Efficient Technique for Network Lifetime Enhancement by Cleaning Dirty Data

Komal V. Shiyale[1], Pranay D. Saraf[2]

[1, 2] Computer Science & Engineering, G. H. Raisoni College of Engineering, Nagpur, India

Abstract: ***In the current scenario, wireless sensor and network (WSN) is used in substantial number of applications. It co-operatively sends the sensed data to the base station. This continuous sensing may give rise to the problem of dirty data i.e. missing data, non-ordered data, outlier, redundant data etc. This may lead to energy consumption to large extend and causes inaccuracy in the application where real time and accurate data is most essential factor. Therefore, efficient utilization of power and maintaining the accuracy in data is must in order to use networks for long duration. To address it, an integration of Redundant transmission of sensed data technique and Belief based data cleaning technique has been proposed for improving efficiency in communication and reducing time delay as well as communication overhead in WSN. It also aims at regenerating the missing data, cleaning dirty data, non ordered data, outliers, etc to gather the necessary information which also enhances network lifetime by efficient use of battery power.***

Keywords: dirty data, energy consumption, network lifetime, cleaning

1. Introduction

In recent years it is seen that real-time requirements for timely actions has become the basic need in any network. Therefore, many new techniques are been developed in pervasive computing, communication and sensing technologies and give rise to the emergence of Wireless sensor and actor networks (WSANs). It is a group of sensors (mobile or static) and actors (mostly mobile, e.g. Sub-Kilogram Intelligent Tele-robots (SKITs), Autonomous Battlefield Robot designed for the Army, etc) which are wirelessly connected with each other and perform distributed sensing and actuation tasks [1].

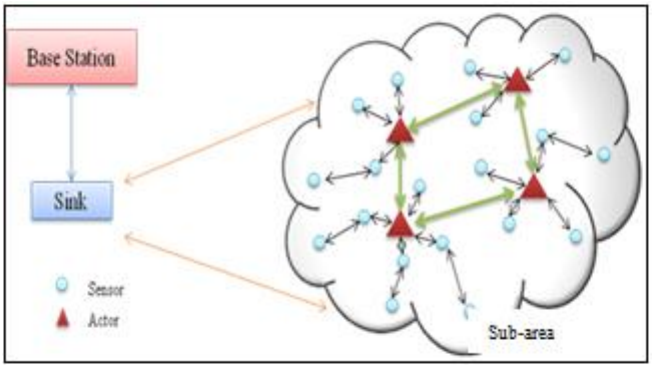

Figure 1: Physical architecture of WSAN

Fig. 1 shows the physical architecture of the WSAN. Here sensors are passive elements with limited energy, low cost small device, and low processing and communication capabilities which gather information from physical environment. Were as, actors are active elements with higher energy, longer battery life, processing and communication capabilities, and is responsible for taking decisions and later performing appropriate actions on the environment. In this way actors remotely interact with the environment. WSAN is the combination of sensor and actor nodes, which is illustarted in equation (1)

Sensors + Actors = WSANs (1)

There are many applications of WSAN like

1) Environmental Applications (e.g. detecting and extinguishing forest fire).
2) Climate control in buildings (e.g. for detection of temperature by the sensors and then trigger the audio alarm actors in that area)
3) Distributed Robotics & Sensor Network (e.g. Mobile robots in sensor network)
4) Battlefield Applications (e.g. detection of mines or explosive substances)

There are many factors because of which WSAN is chosen over WSN. They are as follows:

- Many applications need real time data as well as data must be valid at the time of action.
- Network deployment can be heterogeneous i.e. sensors nodes can be densely deployed and actors nodes can be loosely deployed
- Co-ordination between sensors and actors is the main requirement. There are basically two types of coordination a) sensor-actor and b) actor-actor coordination.

a) **Sensor-Actor coordination:** here sensors sense the data from the environment and then transmit the data to actor nodes.

b) **Actor-Actor coordination:** after receiving the data from the sensors actor may or may not coordinate with the nearby actors and make decision to perform appropriate action.

Energy efficiency and accurate data is an important concern in Wireless Sensor Networks. We have studied various data cleaning approaches and their effectiveness in enhancing lifetime of Wireless Sensor. Improved Cross-Redundant Data Cleaning Algorithm (ICRDC) [8] and Redundancy Elimination for Accurate Data Aggregation (READA) [9] are used to eliminate redundant data. READA also compresses the data before sending and it can also behave as an event detection system and report if an unusual behavior is noted Adaptive Filter-based Data Cleaning [10] and Online Data Cleaning method [11] are used to clean the outliers, missing information, and noise accurately. Extensible Sensor stream Processing (or ESP) [12] and Belief-based cleaning method [2] are used to clean the sensor's unwanted data and the dirty data for the target sensor successfully. Sketch-based data cleaning method [13] and Redundant transmission of sensed data technique [3] will allow the base station to recover the value of missing data.

Based on the analyzed parameters, we tried to integrate two techniques i) Redundant transmission of sensed data technique[3] and ii) Belief based data cleaning technique[2]. The motivation for the proposed system is described in section II. Section III explains our proposed method in detail. The performance and analysis is then illustrated in section IV. The conclusion and future scope is finally shown in section V.

2. Motivation

There are few issues in Wireless Sensor and Actor Network (WSAN). Firstly, when two nodes transmits message or data at the same time this may lead to collision and therefore message will be retransmitted, which will simultaneously lead to wastage of bandwidth and hence energy. Secondly, dirty data (non ordered data, outliers, missing data, noisy data) is caused because of limited resources of sensors, weak wireless multi-hop communication and node mobility.
Therefore, in our proposed work we would focus on:

- To replace the dirty data readings with the reading from group of sensors obtained from the belief table, which is believed to be offering enough reliable readings at specific time interval.
- To reduce the missing data in the network by embedding neighbor nodes sensed value along with sensors own value.
- To reduce communication overhead by NCK (negative acknowledgment) only if the packet is not received instead of every time sending ACK (acknowledgement) when is packet is received.
- To evaluate and simulate the performance of the proposed method.

3. Proposed Work

Whole work is basically divided in two main parts. First part mostly deals with reducing the missing data or data loss in the network and also reducing the communication overhead. Second part mostly concentrates on cleaning the dirty data by using belief based cleaning method. Following assumption are done before proceeding:
1) Assume a pre-processing technique to detect the dirty data.
2) Boundaries of the subareas are already known.
3) All the data is correctly received at the base station.

3.1 Reducing Missing Data or Data Loss[3]:

The acknowledgement mechanisms (RTS/CTS/ACK) are not properly utilized in wireless multicast and broadcast transmission because this transmission would cause higher communication traffic and overhead [4]. This may result in more energy consumption. Moreover, when two sensor nodes transmit the data at the same time there are chances of collision. And when collision occurs the same data must be retransmitted to avoid data loss. But retransmission of data may lead to formation of redundant data as well as bandwidth is also wastage. This further reduces the network lifetime.

Therefore, redundant data transmission protocol (RT) [5] is utilized to reduce the communication overhead and increase the network lifetime. Here, neighbor sensor node's sensed value is embedded with its own value and then transmitted. In mica2dot [6] the message format of sensor node is 32 bytes long. Fig. 2 shows the message format of sensor node. This free space of 24 bytes can be utilized to send the neighbor nodes sensed value along with sensor's own sensed value as shown in Fig. 3. Thus, retransmission of messages is avoided and redundant broadcast transmission of messages is done where the sensor node, send data by embedding data received at t, t-1… t-n to the actor node. So there is less chances of message drop and even if one transmission fails, the message reaches the destination (actor) via other message transmission.

Source ID (2 bytes)	Sequence ID (2 bytes)	Destination ID (2 bytes)	Current Sensed value (2 bytes)	Free Space (24 bytes)

Figure 2: Message format of sensor node

Source ID (2 bytes)	Sequence ID (2 bytes)	Destination ID(2 bytes)	Current Sensed value (2 bytes)	Neighbour nodes sensed value (24 bytes)

Figure 3: Message format embedded with neighbor sensed value

Here the problem of communication overhead is also reduced by introducing negative acknowledgment (NCK) concept instead of flooding acknowledgment (ACK) signals [3]. In short, the data transmission protocol in this work, avoids sending ACK for every message. Instead, NCKs are sent for the missing messages.

3.2 Elimination of Dirty Data

As it is already assumed that there is pre-processed technique to detect dirty data. Certain no. of sensors is deployed in a specific area. These sensors are further divided into multiple subareas. Readings of sensors in same subarea is assumed to be similar were as it is different from adjacent subarea. Boundaries of subareas are already known to the base station. Here firstly a group of neighboring sensors is selected for collaborating in the cleaning process. For this belief parameters are used which depends on the consistency of their streaming data correctly received at the base station and sensor trajectories. And then the dirty sample is cleansed based on the distance function in both time and location of sensors.

1) Belief Based Sensor Selection
Sensor selection is done based on belief degree that is how trust worthy a sensor could help cleaning the dirty data at

specific time within the subarea. Belief degree is based on two parameters
a) Alibi degree (A)
b) Detection rate of dirty data (D)

a) Alibi degree (A)
It is the measure of how many times the sensor has been in the given subarea i.e. higher the alibi degree the more no. of times the sensor has been operated in given sub area. It is computed by residence vector and the frequency of existence in the sub-area.

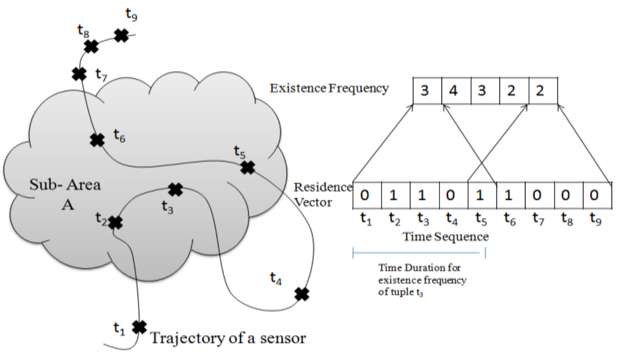

Figure 4: Alibi Degree calculation

- Residence vector: Residence vector is series of existence of sensor located in a sub-area. Sensor existence is computed from trajectory data of each sensor received by the base station. This series is stored in free space in the memory. Depending on the location of the sensors they are differentiated in Boolean types like:
  1: when the sensor is located inside the sub-area
  0: when the sensor is located outside the sub-area
  To calculate residence vector, the locations of the moving sensors are noted in the window of size equals to 9 at time instants from t_1 to t_2 as shown in figure 4. Hence the residence vector is [0 1 1 0 1 1 0 0 0].

- Frequency Vector: It is the frequency at which the sensor appears in the specific subarea. From figure 4 at time instant t_5 the existence frequency is sum of members of residence vector from t_3 to t_7 and so on.

b) Detection rate of dirty data (D)
The quality of the data is highly affected by the trajectory of the sensors. Detection rate of dirty is inversely proportional to reliability of the data i.e. lower the detection rate the more is the data reliable. It is formulated as:

Detection rate of dirty data is formulated =

$$\frac{\text{culuative no.of detected dirty samples}}{\text{all the samples that the sensor measures}} \quad (2)$$

- Belief degree calculation and sensor selection

Belief degree is directly proportional to alibi degree and is inversely proportional to detection rate of dirty data. Depending on alibi degree and detection rate of dirty data belief degree is calculated and is updated to the belief table of specific subarea. Depending on the application a certain threshold belief degree is assumed. The sensors whose belief degree is greater than the threshold belief degree are selected for cleaning purpose.

2) Belief Based Cleaning Method
Depending on the time difference and the distance between the selected sensors and the target sensor, the dirty data is replaced by the cleansed data. Lower the time difference and distance between selected sensors and the target sensor, more similar would be the data between them. In this way the dirty data is replaced by the cleansed and most appropriate data.

4. Performance and Evaluation

In this section, we evaluated the performance of the proposed system. For this, we used NS2 simulator [7], where nearly 90 sensor nodes, 10 actor nodes and 1 base station is deployed as shown in Fig. 5. As it is assumed that the dirty data (outliers, missing data, non-ordered data, etc) is already detected on the sensors which are blue in color. Pink color nodes are actor nodes were as red color node is base station. The sensors selected for cleaning purpose is denoted by yellow color. Here, all the nodes are mobile.

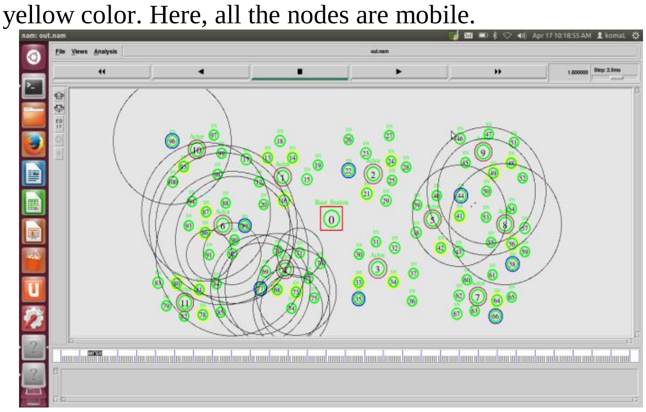

Figure 5: Network Setup

We have integrated two algorithms to clean the dirty data, reduce the redundant data and avoid the missing data. And our system's performance was compared to that of Belief-based cleaning technique. Fig. 6 shows Average Delay Graph, were the delay in sending the packets from source to destination is shown in proposed system and existing system. Fig. 7 shows Average Throughput Graph i.e. the rate at which the data is successfully delivery over a communication channel. Fig. 8 shows Energy Consumption Graph i.e. how efficiently energy is been consumed in both proposed and existing system.

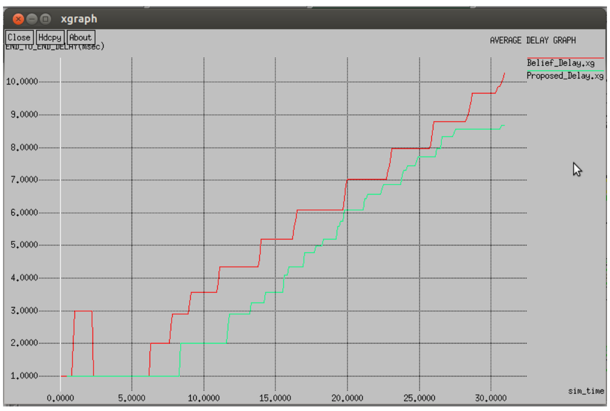

Figure 6: Average Delay Graph

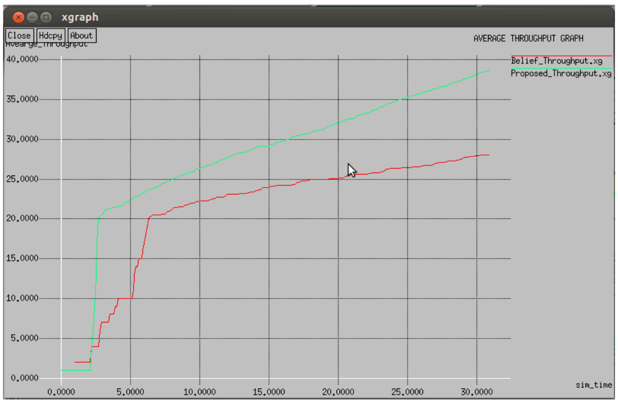

Figure 7: Average Throughput Graph

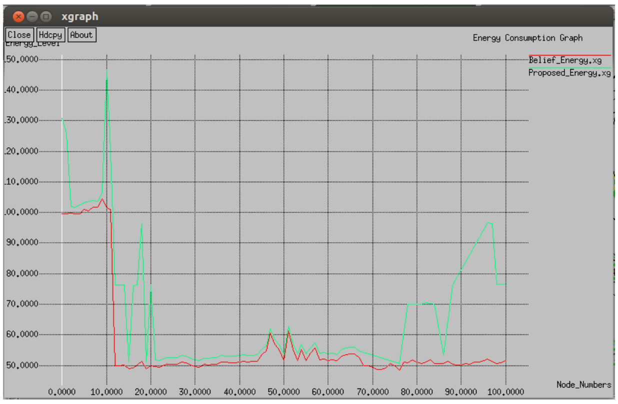

Figure 8: Energy Consumption Graph

5. Conclusion and Future Scope

Efficient utilization of available energy is important concern in Wireless Sensor Network. Unreliable and dirty data is one of the factor which leads to more amount of energy consumption. In the proposed work, we have tried to integrate two algorithms to clean the data efficiently and reduce the consumption of energy to much extend by solving the issue of missing data, dirty data and redundant data in the mobile environment. As we are sending only NCK when the message is not received much communication overhead is also reduced. Our simulation result shows that the amount of energy consumed by our system is reduced by 20% (approximately).

The work can further be extended to develop some improved and less complicated algorithms, which requires less calculations and simulation time.

References

[1] F. Akyildiz, I. H. Kasimoglu, "Wireless Sensor and Actor Networks: Research Challenges," Ad Hoc Networks Journal (Elsevier), Vol. 2, No. 4, pp. 351-367, October 2004.
[2] Sitthapon Pumpichet, Niki Pissinou, Xinyu Jin and Deng Pan, "Belief-based Cleaning in Trajectory Sensor Streams," in IEEE ICC - Ad-hoc and Sensor Networking Symposium, 2012.
[3] Selvaradjou, K. Bharathi Manjula, "Redundant Transmission of Sensed data in Wireless Sensors and Actor Networks for Improved Reliability," in International Journal of Emerging Science and Engineering (IJESE), Volume-1, Issue-7, May 2013.
[4] M. Demirbas, S. Balachandran, "ROBCAST: A Single hop Reliable Broadcast Protocol for Wireless Sensor Networks", in Proceedings of The Sixth International Workshop on Assurance in Distributed Systems and Networks (ADSN2007).
[5] Kiyohiro Morita, Kenichi Watanabe, Naohiro Hayashibara, and Makoto Takizawa, "An Efficient Data Transmission Protocol in a Wireless Sensor-Actuator Network", in Proceedings of the First International Conference on Complex, Intelligent and Software Intensive Systems (CISIS'07), 2007.
[6] Wireless sensor networks Getting Started Guide, Crossbow Technology, Inc., 2004. Available at: https://www.eol.ucar.edu/rtf/facilities/isa/internal/CrossBow/Doc/Getting_Started_Guide_7430-0022-04_A.pdf
[7] Sirwan A.Mohammed, Dr. Sattar B.Sadkhan, " DESIGN OF WIRELESS NETWORK BASED ON NS2," Journal of Global Research in Computer Science, Volume 3, No. 12, December 2012

Author Profile

Komal V. Shiyale is pursuing Masters of Engineering in Mobile Technology (MT) Department of Computer Science and Engineering (CSE) from G.H. Raisoni College of Engineering, Nagpur, MS, India. She received Bachelor of Engineering (BE) degree in Electronics Department from Rajiv Gandhi College of Engineering, Research and Technology, Chandrapur, MS, India. Her research interests include Wireless Sensor Network, Radio Frequency Identification etc.

Pranay D. Saraf received his Bachelor of Engineering (BE) degree in Information Technology from.RTMNU Nagpur, MS, India. He received his Master of Engineering degree in Mobile Technology (MT) Computer Science and Engineering (CSE) from G.H. Raisoni College of Engineering, Nagpur, MS, India. He is currently working as an Assistant Professor in the Department of Computer Science and Engineering. He has totaled 6 National and International papers published. His research interest includes VANET, Mobile Communication.