collision-free schedule with the least number of time slots. In this paper, authors designed FAST as a novel collision-free minimum latency data aggregation scheduling algorithm for tree-based structures. To get a best time slot assignment, FAST simultaneously execute the aggregation scheduling and balanced C3DS-based tree construction.

Energy: The significant challenge in WSN is to reduce the energy consumption in data gathering and forward from sensor node to the base. In following section, we describe approaches for energy efficiency in data aggregation schemes.

MADG-SIA [10] – Wireless sensor networks adopting static data gathering may suffer from uneven energy use due to non-uniform packet relay. Even though mobile data aggregation provides a reasonable approach to solving this difficulty, it inevitably introduces longer data gathering latency due to the use of mobile data collectors. In the meanwhile, energy harvesting has been considered as a promising result to relieve energy limitation in WSN.

In this paper, authors consider a joint design of these two schemes and proposed a novel two layer heterogeneous architecture for wireless sensor networks, which consists of two types of nodes: sensor nodes which are static and powered by solar panels and cluster heads that have limited mobility and can be wirelessly recharged by power transporters. Based on this network architecture, they present a data gathering design, called mobility assisted data gathering with solar irradiance awareness (MADG-SIA), where nodes are clustered around cluster heads that adaptively adjust their positions according to solar irradiance and the sensing data are send to the data sink by these cluster heads functioning as data aggregation points. The performance of the proposed method by extensive simulations and the results demonstrate that MADG-SIA provides major development in terms of balancing energy utilization and the amount data gathered.

Efficient clustering-based data aggregation [11]-In wireless sensor network applications for surveillance and reconnaissance, large amounts of repeated sensing data are regularly generated. It is important to control these data with efficient data aggregation techniques to decrease energy consumption in the network. To overcome these problems, authors proposed two hybrid clustering based data aggregation mechanisms. The collective clustering based data aggregation mechanism can apply multiple clustering techniques simultaneously in a single sensor network depending on the network environment. The adaptive clustering-based data aggregation mechanism can adaptively choose an appropriate clustering technique, depending on the status of the network. The proposed mechanisms can enlarge the data aggregation efficiency as well as improve energy efficiency and other significant problems.

Traffic Load: In WSN, the task of data aggregation is given to the cluster head. If this task of cluster head is fixed to one node then that node suffers from serious traffic load. The node wastes maximum of its energy due to this traffic.

Following methods describe traffic load due to data aggregation scheme.

ADANC [12]-In an energy-constraint WSN, the sensor nodes close to the sink drain their energy rapidly due to serious traffic and create a bottleneck zone. In a huge monitoring region, the sensors in a cluster may communicate with the cluster head using multi-hop approach. A bottleneck effect can also be formed near the cluster heads in a cluster based WSN with a number of clusters. The clustering system exploits the data correlation among the sensors' data and provides opportunities for decrease of energy use in WSN. Integrating an energy efficient system, namely, network coding, with clustering and duty cycling may facilitate the design of a new cluster based data collection scheme. The author proposed an energy efficient adaptive data aggregation policy using network coding (ADANC) which improves the energy efficiency in a cluster based duty-cycled WSN. A of nodes act as network coder nodes and rest of the nodes operates as simple relay nodes in a cluster. The network coder nodes also act as aggregation points, opportunistically, based on the level of data correlation. The proposed ADANC strategy provides decrease of traffic within a cluster and thus raises the energy efficiency of the bottleneck zone.

A Stochastic Time-Domain Model [13] - In numerous network applications, the nature of traffic is of burst type. Often, the transient reply of network to such traffics is the result of a sequence of mutually dependent events whose occurrence forecast is not a minor task. The previous efforts in IEEE 802.15.4 networks often followed top-down designs to form those sequences of actions, i.e., through making top-view models of the whole network, they tried to follow the transient reply of the network to burst packet arrivals. The problem with such methods was that they were unable to give station-level views of network response and were usually difficult. In this paper, authors proposed a non-stationary analytical model for the IEEE 802.15.4 slotted CSMA/CA medium access control (MAC) protocol under burst traffic arrival assumption and without the optional acknowledgements. They developed a station-level stochastic time-domain method from which the network-level metrics are extracted. The bottom-up method makes finding station-level details such as delay, collision and failure distributions feasible. Moreover, network-level metrics like the average packet failure or transmission success rate can be extracted from the model.

Accuracy: In data aggregation process aggregator node or cluster head performs aggregation. The aggregated data is then forwarded in to the network. So if this aggregator node gets compromised, may lead to inaccuracy.

Following method validate the sensor nodes and allows only valid node to forward data in order to provide accuracy in aggregation process.

Lossy data aggregation integrity scheme [14]-Lossy data aggregation assumes that when an aggregation node gathers data from other nodes, some or all sensed data of a node may be dropped. Thus the final data received by the base station are fragmented. They assume that the measurements of different nodes have different weights for their contributions to aggregated data. In this paper, authors proposed a secure identity-based lossy data aggregation integrity method based on homomorphic hashing and identity-based aggregate signature. The system enables the base station to share a distinct key with all sensor nodes, such that all aggregation nodes besides the base station can validate the authority of aggregated data.

In this scheme, both the aggregation nodes and the base station can compute the weights of sensor nodes, neither via an established agreement nor from explicit indications attached to data.

Comparative Study of Data Aggregation Methods: In this section we have done comparative study of data aggregation methods. These methods are mainly focusing

Table 2: Comparatision of Data Aggregation Methods

Methods	Approach	Redundancy	TL	EC	Accuracy
TTCDA	Cluster	Less	Moderate	Less	Moderate
BECPA	Cluster	Less	Moderate	Less	Moderate
BECDA	Cluster	Less	Moderate	Less	Moderate
Fast Aggregation Scheduling	Tree	Moderate	Moderate	Less	Moderate
Efficient Cluster-Based data aggregation	Cluster	Moderate	Moderate	Less	Moderate
ADANC	Cluster	High	Less	Moderate	Moderate
Stochastic Time- Domain Model	Cluster	Less	Less	Less	Moderate
Lossy data aggregation integrity scheme	Cluster	Moderate	Moderate	Less	High

on to address the above stated issues. Table 2 Shows the comparison of various methods based on approaches, redundancy, traffic load(TL), energy consumption(EC) and accuracy.

CONCLUSION

We have presented a complete review of data aggregation in WSN. We have labelled the aggregation functions, network Architecture for data aggregation. Also we have conferred data aggregation approaches and issues. Data aggregation technique has issues like redundancy, delay, energy, accuracy and traffic load that are needed to be addressed. Further we presented aggregation methods to solve the issues and made comparative study of various data aggregation methods.

REFERENCES

1. Jennifer Yick, Biswanath Mukherjee and Dipak Ghosal, 2008. Wireless sensor network survey. Computer Networks, 52(12), 22 August 2008, pp: 2292-2330.
2. Kiran Maraiya, Kamal Kant and Nitin Gupta, 2011. Wireless Sensor Network: A Review on Data Aggregation. International Journal of Scientific and Engineering Research, 2(4), April-2011 ISSN 2229-5518.
3. Susmedha Sirsikar and Samarth Anavatti, 2015. Issues of Data Aggregation Methods in Wireless Sensor Network: A Survey. Proceedings of 4th International Conference on Advances in Computing, Communication and Control (ICAC3'15), 49: 194-201.
4. Paulo Jesus, Carlos Baquero and Paulo Sérgio Almeida, 2015. A Survey of Distributed Data Aggregation Algorithms. IEEE communication surveys and tutorials, 17(1), first quarter 2015.
5. Elena Fasolo, Michele Rossi, Jorg Widmer and Michele Zorzi, 2007. In-network Aggregation Techniques for Wireless Sensor Networks: A Survey. Wireless Communications, 14(2): 70-87, April 2007 ISSN 1536-1284.
6. Dnyaneshwar Mantri, Neeli Rashmi Prasad and Ramjee Prasad, 2013. Two Tier Cluster Based Data Aggregation (TTCDA) for Efficient Bandwidth Utilization in Wireless Sensor Network. Wireless Pers Commun, pp: 2589-2606, November 2013.
7. Dnyaneshwar Mantri, Neeli Rashmi Prasad and Ramjee Prasad, 2014. BECPA: Bandwidth Efficient Cluster Based Packet Aggregation in Wireless Sensor Network. Wireless Pers Commun, March 2014.
8. Dnyaneshwar Mantri, Neeli Rashmi Prasad and Ramjee Prasad, 2014. BECDA: Bandwidth Efficient Cluster Based Data Aggregation in Wireless Sensor Network. computer and Electrical Engineering, pp: 256-264, September 2014.
9. Hamed Yousefi, Marzieh Malekimajd, Majid Ashouri and Ali Movaghar, 2015. Fast Aggregation Scheduling in Wireless Sensor Networks. IEEE Transactions on Wireless Communications, 14(6): 3402-3414, JUNE 2015.
10. JiLi, Yuanyuan Yang and Cong Wang, 2015. Mobility assisted datagathering with solar irradiance awareness in heterogeneous energy replenishable wireless sensor networks. Computer Communications, pp: 88-97, July 2015.
11. Woo-Sung Jung, Keun-Woo Lim, Young-Bae Ko and Sang-Joon Park, 2011. Efficient clustering-based data aggregation techniques for wireless sensor networks. Wireless Netw, pp: 1387-1400, May 2011.
12. Rashmi Ranjan Rout, A. and Soumya K. Ghosh, 2013. Adaptive data aggregation and energy efficiency using network coding in a clustered wireless sensor network: An analytical approach. Computer Communications, pp: 65-75, December 2013.

13. Mohammad Sayad Haghighi, Yang Xiang, Vijay Varadharajan and Barry Quinn, 2015. A Stochastic Time-Domain Model for Burst Data Aggregation in IEEE 802.15.4 Wireless Sensor Networks. IEEE Transactions on Computers, 64(3), MARCH 2015.
14. Shufen Niu, Caifen Wang, Zhixuan Yu and Suzhen Cao, 2012. Lossy data aggregation integrity scheme in wireless sensor networks. Computers and Electrical Engineering, pp: 1726-1735, December 2012.