Effect of Data Selection on Data Aggregation in a Wireless Sensor Network

Indika Suranjith Abeyweera Naoki Wakamiya Masayuki Murata

Graduate School of Information Science and Technology, Osaka University
1-5 Yamadaoka, Suita, Osaka 565-0871, Japan
E-mail: a-indika, wakamiya, murata@ist.osaka-u.ac.jp

1 Introduction

In a wireless sensor network (WSN), data aggregation where multiple data are fused into one or more data of smaller size at a node contributes to saving energy and bandwidth [1]. However, the aggregation efficiency and accuracy depend on selection of data to aggregate. In this paper, we compare four different selection methods through simulation experiments.

2 Data Selection Methods

Consider data $d_i(t)$ and $d_j(t)$ obtained by sensor nodes i and j, respectively. Each of the sensor data defines the tolerance for aggregation, denoted as $a_i(t)$ and $a_j(t)$ respectively. We assume that the similarity can be defined between a pair of data. When similarity $s_{i,j}(t)$ between $d_i(t)$ and $d_j(t)$ satisfy both conditions of $s_{i,j}(t) \leq a_i(t)$ and $s_{i,j}(t) \leq a_j(t)$, they can be aggregated.

An aggregation rule can take a form of averaging, maximum, minimum, median, and any other mathematical or statistical operations but in some aggregation rules, the order of selection of data to aggregate at a node affects the aggregation efficiency and accuracy. For example, assume that there are three data $(d_i(t), a_i(t)) = (1, 3)$, $(4, 4)$, and $(5, 5)$ and the similarity is defined as $s_{i,j}(t) = |d_i(t) - d_j(t)|$. As an aggregation rule, consider averaging. The tolerance of an aggregated data takes a smaller tolerance of the original two data. Because of the difference, (1,3) and (5,5) cannot be directly aggregated. However, aggregation of $(1, 3)$ and $(4, 4)$ results in a new data $(2.5, 3)$, which can further be aggregated with $(5, 5)$. Consequently, $(1, 3)$ is aggregated with $(5, 5)$.

In this paper, we consider four selection methods. With *Similarity First*, a node begins with a pair of data with the smallest similarity. *Accuracy First* first tries a pair of data with the lowest tolerance. *Tolerance First* first chooses a pair of data with the largest tolerance. *Random* randomly selects data. A node repeatedly tries aggregation until no pair of data can be aggregated.

3 Simulation Results

We randomly distribute 300 sensor nodes with communication range 50 m in a 500×500 m^2 field. They organize a tree rooted at a server at the corner. All sensor node are assigned random data and random accuracy ranging from 0 to 1. The similarity is defined as the absolute difference and an aggregation rule is averaging. Figure 1 shows the average number of sensor data that nodes at the depth at x-axis sends. Results are obtained from 1000 experiments. Although the difference is small, *Tolerance First* leads to the smallest number of sensor data. A reason that *Similarity First* cannot effectively reduce the number of data is that the tolerance of an aggregated data becomes smaller than that of *Tolerance First*. When we compare the accuracy in terms of the average and standard deviation of errors, which is defined as $\tilde{d}_i(t) = |\delta_i(t) - d_i(t)|/a_i(t)$ and $\delta_i(t)$ is data received at server for node i, as expected, Similarity First has the smallest error. The average error are 0.212, 0.243, 0.307, and 0.266 with *Similarity First*, *Accuracy First*, *Tolerance First*, and *Random*, respectively.

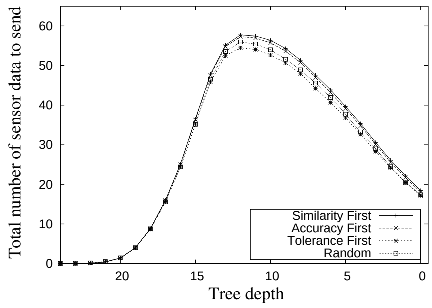

Fig. 1 Comparisons of four data selection methods

4 Conclusion

In this paper, we compared four data selection methods for data aggregation in a WSN. Our future research includes aggregation-aware routing in a densely deployed WSN.

Acknowledgments

This research was supported in part by "Global COE (Centers of Excellence) Program" of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] R. Rajagopalan and P. Varshney, "Data-aggregation techniques in sensor networks: a survey," *IEEE Communications Surveys and Tutorials*, vol. 8, no. 4, pp. 48–63, 2006.