QoS Compromise in Data Gathering for WSN

Jun-Zhao Sun

Academy of Finland

Department of Electrical and Information Engineering, University of Oulu, 90014, Finland

junzhao.sun@ee.oulu.fi

Abstract—**The problem of adaptive determination of data granularity for QoS-constraint query execution is addressed. An application specifies QoS requirements with a query. Then, each node can choose the optimum data granularity for local data collection and transmission. The proposed algorithm is in a distributed fashion, and executed at each local sensor node.**

Index Terms—**wireless sensor networks, QoS, data gathering**

I. INTRODUCTION

The QoS requirements of a WSN application are usually multifold, and conflict. In particular, it is common that an application wishes the network lifetime to be as longer as possible, and at the same time, the data accuracy the higher the better. Thus, compromise in needed in dealing with the two conflict requirements. On the other hand, data collection algorithm should be distributed and adaptive, by taking into consideration both conditions and condition changes of local sensor node.

Many query processing algorithms have been proposed in the literature to retrieve sensor data from the network in an energy-efficient fashion [1-5]. To develop a data collection algorithm for wireless sensor networks, application-specific requirements need to be deeply understood and carefully considered. In particular, data granularity as the representation of accuracy must be determined by jointly considering the QoS factors and constrains specified with each specific task. Finally, the data collection algorithm should be distributed and adaptive to the local situation of each individual sensor node.

This paper presents an adaptive method for determining data granularity used in data collection. Algorithm is proposed for choosing the best granularity for data gathering, by taking into account both application specified QoS constraints and the local node's conditions. The key novelty of the method lies in the compromise consideration of the requirement on accuracy along with network lifetime. Applications specify the minimum lifetime and accuracy, as well as the relative importance of the two factors in the query. By using this information, the proposed algorithm can calculate the best data granularity, so that the data transmission can be minimized while the data accuracy is kept above certain level. The proposed algorithms are both executed in a distributed fashion at each local sensor node.

II. MODELING

A continuous query contains temporal information, which specify the time to start sampling (t_0) and sample rate r in number of samples per second ($r=\Delta t$, Δt is the sample interval). For a sensor node s_i ($i \in [1, n]$), s_i will generate two types of traffic into the network as follows.

(1) s_i originally generate a local sample stream $\vec{x}_i$, in a rate of r samples per second, with each sample a size of b bits.

(2) s_i received samples collected from all the children nodes in a number of c_i, to be relayed to the sink node. The samples are received from all the one-hop children nodes $s_{i,h}$ ($h = 1, 2, \ldots, i_c$) in the form of merged/compressed packets, $s_{i,0} = s_i$. Node s_i then restores the packets into the raw data, add local data in (1), merge/compress them all into one packet and send it to its parent node.

Let the data reduction coefficient at s_i be a ratio γ_i, γ_i depends on the technique (algorithm, function, etc.) utilized at local node s_i.

Application specifies the QoS requirements on both the minimum lifetime of the network, T_{min} and the minimum accuracy of collected data, A_{min}, as well as the compromise weight between lifetime and accuracy, ω ($\omega<1$). Let $E_{i,t}$ be the residual energy of the node s_i at time t, and $E_{i,\Delta t}$ be the energy dissipation for Δt time long at node s_i. In particular, E_{i,t_0} is the initial or full energy of node s_i. At time t, lifetime of node s_i, $T_{i,t}$ can be obtained by:

$$T_{i,t} = \frac{E_{i,t}\Delta t}{E_{i,\Delta t}} \text{ , and} \tag{1}$$

$$E_{i,\Delta t} = E_{i,u\_sense}b + E_{i,u\_Rx}\sum_{h=1}^{i_c} \gamma_h(L_H + (b + L_{ID\_Seq})\sum_{k=0}^{c_h} \text{Pr}_{k,send})$$

$$+ (E_{i,u\_dp} + E_{i,u\_Tx}\gamma_i)(L_H + (b + L_{ID\_Seq})\sum_{h=0}^{c_i} \text{Pr}_{h,send}), \tag{2}$$

where E_{i,u_sense}, E_{i,u_Rx}, E_{i,u_dp}, and E_{i,u_Tx} are the unit energy needed to sense one bit, receive one bit from a child node, process one bit of data, and send one bit to the parent node over a distance d_i using the radio model. E_{i,u_Tx} consists of two parts: the unit energies spent at the transmitter circuitry E_{i,u_TC} and at the RF transmission amplifier $E_{i,u\_amp} = c_i d_i^{\varepsilon}$. c_i is a constant value related to SNR, radio fading and the poser amplifier, d_i is the distance from node s_i to its parent node, and ε is the path loss exponent ($2<\varepsilon<4$); L_H is the size of a packet header; L_{ID_Seq}: the size of a node ID plus a data sequence number denoting the time stamp; $\text{Pr}_{h,send}$: the probability that a local generated sample at node s_i is.

According to the specification in query Q, the following must be fulfilled:

$$\forall i \in [1, n] \text{ and } \forall t > t_0, T_{i,t} \geq T_{min} - t + t_0. \tag{3}$$

For accuracy, let $A_{i,t}$ be the absolute data accuracy of the node s_i at time t, and is given by:

$$A_{i,t} = |\vec{x}_i(t) - \vec{x}_i'(t)|, \tag{4}$$

where $\vec{x}_i(t)$ and $\vec{x}_i'(t)$ are the sample value and report value at time t respectively. A report value is the data value at the collector center, to represent the observation at a sensor node

and at a particular time. It is not necessarily the sample value and not necessarily be sent from the local node either.

The minimum accuracy of collected data that is acceptable by the application has been specified to be A_{min}, the following must be fulfilled:

$$\forall i \in [1, n] \text{ and } \forall t > t_0, A_{i,t} \leq A_{min}. \qquad (5)$$

III. ALGORITHM

The problems to be solved by the data gathering algorithm can be summarized as follows.

Problem 1: For each sensor node, how to find the best data granularity, Δx_{opt}, so that the QoS trade-off denoted by the weight ω leads to optimum result.

Problem 2: For each sensor node, when and how to update best data granularity, Δx_{opt}, so that the changed condition at the local node can be adapted.

The detailed algorithm contains three stages, setup stage, collection stage, and update stage.

1. At setup stage, to address Problem 1, we create a cost function with the data granularity Δx the parameter to be determined, by:

$$C(\Delta x) = \begin{cases} \omega \dfrac{T_{\min} - t + t_0}{T_t(\Delta x)} + (1-\omega) \dfrac{A_t(\Delta x)}{A_{\min}}, t < T_{\min} + t_0 \\ \omega \ln \dfrac{1}{T_t(\Delta x)} + (1-\omega) \ln A_t(\Delta x), \text{otherwise} \end{cases} \qquad (6)$$

The tentative optimum data granularity $\Delta x'_{opt}$ is given by:

$$\Delta x'_{opt} = \arg \min_{\Delta x} C. \qquad (7)$$

Network lifetime $T_t(\Delta x)$ can be estimated by using equation (1) and (2). $\Pr_{send}$ is given by $\Pr_{send} = 1 - \sum_i p_{i,i}$. Data accuracy $A_t(\Delta x)$, by using equation (4), is given by $A_t(\Delta x) = 1/2 \ \Delta x$.

According to formula (3), the lower boundary of data granularity, Δx_{min} is given by:

$$T_t(\Delta x_{min}) = T_{min} - t + t_0. \qquad (8)$$

According to formula (5), the upper boundary of data granularity, Δx_{max} is given by:

$$A_t(\Delta x_{max}) = 1/2 \ \Delta x_{max} = A_{min}. \qquad (9)$$

Finally, by jointly considering the results of the three conditions, i.e. equation (7), (8) and (9), the optimum data granularity Δx_{opt} can be obtained.

2. At data collection stage, at local node at time $t_j = t_0 + j\Delta t, j = 1, 2, \ldots$, if the distance between $x(t_j)$ and $x(t_{j-1})$ is more than Δx_{opt}, then the report data to be sent as the representative of originally generated sample, $x'(t_j) = a_k = x_k + 1/2\Delta x$. Otherwise, no originally generated data is sent.

3. At update stage, the purpose is to choose a new data granularity Δx, to adapt the changed condition at local node. At time t ($t \geq t_0$) for node s, the algorithm is described by:

i) Starting from t_0, at any time t, node s stores a number of W most recent samples $X_{win} = \{x_m = x(t - m\Delta t)\}_{m=1}^{W}$, and keeps updating the set of window samples by replacing the oldest one with the newest one. The number of W depends on the memory size of the local node.

ii) When update is starts, at time t, estimate the Δx_{max} in equation (9), Δx_{min} in equation (8), as well as the $\Pr_{send}$ by using stored window samples X_{win}, and obtain the new optimum data granularity Δx_{opt}. By taking into the consideration that $\Pr_{send}$ is a monotonic increasing function of

Δx, the estimation algorithm can be easily constructed.

iii) After the update has been completed, node s sends the new sum of the $\Pr_{send}$ of this sub-tree to its parent node.

IV. EXPERIMENTS

A sensor network consisting of 30 sensor nodes and one collector center is considered. Table 1 lists the parameters and the values used in the following experiments. We study the influence of different settings of weight ω on the decision of the optimum data granularity. $\Pr_{send}$ is set to be $1/(2\Delta x + 1)$. Numerical results are summarized in Table 2. The results show the determination of the optimum data granularity to be used.

Table 1. Parameters and values.

parameter	value	parameter	value
Δt	1 min	E_{u_dp}	20 nJ/bit
b	32 bits	E_{u_TC}	50 nJ/bit
γ	1	c	100 pJ/bit/m$_2$
T_{min}	2 months	d	1 m
A_{min}	0.5	ε	2
ω	0.8, 0.5, 0.2	E_{t_0}	8×10^3 mJ
E_{u_sense}	20 nJ/bit	L_H	10 bytes
E_{u_Rx}	50 nJ/bit	L_{ID_seq}	2 bytes

Table 2. Numerical results.

	End node		Level 2 node		Level 1 node	
	Δx	T_{t_0}	Δx	T_{t_0}	Δx	T_{t_0}
$\omega_1 = 0.8$	0.04	86869	0.41	227909	0.87	178049
$\omega_2 = 0.5$	0.04	86869	0	153252	0.17	103911
$\omega_3 = 0.2$	0.04	86869	0	153252	0.04	86869

V. CONCLUSIONS

Having a priori knowledge about the result data helps very much on choosing the best methods. Data accuracy may be represented by relative form and consider deviation of data. These two issues are the problems for future investigation.

REFERENCES

[1] Coman, Jorg Sander, and Mario Nascimento, " An Analysis of Spatio-Temporal Query Processing in Sensor Networks", In Proceedings of the 21st International Conference on Data Engineering Workshops, April 2005, 1190- 1190.

[2] Jonathan Beaver, Mohamed A. Sharaf, Alexandros Labrinidis, and Panos K. Chrysanthis, " Power-Aware In-Network Query Processing for Sensor data," In Proceedings of the 2nd Hellenic Data Management Symposium (HDMS'03), Athens, Greece, September 2003

[3] Samuel Madden, "Query Processing for Streaming Sensor Data," Ph.D. Qualifying Exam Proposal, http://db.lcs.mit.edu/madden/html/madden_quals.pdf

[4] Niki Trigoni, Yong Yao, Alan Demers, Johannes Gehrke, and Rajmohan Rajaraman, "Multi-query Optimization for Sensor Networks," In Proceedings International Conference on Distributed Computing in Sensor Systems, June 2005.

[5] Jin Zhu, Symeon Papavassiliou, and Jie Yang, "Adaptive Localized QoS-Constrained Data Aggregation and Processing in Distributed Sensor Networks," IEEE Transactions on Parallel and Distributed Systems, Vol. 17, No. 9, 923-933, September 2006.