Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 8, August 2014, pg.34 – 40

RESEARCH ARTICLE

Distributed Data Collection Scheme for Store and Forward Information in Wireless Sensor Network

Madhavi S. Kukade[1], Prof. Kapil N. Hande[2]

[1]M.Tech.-CSE, Priyadarshini Bhagwati College of Engg, Nagpur, Maharashtra, India

[2]CSE Department, Priyadarshini Bhagwati College of Engg, Nagpur, Maharashtra, India

[1] sakshikhamankar@gmail.com; [2] kapilhande@gmail.com

Abstract— Sensors networks are capable of collecting an enormous amount of data over space and time often, the ultimate objective is to "sample, store and forward" that is to sense the data, store it locally and ultimately forward it to accent almost and analysed. Typical sensor nodes are wireless nodes with limited storage and computational power. Furthermore they are prone to "failure" by going out of wireless range, interference running out of battery etc. They can be deployed in isolated or dangerous areas to monitor objects, temperatures, etc. or to detect fires, floods, or other incidents. There has been extensive research on sensor networks to improve their utility and efficiency. The sensor and storage nodes are distributed randomly in some region and cannot maintain routing tables or shared knowledge of network topology. Some nodes might disappear from the network due to failure or battery depletion. A distributed data collection algorithm to accurately store and forward information obtained by wireless sensor networks is proposed. The proposed algorithm does not depend on the sensor network topology or geographic locations of sensor nodes, but rather makes use of uniformly distributed storage nodes. Analytical and simulation results for this algorithm show that, with high probability, the data disseminated by the sensor nodes can be precisely collected by querying any small set of storage nodes.

Keywords— Distributed data collection algorithm, Storage nodes, Sensor nodes, WSNs

I. Introduction

Wireless communication is the transfer of information over a distance without the use of electrical conductors or cables. The distances involved may be short or long. Wireless operations permits services, such as long-range communications, that are impossible with the use of wires. Information is transferred in this manner over both short and long distances. Wireless sensor networks (WSNs) often consist of small devices (nodes) with limited processing ability, bandwidth and power. They can be deployed in isolated or dangerous areas to monitor objects, temperatures, etc. or to detect fires, floods, or other incidents. [16].There has been extensive research on sensor networks to improve their utility and efficiency. For WSN applications, the data in the sensors are often streamed to a control centre (called sink). This process is called data collection [1]

One of the most important functions provided by Wireless Sensor Networks (WSNs) is directly gathering data from the physical world. Generally, data gathering can be categorized as data collection, which gathers all the data from a network without any data aggregation or merging, and data aggregation, which obtains some aggregation values, e.g. MAX, MIN, SUM, and etc. To evaluate network performance, network capacity which reflects transmission/collection/broadcast it usually adopted,[18][19]e.g. multicast capacity unicast capacity, broadcast capacity, and data collection capacity. For data collection capacity, it is defined as the average data receiving rate at the sink, i.e. data collection capacity reflects how fast data been collected by the sink [3]

A distributed data collection algorithm to accurately store and forward information obtained by wireless sensor networks is proposed. The proposed algorithm does not depend on the sensor network topology where there is arrangement of a network, including its nodes and connecting lines,[16] routing tables in contains the information necessary to forward a packet along the best path toward its destination. Each packet contains information about its origin and destination.

Data collection and selection is common to many systems, which vary greatly not only in scale, but in sensing modalities—i.e. what data is collected, and what mechanism is used acquire that data. Some sensor networks, for instance, monitor only simple physical phenomena such as temperature and humidity. In others the data source may be of higher speed and richer content, providing far more information than a slowly changing scalar value. The nature of the data being collected varies as well, from physical processes like temperature or radar to collection of output from computer applications.In a large scale wireless sensor network with a set of sensing nodes and a set of storage nodes. and one of the most important applications for wireless sensor networks (WSNs) is Data Collection,[10][11] where sensing data are collected at sensor nodes and forwarded to a central base station for further processing. The sensing nodes have limited memory and bandwidth, and they might disappear from the network at any time due to limited battery lifetime.[17] The storage nodes have large memory and bandwidth, but they do not sense information about the region.[16]We assume that the data collector (base station) is far away from the nodes as shown in Fig. 1, but it is connected with a set of storage nodes. The sensor nodes are able to sense data and distribute it to the storage nodes.

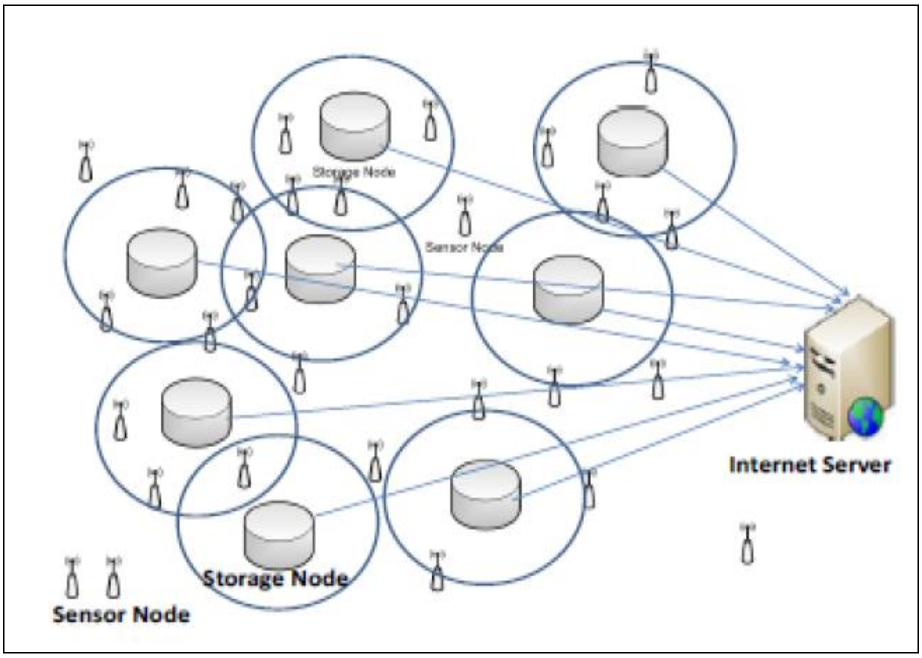

FIGURE 1. NETWORK MODEL REPRESENTING A WIRELESS SENSOR WITH SENSING AND STORING NODES, AND BASE STATION NODES

II. Problem Definition

Sensor networks are especially useful in catastrophic or emergency scenarios such as floods, fires, terrorist attacks or earthquakes where human participation may be too dangerous. However, such disaster scenarios pose an interesting design challenge since the sensor nodes used to collect and communicate data may themselves fail suddenly and unpredictably, resulting in the loss of valuable data.[10] Furthermore, because these networks are often expected to be deployed in response to a disaster, or because of sudden configuration changes due to failure, these networks are often expected to operate in a "zero-configuration" paradigm, where data collection and transmission must be initiated immediately, before the nodes have a chance to assess the current network topology.

Sensors networks are capable of collecting an enormous amount of data over space and time. Often, the ultimate objective is to "sample, store and forward" that is to sense the data, store it locally and ultimately forward it to a central host (or "master node" where data from other sensor nodes is also collected and analyzed.[8][17]A useful example is a traffic sensing network, there being traffic sensors at each intersection that estimate the traffic and relay it to a central processing station. Typical sensor nodes are wireless nodes with limited storage and computational power. Furthermore, they are prone to "failure", by going out of wireless range, interference, running out of battery etc. When a sensor node fails, the data it was storing is lost.

In a cooperative sensor network, it is a good idea to have nodes' data duplicated and spread around the network so it can be recovered from other nodes in case of failure. In particular, every node can store some of its own data as well as data from other nodes up to its storage capacity. The sensor and storage nodes are distributed randomly in some region and cannot maintain routing tables or shared knowledge of network topology.[17][1] Some nodes might disappear from the network due to failure or battery depletion.Traditional error-correcting erasure codes can also be used to achieve the goal of encoding data for collection if some of the encoding symbols are lost, data can still be recovered. Reed-Solomon codes are block erasure codes that have been traditionally used for error correction.[19]

III. Literature Survey

Author Shouling Ji, Zhipeng Cai provided method that functions by Wireless Sensor Networks (WSNs) is directly gathering data from the physical world. Generally, data gathering can be categorized as data collection, which gathers all the data from a network without any data aggregation or merging, and data aggregation, which obtains some aggregation values, e.g. MAX, MIN, SUM, and etc. For data collection capacity, it is defined as the average data receiving rate at the sink, i.e. data collection capacity reflects how fast data been collected by the sink. Without confusion, author use data collection capacity and network capacity interchangeably throughout this paper.

Author Abhinav Kamra, Jon Feldman, Vishal Misra and Dan Rubenstein discuss the benefits of storing combinations of data instead of original data has been studied in various works . Traditional error-correcting erasure codes can also be used to achieve the goal of encoding data such that if some of the encoding symbols are lost, data can still be recovered. Reed-Solomon codes are block erasure codes that have been traditionally used for error correction.

From the "Geographic Protocols in Sensor Networks" written by Karim Seada, Ahmed Helmy described In wireless sensor networks, Geographic protocols, that take advantage of the location information of nodes, are very valuable for sensor networks.

Mu Lin, Chong Luoyz, Feng Liu and Feng Wuz proposed a work for a large scale wireless sensor network where sensor readings are occasionally collected by a mobile sink, and sensor nodes are responsible for temporarily storing their own readings in an energy-efficient and storage efficient way. In the development of compressive data persistence scheme, they design a distributed compressive sensing encoding approach based on Metropolis-Hastings random walk.

Author Xiaohua Xu, Xiang-Yang Li, Min Song shown efficient query-selection algorithm in their paper where carefully selecting a subset of queries such that the total weight of selected queries is at least a constant fraction of the optimum solution when the load of all queries exceed the network capacity (*i.e.*, the WSN is overloaded with queries from control applications) .Here all the above are the different methods related with data collection

IV. Network Model and Assumption

We consider a large scale wireless sensor network with a set of sensing nodes $S = \{s1, . . . , sk\}$.All sensor nodes have the same capabilities such as mobility, homogeneous, limited memory and power. and a set of storage nodes both are distributed randomly and uniformly in a square region R = L×L, where L is the side length. Each storage node has a memory buffer of size *M* and this buffer can be divided into smaller buffers, each of size *c*, such that e = *[M/c]*. For simplicity we assume that all storage nodes have equal memory size *M*. Each storage node *ri* can send multicasting messages to neighboring nodes[16].

Each node *ri* can detect its total number of neighbors by sending a simple flooding query message, and any sensor node that responds to this message will be a neighbor of this node. Every node s_i prepares a packet $packet_{si}$ with its ID_{si} , sensed data x_{si} , and a *flag* that is set to zero or one so that packet contain its id ,sensed data information and flag value which show that sensor node has which data as we will consider two different types of packets depending on the *flag* value: initialization and update packets. If the source node sends a packet and the *flag* is set to zero, then it will be considered as an initialization packet. Otherwise, it will be considered as an update packet.

$$packet_{si}\ (ID_{si},\ x_{si},\ flag)$$

V. Distributed Data Collection Algorithm

We consider a wireless sensor network N with n nodes among which S = {s1, . . . , sn} are sensing node and R = {r1, . . . , rn−k} are storage nodes in network simulator model (NS2). The sensor and storage nodes are distributed randomly in some region R and cannot maintain routing tables or shared knowledge of network topology. Some nodes might disappear from the network due to failure or battery depletion.

In this paper, we will propose a distributed data collection algorithm for overcoming the storage problem by using NS2 simulator which help to create the desired network model The clustering storage algorithm runs in the following phases:

i) Clustering phase: We assume that the sensor network has $k/n \approx 80\%$ sensing nodes, and $(n - k)/n \approx 20\%$ storage nodes. All clusters in the network are established using clustering algorithms In the clustering phase, each storage node sends a flooding beacon message with its ID to all neighboring nodes in the network. Due to the random locations of the sensing nodes, some nodes will be able to receive this message and reply with their IDs to the storage nodes. In addition the sensing nodes will store the IDs of the storage nodes in which they received beacon messages:

$$packet\ ri \rightarrow S(ID_{ri})$$

ii) Sensing phase In the sensing phase, the sensor nodes sense data from the environment. Once the data is collected, they send their packets to the storage node, from which they have received beacon packets:

$$packet_{si} \rightarrow R_{si}\ (ID_{si},\ x_{si}, R_{si},\ flag)$$

where R_{si} is the set of storage nodes with whom s_i is connected. The flag value determines whether the packet contains an update or initially sensed data. The update data from the sensing nodes will occur whenever they sense new information about the surrounding environment

iii) Data collection and storage phase: When a sensing node senses the environment, it sends its packets to its storage nodes. The storage nodes collect the incoming packets and store them encoded in their own buffer. Based on the type of the incoming packets, the storage nodes will store these packets or update the existing data in their buffers.

iv) Querying phase: The query process can be performed by the base station or server that collects all data from the storage nodes. Total number of nodes that must be queried in order to obtain the data sensed by the sensor nodes.

Algorithm

Step: 1 INITIALIZATION
// Note: '-model' means network- model'
INITIALIZE network-model;
INITIALIZE sensor node S = {s_1, . . . , s_k}also acts like source nodes,
Where k is source packet { x_{si}, . . . , x_{sk}}
INITIALIZE n – k storage nodes R = {r_1, r_2, ..., r_{n-k}};
INITIALIZE storage buffers = y_1, y_2. . . y_{n-k} for all storage nodes *R*.
Step 2: LOOP
For each storage node r_i = 1 to n − k do
Generate a beacon packet with its ID_{ri} and send flooding message to all sensing neighbors;
Every sensing node will decide the storage nodes which to connect ;
End the process.
Step 3: FOR DATA COLLECTION
For each *source node* s_i, i = 1 ……. k do
Generate header of x_{si} and flag = 0 ;
Prepare the packet s_i ;
Send the packet s_i to storage nodes for data collection;
End the process
Step 4: IF source packets are remaining then
For each *node* r_j receives packets do
if the flag=0 then
Put x_{si} into r_j ’s buffer ;
End the process

Step 5: ELSE
Update the y_j buffer of the storage node r_j
$y_j = y_j \pm x_{si}$;
End the process.

Fig: 2- Distributed Data-Collection Algorithm

VI. Simulation Results

In third paper we overcome the duplication of data problem as when storage node collect the data first it check whether that data has been already present in its buffer or not if yes then it doesn't take that data and if not then it collect it and update in its own buffer. This is the distributed data collection process where storage node collect data from every sensor node those who reply to it and collect the data and forward to base station , this results are shown with the help of simulation in this module.

In the following figure we see the graphical result how storage node collect the data from each sensor node and store it in its own buffer.

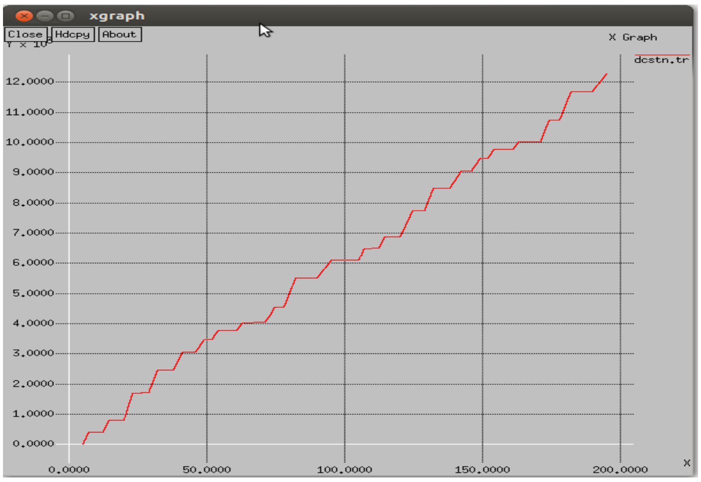

Fig.3 graphical view for storage node while collecting data

In the next step when base station collect the data from storage node it apply query process according to their buffer size of data present in storage node and collect the data from storage node,for this collection base staion done indexing also as on the priority basis. the following figure shows simulation results and the graphical view how base station collect the data from storage node.

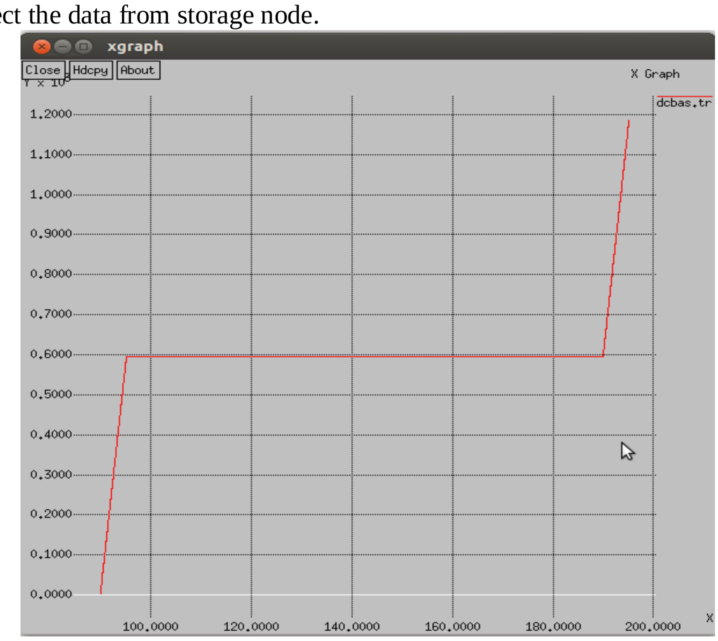

Fig .4 graphical view for base station while collecting data from storage node

VII. Conclusion

From the above paper we concluded that the technology for implementing distributed data collection is progressing well. Numerous investigators have addressed important aspects of this difficult problem, including researchers in data collection techniques. The initial success of these techniques is to collect the data from the sensor nodes which are uniformly distributed in the network, which depend on different domain structure. The acceptability of data collection algorithm is ultimately depending on results of simulation. We proposed the distributed data collection concepts for collect the data from the network where sensor nodes are uniformly distributed. We presented a result with network simulator which determines a performance baseline and the simulation results demonstrate that the proposed model is suitable for large-scale wireless sensor networks also with high probability, querying of the storage nodes with limited or unlimited buffers will retrieve all sensed data gathered by the sensing nodes. We evaluated a distributed data collection algorithm using terms and concepts as storage node collect the data and given to base station.

References

[1] Xiaohua Xu, Xiang-Yang Li, Min Song “Real-time Data Collection Scheduling in Multi-hop Wireless Sensor Networks,” IEEE Global Communications Conference 2013

[2] Mingsen Xu, Wen-Zhan Song, “Ravine Streams: Persistent Data Streams in Disruptive Sensor Networks” IEEE SECON (2013) Key: citeulike:12237835

[3] Shouling Ji, Zhipeng Cai, “Distributed Data Collection and Its Capacity in Asynchronous Wireless Sensor Networks,” IEEE INFOCAM 2012

[4] X. Hou, Z. Sumpter, L. Burson, X. Xue, and B. Tang, “Maximizing Data Preservation in Intermittently Connected Sensor Networks,” in IEEE International Conference on Mobile Ad hoc and Sensor Systems, Oct. 2012.

[5] M. Valero, M. Xu, N. A. Mancuso, W.-Z. Song, and R. Beyah, “EDR2: A Sink Failure Resilient Approach for WSNs,” in IEEE International Conference on Communications, Jun. 2012.

[6] A. G. Dimakis, V. Prabhakaran, and K. Ramchandran. "Decentralized erasure codes for distributed networked storage". IEEE/ACM Transactions on Networking, 14(SI):2809 – 2816, June 2006

[7] A. G. Dimakis, V. Prabhakaran, and K. Ramchandran. "Distributed fountain codes for networked storage". In Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Toulouse, France, May 14-19, 2006.

[8] A. Kamra, V. Misra, J. Feldman, and D. Rubenstein." Growth codes: Maximizing sensor network data persistence". In Proc. 2006 Conference on Applications, Technologies

[9] Z. Kong, S. A. Aly, and E. Soljanin." Decentralized coding algorithms for distributed storage in wireless sensor networks". IEEE Journal on Selected Areas in Communications, 28(2):261– 267, Feburary 2010.

[10] Y. Lin, B. Liang, and B. Li."Data persistence in large-scale sensor networks with decentralized fountain codes". In Proc. Twenty-Sixth Annual Joint Conference of the IEEE

[11] M. Goyeneche, J. Villadangos, J.J. Astrain, M. Prieto, and A. Cordoba. "A distributed data gathering algorithm for wireless sensor networks with uniform architecture". In Proc. 15th Euromicro International Conference on Parallel Distributed and Network-Based Processing (PDP’07), pages 373–380. IEEE Computer Society, Los Alamitos, CA, USA, 2007.

[12] O. Younis and S. Fahmy." Distributed clustering in ad hoc sensor networks: A hybrid, energy-efficient approach". In Proc. Twenty-Third Annual Joint Conference of the IEEE Computer and Communication Societies, volume 1, pages 629–640, Hong Kong, China, March 7-11, 2004

[13] Y. Lin, B. Li, and B. Liang. "Differentiated data persistence with priority random linear code". In Proc. 27th International Conference on Distributed Computing Systems (ICDCS’07), Toronto, Cana

[14] U. Niesen, P. Gupta, and D. Shah, "On Capacity Scaling in Arbitrary Wireless Networks" IEEE Transactions on Information Theory, 55(9): 3959-3982, 2009.

[15] G. Sharma, R. Mazumdar, and N. B. Shroff, "Delay and Capacity Trade- Offs in Mobile Ad Hoc Networks: A Global Perspective", IEEE/ACM Transactions on Networking, 15(5): 981-992, 2007.

[16] Salah A. Aly, Ahmed Ali-Eldin, H. Vincent Poor," A Distributed Data Collection Algorithm for Wireless Sensor Networks with Persistent Storage Nodes " 4th IFIP International Conference on NewTechnologies, Mobility and Security(NTMS) 7-10 February 2011 Paris, France

[17] Karim Seada, Ahmed Helmy," Geographic Protocols in Sensor Networks " The 3rd ACS/IEEEInternational Conference on Computer Systems and Applications, 2005.

[18] Mu Lin, Chong Luo, Feng Liu and Feng Wu.″ Compressive Data Persistence in Large-Scale WirelesSensor Networks″ IEEE Global Telecommunications Conference (GLOBECOM 2010)

[19] Shouling Ji,Zhipeng Cai.”Distributed Data Collection and Its Capacity in Asynchronous Wireless Sensor Networks” IEEE INFOCOM 2012

[20] M. Xu, W.-Z. Song, and D. Heo, “OnCode: Opportunistic In-Network Coding and Delivery in Energy-Synchronized Sensor Networks,” IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013

[21] S. A. Aly, H. Darwish, M. Youssef, and M. Zidan. "Distributed flooding based storage algorithms for large-scale wireless sensor networks.".IEEE International Conference on Communications, Dresden, Germany, June 13-17, 2009.

[22] S. A. Aly, Z. Kong, and E. Soljanin. Fountain codes based distributed storage algorithms for wireless sensor networks. IEEE/ACM Information Processing of Sensor Networks (IPSN’08), pages 171–182, St. Louis, MO, USA, April 19-22, 2008.

