Data Gathering and Location Inaccuracy in Sensor Network

Ahmed Gaffar
Zentrum for Informatics
University of Goettingen
Email: ahmed_gaffar@yahoo.com

September 15, 2004

Abstract

Sensor networks have a lot of applications in our daily life, in medicine, biology, and in the military. This essay will introduce the design and analysis of hybrid indirect transmissions (HIT) for data gathering, and the modeling and analyzing of the impact of location inconsistencies on geographic routing in the sensor networks. This essay will show that HIT is based on hybrid architecture that consist of one or more clusters, and how it can save energy in sensor network, also how is the effect of location inaccuracy on the geographic routing in the sensor networks.

1 introduction

Sensor networks have improved themselves in the world of technical in the last several years. Most of the new research in wireless networks searching in problems of energy saving in data gathering, and location inaccuracy, but what is the definition of data gathering; Sensor networks consisting of nodes with limited battery power and wireless communications are deployed to collect useful information from the field. Gathering sensed information in an energy efficient manner is critical to operating the sensor network for a long period of time. A data collection problem is defined where, in a round of communication, each sensor node has a packet to be sent to the distant base station. There is some fixed amount of energy cost in the electronics when transmitting or receiving a packet and a variable cost when transmitting a packet which depends on the distance of transmission. If each node transmits its sensed data directly to the base station, then it will deplete its power quickly, this is the base of thinking of HIT.

In the side of location inaccuracy; Sensor nodes have to be aware of their location to be able to specify "where" a certain event takes place. Therefore, the problem of localizing the sensors is of paramount importance for many classes of sensor network applications.
As a small description of a sensor network, A wireless sensor network consists of a number of sensors spread across a geographical area. Each sensor has wireless communication capability and some level of intelligence for signal processing and networking of the data. Some examples of wireless sensor networks are the following:(1):Military sensor networks to detect and gain as much information as possible about enemy movements, explosions, and other phenomena of interest [3].(2):Sensor networks to detect and characterize Chemical, Biological, Radiological, Nuclear, and Explosive (CBRNE) attacks and material[4].(3):Sensor networks to detect and monitor environmental changes in plains, forests, oceans, etc [5].(4):Wireless traffic sensor networks to monitor vehicle traffic on highways or in congested parts of a city.(5):Wireless surveillance sensor networks for providing security in shopping malls, parking garages, and other facilities[5]. HIT is one of the research results in this area, and also the effect of contradiction on the geographic routing. The rest of this Essay is organized as follows. In section 2, Backgrounds where the Essay defines the basics models that HIT is based on, and the four inaccuracy metric. In section 3, presents HIT as an approach in data gathering. A study for location inaccuracy will be presented in section 4.In section 5, there will be some Potential directions and outlook. References, where this Essay based on will be delivered in section 6.

2 Backgrounds

In this section,this Essay presents the basic models for HIT, starting with data delivery model, throw radio model, and parallel transmission, also a classification for location inaccuracy, starting with absolute location inconsistency, and ending with relative distance inconsistency.

2.1 Basic Models for Data Gathering

Here is a small definition of each model underlying HIT. The classification of sensor networks depend on the way of data delivery.

2.1.1 Data Delivery Model

the proposed HIT is based on a continuous model of sensor network, which assume that the nodes have always data to transmit, and the data are always sent. It also should be in consider; that the the flow of data packets between the nodes and the observer; this is a routing problem subject to the network protocol. The proposed HIT is using a data aggregation techniques, and reduce the overhead of a broadcast approach for the flow of data from the nodes to the observer. the advantage of this it doesn't depend on a complex network layer protocol for flow chart, addressing, and location management.

2.1.2 Radio Model

In the radio model, The channel is proposed to be the same. this means, the same power to transmit data from node A to B, is the same Power to transmit data from B to A,the variables (d,k) mean the distance between node A and node B, and k-bit message.

$$Transmission = E_{tx}(k, d) = E_{elec} * k + \epsilon_{amp} * k * d^2$$

$$Reception = E_{tx} = E_{elec} * k$$

2.2 Location Inaccuracy and its Variants

A needed definitions of the inconsistency metrics, and classifications of the different types of it, is presented in this section

2.2.1 Absolute Location Inaccuracy

The definition of absolute location inaccuracy for each node i is as follows :

$$\sqrt{(x_i - (x_i + \Delta x_i))^2 + (y_i - (y_i + \Delta y_i))^2} \quad (1)$$

Absolute location inaccuracy is the distance between the true location and the faulty location. There is a possible problem of absolute location inaccuracy which called packet drop, where node i calculate a wrong distance for the destination node d, and send its packet there. In this way Absolute location inaccuracy may conduct relative distance inaccuracy, absolute location inconsistency, and relative location inconsistency.

2.2.2 Relative Distance Inaccuracy

Relative distance inaccuracy for the nodes i and j, and the destination node d is defined as follows:

$$Dist_i = \sqrt{(x_d - x_i)^2 + (y_d - y_i)^2} \quad (2)$$

$$Dist_j = \sqrt{(x_d - x_j)^2 + (y_d - y_j)^2} \quad (3)$$

This means that the relative distance inaccuracy is defined as follows

$$\Delta Dist.\Delta Dist_{error} \leq 0 \tag{4}$$

There is possible problems related to relative distance inaccuracy like; higher packet drop rate, where the node could not find the closer neighbor node to the destination node due to location inaccuracy even though there is in reality closer neighbor node to the destination.

2.2.3 Absolute Location Inconsistency

The definition of absolute location inconsistency is as follows:

$$\sqrt{(x_k^i - x_k^j)^2 + (y_k^i - y_k^j)^2} \tag{5}$$

Absolute location inconsistency represents the difference of the same destination node k perceived by two nodes. A potential problem of absolute location inconsistency is when a sender node (i) sends a packet including the destination location in the header as location (d), however a forwarding node,(j), at the last hope perceives the destination at a different location (d).

2.2.4 Relative Distance Inconsistency

The relative distance inconsistency perceived by nodes i, and j as follows:

$$\Delta Dist_{error}^{i} = Dist_{i-error}^{i} - Dist_{j-error}^{i} \tag{6}$$

And for j defined by:

$$\Delta Dist_{error}^{j} = Dist_{i-error}^{j} - Dist_{j-error}^{j} \tag{7}$$

This means that the relative distance inconsistency is be defined as follows:

$$\Delta Dist_{error}^{i}.\Delta Dist_{error}^{j} \leq 0 \tag{8}$$

There is some problems shining with relative distance inconsistency like: A routing loop case, which means a node (i)thinks that node (j) is the closet node to the destination within the transmission range and send the packet to (j) and (j) send the packet to (i) because in thought it is the closet node to the destination, node (i) send it back to node (j) because it thinks that node (j) is the closet, and this called routing loop. There is also another problem which is called multi-hop routing loop, node (s) sends the packet to node (i), (i) sends the packet to node (j), and (j) send it to node (k), and node (k) send it back to (i) again and so on.

3 HIT – An Approach for Data Gathering

In this section there will be a description of HIT, a simulation of how HIT can save energy, and analysis of the experiment.

3.1 Analysis of Direct and Indirect Transmission

In another cluster-based protocol as LEACH, each node transmit data directly to the cluster head. HIT transmit data indirectly to the cluster head using a multi-hop plane. A small calculation between direct and indirect shows that: the energy cost to send a packet directly is:

$$E_{elec} * k + \epsilon_{amp} * k * (nr)^2 + (E_{elec} * k) \tag{9}$$

In case that the node chooses to send the packet indirectly, this means that the energy cost will be equal to the sum of n transmissions over distance r, and n receptions:

$$n * (E_{elec} * k + \epsilon_{amp} * k * r^2) + n * (E_{elec} * k) \tag{10}$$

Now its easy to see in equation (1) there is more energy saving than in equation (2). HIT using the indirect case assuming data fusion, this means HIT assumes that each intermediary node has its own packet to transmit to the cluster-head, and it can be fused with received packet. The energy coast in this case will be:

$$\sum_{d=1}^{n} [E_{elec} * k + \epsilon_{amp} * k * (dr)^2] + n * (E_{elec} * k) \tag{11}$$

3.1.1 Parallel Transmissions

HIT has a higher level of parallelism than other protocols like LEACH or PEGASIS, which allows multiple,parallel indirect transmissions, because delay is an important thing for a routing protocol, and reducing the delay is throw the parallel transmissions.

3.1.2 Analysis of TDMA against CSMA

TDMA is Time-Division Multiple Access, and CSMA is Carrier-Sensing Multiple Access. HIT using CSMA for its steady-state phase. CSMA lose energy in an energy-constrained sensor network, because of collisions, overhearing, control packet overhead, and idle listening [7]. The vantages of TDMA in sensor network using a continuous data delivery model are no collisions, little overhead, and high energy-efficiency.

3.2 Description of HIT Protocol

Hybrid indirect transmission (HIT) is contains of two state, the first one is the cluster setup, and the second one is the steady state, the steady state is the long one. Each state is broken to several phases, before this essay go on to describe HIT, there is some points have to be assumed :
1- Nodes are distributed randomly.
2- The nodes are able to communicate with CSMA.
3- Packet of data are able to be fused.
4- Nodes are able to estimate distance of other nodes 5- Each sensor node has a unique ID.

3.2.1 Cluster Head Selection

In this phase, cluster-heads will be selected, one or more, each cluster has one cluster-head. This phase can be describe in four selection schemes:
1- Single cluster, rotation: this is the most energy-efficient to have a single cluster-head: then only one node needs to make the large energy expenditure to transmit to the remote base station.
2- Single cluster, rotation, addition selection criteria: in some networks it may preferable to select cluster-head through a certain conditions, like one with a certain level of connectivity to the remote base station, or with a power level that is sufficient to complete an entire steady-state phase of communication.
3- Multiple cluster, random: the use of multiple clusters will reduce the gathering delay of the network, at the expense of the additional energy consumption required for multiple cluster-heads to transmit directly to the remote base station.
4- Multiple clusters, random, additional selection criteria: like the second one, it puts further conditions on the set of candidate node.

3.2.2 Cluster Head Advertisement

in this phase, the selected cluster-heads send their status in the advertisement, the non-cluster-heads have to be on, so they can listen to the advertisement. The non-cluster-heads then compute the distance to the cluster-head.

3.2.3 Cluster Setup

In this phase, some clusters are formed. The message is broad casted at the fixed transmission power. Each non-cluster-head listen for the membership broadcasts and estimate the distance from their position to the sender of the broadcast. After the end of this phase each node (i) knows distance information of any other node (j) and it's cluster-head (H).

3.2.4 Route Setup

In this phase, all nodes broadcast the information at the fixed strength using CSMA MAC, and include their estimate of the distance to their neighbor. This allows all other nodes to know the upstream neighbor of every other node.

3.2.5 Blocking Set Computation

In this phase, each node computes the blocking set for its downstream neighbors. Finally each node merges the knowledge of the other nodes downstream neighbors.

3.2.6 TDMA Schedule Setup

In this stage, each node computes a TDMA schedule that allows close to maximum number of nodes to communicate in parallel while maintaining a collision guarantee. Each one computes the same schedule independently and in parallel.

3.2.7 Data Transmission

This phase is a long steady state phase where each node sense the environment and send the information to their upstream neighbor following the TDMA schedule from the last phase.
A node that receives data from its downstream node will fuse it together with its own, and send it to its own upstream neighbor.the data will reach the cluster-head, which send it to the base station.

3.3 Simulations and Results

In this section of the essay, there will be a small description of the simulation made for HIT, to be sure that this protocol is really energy efficient, and a simulation for location inaccuracy, to watch the result on different aspect on the sensor network.

3.4 Simulation for HIT Protocol

There is some details for the simulation made to test HIT when compared to other protocols, there will be description for the simulation parameters, and the results of the simulation.
1- the simulation parameters:
a- All nodes were bestowed with 20j of initial energy.
b- The data rate of the network is 1 Mbit/s.
c- The number of steady state loops for each round was fixed at 10,000.
d- The coordinates for the base station are (l/2, -200), where l is the area side length.
2-The simulation results are in more than one criteria, the first one is in network longevity, network delay, energy dissipation, and energy delay.

4 A study On Location Inaccuracy In Sensor Network

4.1 Approach: Simulation

We start the first simulation, to know the effect of location inaccuracy on geographic routing. Figures in [2] , shows the simulation parameters. It start with 250m of radio range and Gaussian distribution to generate location inaccuracy. Node degree is defined as the average number of neighbor nodes.
1- Impact on Drop Rate It is clear like in figure 9 in [2], that the error in packet drop becomes higher, as soon as the error in location becomes higher.
2-Impact on Optimal Path In figure 13 in [2] we see the impact of location inaccuracy on optimal path rate. There is higher non-optimal path rate in dense network reaching up to 53 percent.
3-Impact on Looping in sensor network happens because of relative distance inconsistency. Figure 16 in [2] shows the result of the simulation on loop.

4.2 Main Results

(I) For greedy forwarding, packet drops occur mainly because of (a) routing loops caused by relative distance inconsistency and (b) no-route within destination caused by absolute location inconsistency. The packet drop rate is affected by the node degree and the network diameter. In denser network, there was more absolute location inconsistency due to shorter distance between neighbors which resulted in higher packet drop. In sparser network, there was more routing loops due to higher relative distance inconsistency, which resulted in packet drop. For the same network density, larger

network diameter affects geographic routing more negatively in the presence of location in accuracy. This is because as the average path length increases, there is higher probability to incur relative distance inconsistency and absolute location inconsistency. Non-optimal path is caused by relative distance inaccuracy.Relative distance inaccuracy causes wrong greedy neighbor selection. In addition, relative distance inaccuracy causes false local maximum which invokes local maximum resolution. Non-optimal path is affected by node degree and network diameter due to reasons similar to those of packet drops. Non-optimal path is especially undesirable in wireless networks because it consumes valuable power.
(II) Location inaccuracy affects the correctness of graph-based local maximum resolution scheme. In the analysis of the perimeter mode (GG graph) of GPSR figure 18 [2],shows that in the presence of location inaccuracy, the planar graph is highly unlikely to be constructed correctly. Even with small absolute location inaccuracy, It should observed high probability of planar graph collapse.Location inaccuracy degrades performance of perimeter mode in terms of packet drop, optimal path and routing loop rate.

5 Potential Directions

In this Essay, we see that, the two papers searched in two different criteria in the sensor networks, the first one is the matter of location inaccuracy, and the second one searched in the problem of energy saving. Both had the same way of presentation. the same flow of data. both were good presented.I think that both had made a good simulations for the problems presented, with full list of references, but it would be more better if both made some more background in the begin of the papers, both made a very good smoothly presentation,it would also be better if both made some explanation in the matter of data gathering, and in location inaccuracy, but this still our own mind,

6 References

[1] B.J. Clupepper, Lan Dung, M. Moh ' Design and Analysis of HIT for Data Gathering in Wireless Micro Sensor Networks' ACM SIGMOBILE Mobile Computing and Communications Review, Volume 8 , Issue 1 (January 2004).
[2] Y. Kim, Jae-Joon Lee, and Ahmed Helmy 'Modeling and Analyzing the Impact of Location Inconsistencies on Geographic Routing in Wireless Networks' ACM SIGMOBILE Mobile Computing and Communications Review, Volume 8 , Issue 1 (January 2004).
[3]http://robotics.eecs.berkeley.edu/ pister/29Palms0103/

[4]A. Mainwaring, J. Polastre, R. Szewczyk, and D. Culler ‘Wireless Sensor Networks for Habitat Monitoring ‘ ACM International Workshop on Sensor Networks and Applications (WSNA 02), Atlanta, GA, September, 2002.
[5] V. Kottapalli, A. Kiremidjian, J. Lynch, E. Carryer,T. Kenny, K. Law, and Y. Lei, Twotiered wireless sensor network architecture for structural health monitoring,(SPIE 10th) Annual International Symposium on Smart Structures and Materials, San Diego, CA, USA, March 2- 6, 2003.