

Mobile Filter: Exploring Migration of Filters for Error-Bounded Data Collection in Sensor Networks

Dan Wang*, Jianliang Xu†, Jiangchuan Liu‡, Feng Wang‡

* Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong,
Email: csdwang@comp.polyu.edu.hk

† Department of Computer Science, Hong Kong Baptist University
Kowloon Tong, Hong Kong, Email: xujl@comp.hkbu.edu.hk

‡ School of Computing Science, Simon Fraser University, Burnaby, British Columbia, Canada,
V5A 1S6, Email: {jcliu, fwa1}@cs.sfu.ca

Abstract—In wireless sensor networks, filters, which suppress data update reports within predefined error bounds, effectively reduce the traffic volume for continuous data collection. All prior filter designs, however, are *stationary* in the sense that each filter is attached to a specific sensor node and remains stationary over its lifetime. In this paper, we propose *mobile filter*, a novel design that explores migration of filters to maximize overall traffic reduction. A mobile filter moves upstream along the data collection path, with its residual size being updated according to the collected data. Intuitively, this migration extracts and relays unused filters, leading to more proactive suppressing of update reports. While extra communications are needed to move filters, we show through probabilistic analysis that the overhead is outrun by the gain from suppressing more data updates.

I. INTRODUCTION

Wireless sensor networks have recently been used for many applications to continuously collect the sensed data in the operational field, so that the field’s properties of interest can be monitored. In this paper, we are interested in continuously gathering data distribution of the sensor field. For example:

Q: Monitor the population of wildlife at difference places every 4 hours for the next 12 months.

Such complex queries, though clearly more difficult to answer, reveal richer information than a simple aggregate such as sum or average. For example, a (consistent) change of the population distribution of the wildlife may be an indication of the change of the surrounding environment [3].

In sensor networks, energy is a severely limited resource, and communication dominates energy consumption. To obtain the distribution information aforementioned, the base station needs to continuously collect data from each sensor node. This is obviously very energy expensive. Fortunately, approximate results are usually acceptable as long as the error is bounded by a certain threshold. Thus, a trade-off between energy consumption and data quality can be explored. Data filtering, by exploiting temporal data correlation, is an effective in-network processing scheme towards this goal. Intuitively, if the difference between the new reading and the previous reading in a sensor node is small, the node should not report the new

reading. Olston *et al.* [4] first generalizes this idea to a filter design for continuous data collection. In their work, a filter is allocated to each sensor node such that the total filter size obeys the user-specified error bound. In each round of data collection, a node will *suppress* its data update report if the difference from the previous report is less than its filter size. There have been a flourish of follow-ups with more intelligent filter allocation strategies (e.g., [2][5]).

All these prior filter designs, however, are *stationary* in the sense that each filter is attached to a specific node and remains stationary during a round of data collection. Thus, unused filters in the current round of data collection might be wasted, limiting the filtering capability. In this paper, we propose *mobile filter*, a novel design that explores migration of filters to reduce network traffic for error-bounded data collection. A mobile filter moves upstream along the data collection path, with its residual size being updated according to the collected data. Intuitively, this migration extracts and relays unused filters, leading to more proactive suppressing of data reports. While extra communications are needed to move filters, we show through probabilistic analysis that the overhead is outrun by the gain from suppressing more data transmissions. The overhead can be further reduced by piggy-backing the filter information in data update reports.

An Example. To illustrate the effect of our mobile filtering scheme, we compare it with a basic stationary filtering scheme in a toy example in Figs. 1 and 2. Consider a sensor network of chain topology (s_4 through s_0). The base station s_0 needs to record the data for each sensor node in each round (or use the previously recorded data if it does not hear from the node). Assume L_1 distance is used for bounding data errors [1],¹ and the total user-allowed filter size (error bound) is 4. The previously reported reading of each sensor is shown in Fig. 1(a). In the current round, each sensor acquires a new reading, as shown in Fig. 1(b). Using the stationary filtering scheme, a filter is allocated to each node and one possible (uniform)

¹ L_1 distance is the sum of the absolute differences between paired values in two datasets. See Section II for a detailed definition.

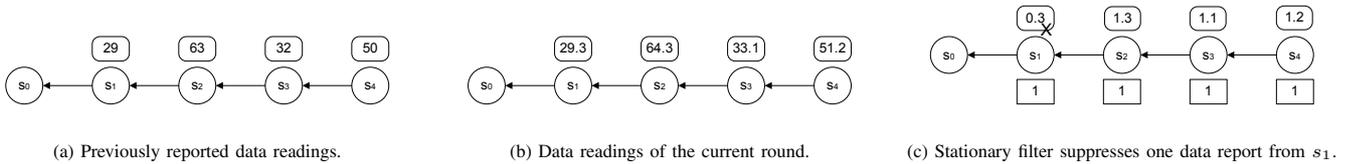


Fig. 1. An example of a stationary filtering scheme. Total user allowed filter size (error bound) is 4. Node s_0 is the base station.

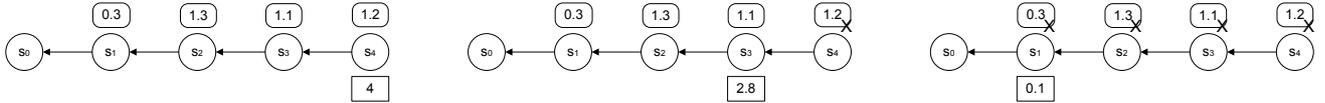


Fig. 2. An example of a mobile filtering scheme. Total user allowed filter size (error bound) is 4. Node s_0 is the base station.

allocation is shown below each sensor in Fig. 1(c). We can see that the stationary filters can suppress only one data update report from s_1 . All other updates need to be reported, and overall it incurs $2+3+4 = 9$ link messages. As a contrast, we now employ the mobile filtering scheme for the same scenario. The entire filter size is assigned to s_4 at the beginning of the current round, as shown in Fig. 2(a). The filter suppresses s_4 's data update report and the residual filter moves upstream as shown in Fig. 2(b), which further suppresses s_3 's update report. In general, the filter suppresses update reports while it moves along the path. Eventually, all four update reports are suppressed, as shown in Fig. 2(c). The total number of link messages incurred is 3 (for the mobile filter transmission).

Intrinsically, one may consider the filter (i.e., the error bound allowed by the user) a valuable resource that can be exploited for conserving energy. In the stationary filtering scheme, each filter has to make an independent decision about data suppressing. The filters have no knowledge of how other filters are used by other sensor nodes. Therefore, the utilization of the filter resource is not optimized; for example, the filters on s_2 through s_4 are wasted in the above example. The mobile filtering scheme, on the other hand, is able to adapt to the current data readings and allocates filters on the fly to optimize the utilization. This intuition will be formalized in our analysis.

II. SYSTEM MODEL AND PROBLEM DEFINITION

In our system, the readings of all sensors need to be periodically collected by the base station to evaluate complex queries, and we call each data collection a *round*. In the first round, all the sensor nodes report their readings. In the subsequent rounds, the sensor nodes report readings that are not suppressed. If the base station does not receive a report from a sensor node, its previously reported reading will be treated as collected data and used for current query evaluation.

To facilitate our presentation, in this paper we employ L_1 distance as the error bound model. Specifically, let the true readings of the sensor nodes be x_1, x_2, \dots, x_N and let the readings collected by the base station be x'_1, x'_2, \dots, x'_N ; the L_1 distance is then $L_1 = \sum_{i=1}^N |x_i - x'_i|$. If the user-specified precision requirement is E , the error-bounded data collection must guarantee $L_1 = \sum_{i=1}^N |x_i - x'_i| \leq E$. L_1 distance is commonly used to measure the distance between complex distributions [1]. It is worthwhile to note that our mobile

filtering scheme is not limited to the L_1 model; and discussions on different error models can be found in [6].

To bound the error of data collection, data filters are installed on sensor nodes in the network. Each filter is associated with a deviation bound (hereafter referred to as *filter size*) and the total filter size should not exceed the bound E . Thus, our problem is to design a mobile filtering scheme that maximizes the overall data suppression while maintaining the user-defined error bound in each round. The mobile filter scheme includes the initial placement of the mobile filters, the mobile filter migration strategy, and the data filtering strategy.

III. MOBILE FILTERING ON A CHAIN TOPOLOGY

For a chain topology, we first show that the mobile filter should initially be placed at the leaf node.

Theorem 1: For a chain topology, the filter should be allocated as a whole to the leaf sensor node in order to minimize the total communication cost.

The proof can be found in [6]. Following this theorem, given a total error bound of E , the filter size initially allocated to the leaf node is E and the filter sizes allocated to all other nodes are zero. By the end of each round of data collection, the leaf node resets the filter size to E and all other nodes reset the filter sizes to zero. It is worth noting that resetting the filter sizes does not incur any communication cost.

We now give a formal cost analysis of the performance for both the stationary and mobile filtering schemes. While this analysis is necessarily simplified, it provides a clearer view of the benefit of mobile filtering. Assume that the data changes follow a standard normal distribution. Let E be the total filter size, and $X_i (i = 1, 2, \dots, N)$ be the random variable for the sensor value change of node s_i . Further assume that the change for each node is i.i.d. and that X_i follows a normal distribution of $N(0, 1)$. Without loss of generality, we consider the case where only the upper bound of the filter is violated.

For stationary filtering, we apply a uniform allocation where each node is assigned a filter size of $\frac{E}{N}$. The probability that the filter is violated at node s_i is $p_i = Pr[X_i > \frac{E}{N}] = 1 - Pr[X_i \leq \frac{E}{N}] = \frac{1}{2}(1 - \text{erf}(\frac{E}{\sqrt{2N}}))$. Define an indicator random variable Y_i such that

$$Y_i = \begin{cases} 1 & \text{if } X_i > \frac{E}{N}; \\ 0 & \text{otherwise.} \end{cases}$$

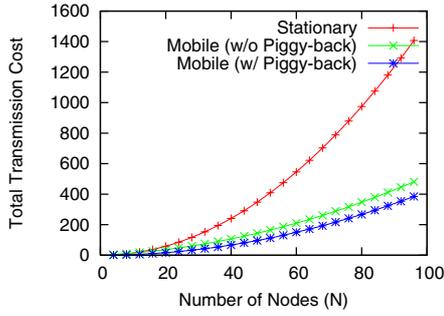


Fig. 3. Expected cost as a function of the number of nodes N . $E = \frac{1}{2}N$.

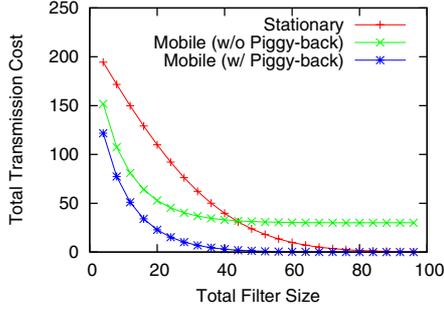


Fig. 4. Expected cost as a function of the total filter size E . $N = 30$.

We have $E[Y_i] = p_i \times 1 + (1 - p_i) \times 0 = p_i = \frac{1}{2}(1 - \text{erf}(\frac{E}{\sqrt{2}N}))$. Given node s_i 's update cost of i , the expected transmission cost is $E[\sum_{i=1}^N Y_i \times i] = E[Y_i] \sum_{i=1}^N i = \frac{N(N+1)}{2} E[Y_i]$.

For mobile filtering, the filter simply migrates upstream and suppresses the data reports as long as the residual filter size is larger than the data change. The probability that the filter is violated at node s_i is $p_i = \Pr[X_i > E - \sum_{j=i+1}^{N+1} X_j] = \Pr[\sum_{j=i}^{N+1} X_j > E]$ (Define $X_{N+1} = 0$). Let $Z_i = \sum_{j=i}^N X_j$. Since X_i 's are i.i.d, Z_i is also a normal distribution of $N(0, N - i + 1)$. Define an indicator random variable Y_i

$$Y_i = \begin{cases} 1 & \text{if } Z_i > E; \\ 0 & \text{otherwise.} \end{cases}$$

We have $E[Y_i] = p_i = \frac{1}{2}(1 - \text{erf}(\frac{E}{\sqrt{2}(N-i+1)}))$. The expected cost of mobile filtering is then $E[\sum_{i=1}^N Y_i \times i] = \sum_{i=1}^N E[Y_i]i$. If the filter migration is not piggy-backed with data reports, there is at most an additional cost of N .

In Figs. 3 and 4, we plot numerical results for the two filtering schemes. For mobile filtering, the results are shown for two versions, namely, with and without piggy-back. Their performance differs by at most N . In Fig. 3, the total error bound is fixed to $\frac{N}{2}$ and a filter of size 0.5 is attached to each node for stationary filtering. Since the data changes follow a standard normal distribution, each filter will suppress the updates with a probability of approximately 40%. We can see that mobile filtering greatly outperforms stationary filtering. We also see that when N increases, the cost of stationary filtering increases much faster than that of mobile filtering, implying that mobile filtering is more scalable.

In Fig. 4, we fix the number of sensors to $N = 30$ and vary the total error bound E from 0 to 100. With piggy-back, mobile filtering performs better than stationary filtering for all cases tested. Even without piggy-back, only after $E = 50$ does mobile filtering perform worse than stationary filtering. Note that when $E = 50$, each node obtains a filter size of $50/30 = 1.67$ for stationary filtering, which implies that the probability of a new reading being suppressed is as high as 90%. In other words, the error bound is extremely large in this case, which may not provide meaningful results and is not desirable for most applications.

The above analysis is simplified in many ways. First, the mobile filter migration strategy is to move the filter all the way to the base station. The mobile filter, however may stop migration, if the residual filter size is small. Second, the data filtering strategy is to simply suppress all the data as long as the mobile filter can. This greedy strategy may not lead to an optimal solution. The intuition is that suppressing data updates with large changes consumes its filter size and may restrict the mobile filter's ability from suppressing more data updates upstream. Third, the underlying network topology is simplified as a chain. Nevertheless, the analysis has shown that substantial gain can be expected by mobile filtering. For a comprehensive study of mobile filtering under general tree-structured sensor networks, the reader is referred to [6].

IV. CONCLUSION

In this paper, we have proposed a novel mobile filtering scheme for error-bounded non-aggregate data collection in sensor networks. By exploring the migration of filters, a mobile filter extracts and relays unused filters in the network to suppress as many data update reports as possible. An analytical study has been performed to quantify the performance benefit of mobile filtering against the conventional stationary filtering; where substantial gain is shown to be expected.

ACKNOWLEDGMENT

J. Xu's work was supported by the Research Grants Council of Hong Kong under Project Nos. HKBU211505 and HKBU211307. J. Liu's work was supported by a Canada NSERC Discovery Grant and an NSERC Strategic Grant.

REFERENCES

- [1] T. Batu, L. Fortnow, R. Rubinfeld, W. Smith, and P. White, "Testing That Distributions Are Close", in *Proc. IEEE FOCS'00*, Redondo Beach, CA, Nov. 2000.
- [2] A. Deligiannakis, Y. Kotidis, and N. Roussopoulos, "Hierarchical In-Network Data Aggregation with Quality Guarantees", in *Proc. EDBT'04*, Heraklion, Greece, Mar. 2004.
- [3] T. He, S. Ben-David, and L. Tong, "Nonparametric Change Detection and Estimation in Large Scale Sensor Networks", *IEEE Trans. Signal Processing*, vol. 54, no. 4, pp. 1204-1217, Apr. 2006.
- [4] C. Olston, J. Jiang, and J. Widom, "Adaptive Filters for Continuous Queries over Distributed Data Streams", in *Proc. ACM SIGMOD'03*, San Diego, CA, June 2003.
- [5] X. Tang and J. Xu, "Extending Network Lifetime for Precision-Constrained Data Aggregation in Wireless Sensor Networks", in *Proc. IEEE INFOCOM'06*, Apr. 2006.
- [6] D. Wang, J. Xu, J. Liu, and F. Wang, "Mobile Filter: Exploring Migration of Filters for Error-Bounded Data Collection in Sensor Networks", Technical Report, Simon Fraser University, July 2007.