

# Information Quality Aware Routing in Event-Driven Sensor Networks

Hwee-Xian Tan\*, Mun-Choon Chan\*, Wendong Xiao†, Peng-Yong Kong† and Chen-Khong Tham†

\*School of Computing, National University of Singapore (NUS)

†Institute for Infocomm Research (I<sup>2</sup>R), Singapore

{hweexian, chanmc}@comp.nus.edu.sg, {wxiao, kongpy, cktham}@i2r.a-star.edu.sg

**Abstract**—Upon the occurrence of a phenomenon of interest in a wireless sensor network, multiple sensors may be activated, leading to data implosion and redundancy. Data aggregation and/or fusion techniques exploit spatio-temporal correlation among sensory data to reduce traffic load and mitigate congestion. However, this is often at the expense of loss in Information Quality (IQ) of data that is collected at the fusion center.

In this work, we address the problem of finding the least-cost routing tree that satisfies a given IQ constraint. We note that the optimal least-cost routing solution is a variation of the classical NP-hard Steiner tree problem in graphs, which incurs high overheads as it requires knowledge of the entire network topology and individual IQ contributions of each activated sensor node. We tackle these issues by proposing: (i) a topology-aware histogram-based aggregation structure that encapsulates the cost of including the IQ contribution of each activated node in a compact and efficient way; and (ii) a greedy heuristic to approximate and prune a least-cost aggregation routing path. We show that the performance of our IQ-aware routing protocol is: (i) bounded by a distance-based aggregation tree that collects data from all the activated nodes; and (ii) comparable to another IQ-aware routing protocol that uses an exhaustive brute-force search to approximate and prune the least-cost aggregation tree.

## I. INTRODUCTION

Wireless Sensor Networks (WSNs) comprise of large numbers of sensor nodes that are deployed in the terrain to sense physical phenomena of interests (PoIs) [1] and transmit the sensed data to one or more centralized sinks for real-time processing and decision-making [2]. These networks can be used for a multitude of applications, such as healthcare monitoring and tactical surveillance. However, the successful deployment and maintenance of sensor networks is challenged by several factors, such as temporal link connectivity as well as limited bandwidth, energy and processing capabilities.

We focus on a particular class of sensor networks that is deployed specifically for the detection of PoIs. Such networks have convergecast traffic characteristics [3] and are considered to be **event-driven** as sensory data is typically generated only when the PoI is detected. Upon the occurrence of a PoI (such as a fire hazard or an elderly person falling in a monitored home environment), multiple sensors may be activated concurrently. This may lead to severe data implosion and redundancy [4], especially in extremely dense networks.

To mitigate the effects of congestion and reduce medium access contention, data aggregation and/or fusion techniques are often used to suppress data from multiple sensor sources.

These schemes exploit spatio-temporal correlation among sensory data to reduce traffic load and energy consumption. However, this often comes at the expense of loss in information quality (IQ) of data that is collected at the fusion center. In event-driven sensor networks, this is equivalent to a loss in (event) *detection accuracy*, which results in overall lower system and network reliability. As such, there exists an obvious trade-off between information quality and energy efficiency in data aggregation and/or fusion schemes [5].

We address the problem of finding a least-cost routing tree that satisfies a given IQ constraint, noting that the optimal least-cost routing solution is a variation of the classical NP-hard Steiner tree problem in graphs [6]. In the context of event-driven sensor networks, the original Steiner tree problem is to find a Steiner Minimum Tree that spans the fusion center and entire set of activated sensor nodes (denoted as  $V_a$ ) that detect the PoI. However, it is unnecessary and expensive for all the sensors in  $V_a$  to transmit their data to the fusion center.

Ideally, an IQ-aware event-driven routing scheme only needs to aggregate sufficient sensory data from a subset of activated nodes  $V_\tau \subseteq V_a$  to meet the IQ constraint. We refer to this as a *subset- $\tau$*  Steiner Tree (SST) problem. Clearly, the latter is a harder problem than the Steiner tree, as the: (i) set of activated nodes  $V_a$  and IQ contribution of each activated node are not known *a priori* until the PoI occurs; and (ii) subset of activated nodes  $V_\tau$  whose aggregated IQ meets the given constraint is not unique. Knowledge of the entire network topology and individual IQ contributions of each activated sensor node are quintessential to find the least-cost SST that satisfies the required IQ. Unfortunately, this incurs extensive computational, storage as well as communication overheads, and is not a feasible approach in resource-constrained WSNs.

We tackle the fundamental issues in constructing an optimal least-cost SST in an IQ-aware event-driven sensor network by proposing: (i) a topology-aware histogram-based aggregation structure that encapsulates the cost of including the IQ contribution of each activated node, in a compact and efficient way; and (ii) a greedy heuristic to approximate and prune a least-cost aggregation routing path. The proposed **IQ-Aware Routing (IQAR)** protocol constructs an initial distance-based aggregation tree that spans *all* the sensors in the WSN. When a PoI occurs, activated sensors forward their data to the fusion center using the underlying pre-built distance-based aggregation tree. At each hop along the routing path, the IQ

contributions of each activated sensor and its downstream (forwarding) nodes are discretized and incorporated together, to form a topology-aware histogram-based aggregation structure.

When data packets (with incorporated histograms) reach the fusion center, it utilizes a greedy heuristic to prune the original aggregation tree such that: (i) aggregated IQ of the resulting pruned tree satisfies a given IQ constraint; and (ii) total cost of collecting data from activated nodes in the pruned tree is minimized. The pruning process is recursively executed at each forwarding node along the initial aggregation tree.

Activated sensor nodes that are not part of the pruned routing tree suppress their data for a time epoch. When the time epoch expires, activated nodes whose data have been suppressed resume the forwarding of their data to the fusion center. This allows the routing protocol to be adaptive towards dynamic changes in the network and phenomena of interest. We show that the performance of IQAR is upper bounded by a distance-based aggregation tree that collects data from all the activated nodes, and comparable to another IQ-aware routing protocol that uses an exhaustive brute-force search to approximate and prune the least-cost aggregation tree.

The rest of this paper is organized as follows: Section II discusses background and related work. The system model is described in Section III. Details of the topology-aware histogram-based aggregation structure and IQAR are presented in Sections IV and V respectively. We evaluate the performance of IQAR in Section VI and conclude in Section VII.

## II. BACKGROUND AND RELATED WORK

### A. Aggregation and/or Fusion based Routing

Data aggregation and/or fusion based-routing has gained significant popularity in WSNs due to the obvious benefits brought about by in-network processing [4] [7]. These schemes exploit presence of spatio-temporal correlation [8] among sensory data to reduce traffic load, which subsequently alleviates medium contention as well as reduces transmission costs and end-to-end delays. Generally, such techniques can be classified into two main categories, viz. structure-less and structured.

In structure-less techniques [9] [10], data aggregation occurs *opportunistically* only when data flows happen to meet at the same time at the same intermediate forwarding node. There is no deliberate attempt to delay any transmission or re-route packets such that these encounters take place. Consequently, structure-less approaches incur shorter delays when the network is lightly loaded, as data is forwarded towards the fusion center using an underlying shortest-path or least-cost routing algorithm. However, such approaches do not scale well with large networks or high traffic volumes as aggregation opportunities are not maximized.

In structured techniques [11] [12] [13] [14], routing paths are computed and maintained to allow efficient data aggregation. The routing path is influenced primarily by the amount of data reduction that can be achieved by data compression before it is forwarded to the fusion center. Such techniques incur relatively higher overheads to maintain the network structure, and are associated with a delay factor, as intermediate

forwarding nodes have to wait for their upstream nodes to send data to them, before aggregating these data packets and forwarding them to the fusion center. Nevertheless, structured techniques can achieve significant energy savings, as data is maximally aggregated along the forwarding paths. They are well-suited for WSNs with slow-varying traffic characteristics, such as periodic monitoring. Many structured schemes adopt a clustering approach [15] [16] [17], whereby sensory data is first transmitted to a cluster head for aggregation before being forwarded to the fusion center.

These existing aggregation-based approaches require all sensory data to be collected at the fusion center, which results in high data redundancy and energy costs. As not all the data has to be transmitted to the fusion center for reliable PoI detection in IQAR, it can achieve significant energy savings over generic aggregation-based routing schemes.

### B. Information-Quality (IQ) Aware Routing

Unlike aggregation-based routing schemes, IQ-aware routing schemes consider the information content of data during data aggregation and forwarding. Information-directed approaches such as IDSQ and CADR [18] [19] [20] handle data querying and routing using energy-efficient techniques, while minimizing delay and bandwidth consumption. At each step along the routing path, the neighboring node with the highest predicted information gain is selected to be the next-hop, and the fused data is transmitted to the fusion center as soon as it satisfies a given IQ threshold. These existing schemes differ from IQAR in that they are query-based, targeted at tracking applications, and data flows along a single path which is initiated from the fusion center. However, IQAR adopts a tree-based approach and is targeted at event detection in WSNs.

The emphasis of existing literature on event-detection in sensor networks [1] [21] [22] [23] is on designing energy-efficient hypothesis testing models to detect the presence of the PoI. Most of these schemes do not consider routing, and are based on a centralized one-hop sensor network topology. In contrast, IQAR addresses both event detection and multi-hop routing simultaneously.

## III. SYSTEM MODEL

In an event-detection WSN, the *information quality* of concern is related to the detection accuracy of the system. In this section, we detail how IQ is mapped onto the targeted detection and false alarm probabilities  $P_d$  and  $P_f$  using sequential detection [24] [25]. We also describe the Likelihood Ratio Test (LRT) [26], which has been shown to be the optimal detection scheme that maximizes detection probability.

The network is modeled as a graph  $G = \{V, E\}$ , where  $V = \{v_0, v_1, v_2, \dots, v_n\}$  denotes the set of  $n$  sensor nodes and fusion center  $v_0$ , and  $E$  denotes the set of edges (or links) between any two nodes. An edge  $e_{ij} \in E$  represents the existence of a communication link between two arbitrary sensors  $v_i$  and  $v_j$ .

We let hypothesis  $H_1$  denote the presence of a PoI in the sensor network;  $H_0$  denotes the corresponding absence of the PoI. The probabilities  $P(H_1) = p$  and  $P(H_0) = 1 - p$ , where

$0 < p < 1$ , are known *a priori*. Each node *independently* senses and collects data about the environment periodically. When conditioned upon the hypothesis  $H_i$ ,  $i \in \{0, 1\}$ , sensor observations are assumed to be independently and identically distributed (i.i.d.) at each sensor as well as across sensors [25].

#### A. Event Detection at Sensor

The independent signal  $y_i$  observed by a node  $v_i$  is:

$$y_i = \begin{cases} w_i & \text{if } H_0 \text{ (PoI is absent);} \\ f(r_i) + w_i & \text{if } H_1 \text{ (PoI is present),} \end{cases} \quad (1)$$

where  $w_i \sim \mathcal{N}(0, \sigma_w^2)$  is the noise seen by  $v_i$  that follows a normal distribution with mean 0 and standard deviation  $\sigma_w$ ;  $r_i$  is the the distance between  $v_i$  and the PoI; and  $f$  is a function that monotonically decreases with increasing  $r_i$ . For each sampled signal  $y_i$ ,  $v_i$  makes a **per-sample binary decision**  $b_i \in \{0, 1\}$  such that:

$$b_i = \begin{cases} 0 & \text{if } y_i < \mathbb{T}_i; \\ 1 & \text{otherwise,} \end{cases} \quad (2)$$

where  $\mathbb{T}_i$  is the per-sample threshold of  $v_i$ .

The per-sample probability of false alarm  $p_0^i$  by  $v_i$  is independent of its location, and given by [1]:

$$p_0^i = P(b_i = 1|H_0) = Q\left(\frac{\mathbb{T}_i}{\sigma_w}\right), \quad (3)$$

where  $Q(x)$  is the Gaussian Q-function of a standard normal distribution. The corresponding per-sample probability of detection  $p_1^i$  (where  $p_1^i > p_0^i$ ) at  $v_i$  is dependent on the distance  $r_i$  between  $v_i$  and the PoI, and given by:

$$p_1^i = P(b_i = 1|H_1) = Q\left(\frac{\mathbb{T}_i - f(r_i)}{\sigma_w}\right). \quad (4)$$

Although each node senses the environment at periodic intervals, it is infeasible for each sample to be transmitted to the fusion center  $v_0$ , due to limited network bandwidth and energy constraints of the sensors. Consequently, a data packet is generated by  $v_i$  to be transmitted to  $v_0$  only if  $v_i$  detects the presence of a PoI and becomes **activated** (when  $b_i = 1$ ).

#### B. Event Detection at Fusion Center

The role of the fusion center  $v_0$  is to detect the presence of the PoI by making a **global binary decision**  $\hat{H} = \{H_0, H_1\}$ , based on the data that it has received from the set of activated nodes  $V_a$ . Let  $B = \{b_1, b_2, \dots, b_{|V_a|}\}$  be the set of per-sample binary decisions that  $v_0$  receives from each  $v_a \in V_a$  in a time epoch. The optimal fusion rule for  $v_0$  using data from *all* the activated nodes is the Likelihood Ratio Test (LRT) [25] [26]:

$$\Lambda(B) = \frac{P(b_1, b_2, \dots, b_{|V_a|}|H_1)}{P(b_1, b_2, \dots, b_{|V_a|}|H_0)} \stackrel{H_1}{\geq} \frac{1-p}{p}. \quad (5)$$

Hence,  $v_0$  makes the decision that the PoI is present ( $\hat{H} = H_1$ ) if  $\Lambda(B) \geq \frac{1-p}{p}$ . Notice that for small values of the *a priori* probability  $p$ , the likelihood ratio  $\Lambda(B)$  required for the PoI to be detected is much larger than for bigger values of  $p$ . As a numerical example, consider  $p = 0.01$ ; then, the PoI can be detected only when  $\Lambda(B) \geq \frac{1-0.01}{0.01} = 99$ . In contrast, if  $p = 0.1$ , the PoI can be detected when  $\Lambda(B) \geq \frac{1-0.1}{0.1} = 9$ .

#### C. Sequential Detection

To reduce the amount of data that is collected for  $v_0$  to make an accurate global binary decision  $\hat{H}$ , we adopt the sequential detection model which is based on the Sequential Probability Ratio Test (SPRT) proposed by A. Wald [24]. In SPRT, the amount of data required is a random variable dependent on the prior data that has been obtained thus far.

Let  $X_a = \{x_1, x_2, \dots, x_{|V_a|}\}$  be the *sequence* whereby data is collected from each activated node  $v_a \in V_a$ . Using sequential detection, data acquisition can terminate at the earliest subsequence of fused local data  $X_\tau = \{x_1, x_2, \dots, x_\tau\} \subseteq X_a$  when the decision  $\hat{H} = \{H_0, H_1\}$  can be made., thus minimizing the cost of data acquisition and PoI detection.

We consider an arbitrary node  $v_i$  with hopcount  $h$  along the routing path. The neighboring node  $v_j$  with hopcount  $h + 1$  and which uses  $v_i$  to forward packets to the fusion center is considered an **upstream** node of  $v_i$ ; the set of upstream nodes of  $v_i$  is denoted as  $V_i^u$ . In the same manner,  $v_i$  is known as the **downstream** node of  $v_j$ . In Figure 1,  $v_1$ ,  $v_2$  and  $v_3$  are the upstream nodes of  $v_0$  such that  $V_0^u = \{v_1, v_2, v_3\}$ .

Since observations across sensor nodes are i.i.d., the **cumulative log-likelihood ratio**  $S_0$  at the fusion center  $v_0$  is:

$$S_0 = \log \Lambda(B) = \log \prod_{i=1}^{|V_a|} \Lambda(b_i). \quad (6)$$

The corresponding cumulative log-likelihood ratio  $S_i$  at  $v_i$  comprises of its log-likelihood ratio and the cumulative log-likelihood ratios of each upstream node  $v_j \in V_i^u$ , such that:

$$S_i = \log \Lambda(b_i) + \sum_{v_j \in V_i^u} S_j. \quad (7)$$

The **stopping rule**  $\gamma_i = \{0, 1\}$  is computed after each incorporated data from  $v_i$ , and is dependent on the targeted detection and false alarm probabilities  $P_d$  and  $P_f$ . It determines if the current sequence of incorporated data along the routing path is sufficient for the global decision  $\hat{H}$  to be made at  $v_0$ , and is given by Wald's Equation [24]:

$$\gamma_i = \begin{cases} 0 & \text{if } \mathbb{A} < S_i < \mathbb{B}; \\ 1 & \text{otherwise,} \end{cases} \quad (8)$$

where  $\mathbb{A} = \log\left(\frac{1-P_d}{1-P_f}\right)$ ; and  $\mathbb{B} = \log\left(\frac{P_d}{P_f}\right)$ . This stopping rule is considered to be optimal in sequential detection, as it results in the least possible amount of data required for decision making.

If  $\gamma_i = 0$ , the current sequence of data collected is insufficient for a global decision  $\hat{H}$  to be made and more data samples have to be acquired. However, when  $\gamma_i = 1$ , the decision  $\hat{H}$  can be made based on the current sequence of incorporated data, according to:

$$\hat{H} = \begin{cases} H_0 & \text{if } S_i \leq \mathbb{A}; \\ H_1 & \text{otherwise } (S_i \geq \mathbb{B}). \end{cases} \quad (9)$$

Hence, additional data need not be collected from other sensor nodes and data acquisition can be terminated to minimize the overall cost while satisfying the  $P_d$  and  $P_f$  constraints.

Based on this sequential detection model, the minimum cumulative log-likelihood ratio required for  $v_0$  to detect the PoI with sufficient accuracy is  $S_0 \geq \mathbb{B}$ . The IQ threshold  $I_T$  can then be directly mapped to  $\mathbb{B}$  such that  $I_T = \mathbb{B}$ . The corresponding IQ provided by each node  $v_i$  is hence given by  $q_i = \log \Lambda(b_i)$ . As an illustration, the IQ provided by  $v_7$  and  $v_8$  in Figure 1 are  $q_7 = \log \Lambda(b_7) = 0.4$  and  $q_8 = \log \Lambda(b_8) = 0.6$  respectively. The corresponding cumulative log-likelihood ratio at  $v_3$  is  $S_3 = \log \Lambda(b_3) + q_7 + q_8 = 1.2 + 0.4 + 0.6 = 2.2$ .

#### D. Cost Model

Each link between a pair of nodes  $v_i$  and  $v_j$  is associated with some cost  $C_{ij}$ . In the absence of power control, the per-hop link cost  $C_{ij}$  is independent of the distance between  $v_i$  and  $v_j$ , and can be computed as a function of: (i) processing energy required to process and perform data aggregation on a data packet; (ii) transmission energy expended by  $v_i$ ; and (iii) reception energy expended by neighbors of the transmitting node  $v_i$  upon reception of the packet in a wireless medium. One implicit assumption in our cost model is that each data packet is of the same size, regardless of the amount of data that has been fused together from different nodes. We explain our assumptions behind this model in Section IV.

#### E. Problem Formulation

Given the network  $G = \{V, E\}$ , set of activated nodes  $V_a$ , IQ contribution  $q_a$  of each activated node  $v_a \in V_a$  and IQ threshold  $I_T$ , our objective is to design an IQ-aware routing protocol that detects the PoI with an IQ of at least  $I_T$  using minimal cost. Formally, we want to find a *subset- $\tau$*  Steiner Tree, which is a Steiner Minimum Tree  $G_\tau = \{V_\tau, E_\tau\} \subseteq G$  that spans the fusion center  $v_0$  and all nodes in  $V_\tau \subseteq V_a$ , such that: (i) aggregated IQ collected from  $V_\tau$  exceeds  $I_T$ ; and (ii) total cost of the aggregation tree is minimum among all possible Steiner trees that meet the IQ constraint, i.e.,

$$\min \sum_{e_{ij} \in E_\tau} C_{ij}; \quad (10)$$

subject to:

$$\sum_{v_i \in V_\tau} q_i \geq I_T. \quad (11)$$

**Lemma** The IQ-aware routing problem which finds the least-cost routing path for a given subset  $V_\tau \subseteq V_a$  of the activated nodes that satisfies the IQ constraint  $I_T$  is NP-hard. The proof for NP-hardness can be found in [27].

In addition to the computational complexity on the fusion center, the acquisition of knowledge on the global network topology and individual IQ contribution of each node incurs high overheads in both communication and computation. Thus, such an approach is impractical in the context of wireless sensor networks which are inherently resource-limited. In the following sections, we describe a compact and efficient way of representing the network topology and IQ contributions of each node, and then propose a heuristic for solving the NP-hard least-cost IQ-aware routing problem.

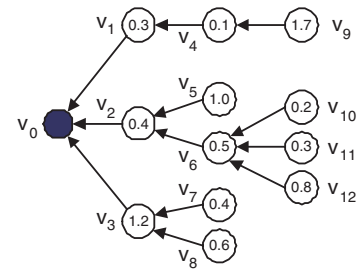


Fig. 1. Fusion center  $v_0$  with three upstream nodes  $v_1$ ,  $v_2$  and  $v_3$ .

## IV. TOPOLOGY-AWARE HISTOGRAM-BASED AGGREGATION

We first illustrate our approach using the network topology in Figure 1, where the fusion center  $v_0$  has three (direct) upstream nodes such that  $V_0^u = \{v_1, v_2, v_3\}$ . All the nodes are assumed to be activated; the number associated with each node  $v_i$  represents its IQ contribution  $q_i = \log \Lambda(b_i)$ . The cost of data transmission across each link is assumed to be of unit cost, i.e.  $C_{ij} = 1 \forall e_{ij} \in E$ , as the packet size remains constant across each link in our IQ aggregation scheme.

#### A. Motivation

Using a direct (or brute force) approach, all activated sensors forward their data to  $v_0$ , which will then determine if the PoI is present. Such an approach is inefficient as data acquisition from all the activated nodes without any aggregation incurs high communication costs and overloads  $v_0$ . Furthermore, even if all these data can be obtained by  $v_0$ , the optimal least-cost routing tree cannot be found efficiently.

Given a global view of the topology and knowledge of IQ contributions of each sensor node in Figure 1, it is easy to see that the minimum cost aggregation tree comprises of only  $\{v_3\}$  if the required IQ  $I_T = 1.0$ . If  $I_T = 2.0$ , the minimum cost aggregation tree can be  $\{v_3, v_7, v_8\}$ ,  $\{v_2, v_3, v_8\}$ ,  $\{v_2, v_3, v_5\}$ ,  $\{v_1, v_3, v_8\}$  or  $\{v_1, v_4, v_9\}$ . Similarly, if  $I_T = 4.5$ , then the minimum cost aggregation tree can be  $\{v_2, v_3, v_5, v_6, v_8, v_{12}\}$  or  $\{v_1, v_2, v_3, v_5, v_5, v_9\}$ . However, it is desirable to utilize an efficient and distributed way of computing a minimum cost aggregation tree that meets the IQ constraint  $I_T$ .

In the proposed approach, upon the detection of a PoI, activated nodes that are further from the fusion center  $v_0$  initiate the transmission of *hints* towards it. These hints are *aggregated* by downstream nodes and further propagated towards  $v_0$ . The aggregated hint conveyed by an arbitrary node  $v_i$  is designed to present useful information about how IQ is distributed in the subtree rooted at  $v_i$ , without providing the detailed IQ values of each node and the actual topology of the subtree. For the purpose of scalability, these *hints* are of constant size. The objective is to design a scheme that generates sufficiently useful hints to  $v_0$ , so that a minimum cost tree can be constructed in the reverse direction, in a distributed fashion. Our approach is based on the concept of a topology-aware histogram-based (hints) aggregation.

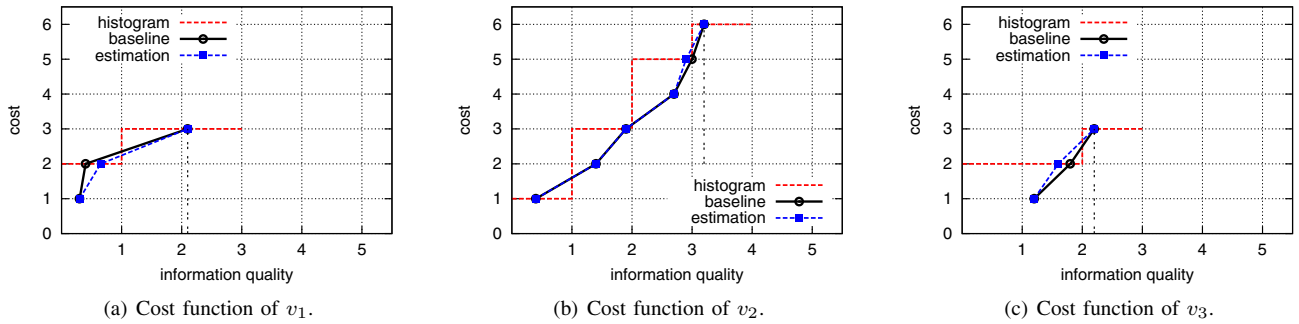


Fig. 2. Cost functions of subtrees rooted at  $v_1$ ,  $v_2$  and  $v_3$  in Figure 1, with IQ threshold  $I_T = 5$  and number of discretization levels  $\phi = 5$ .

### B. Histogram-Based Representation

In our baseline histogram representation, the  $y$ -axis is the cost (total number of transmissions along routing path) and the  $x$ -axis is the IQ that can be accumulated with the given cost. Depending on the routing path that is taken, different IQ values may be accumulated for a given cost. In this baseline representation, the accumulated IQ is the largest possible for a particular cost. Note that the computation of this *maximum* IQ for a given cost is similar to the original routing problem (which finds the minimum cost of obtaining a particular IQ value), and cannot be computed efficiently for a large network.

An exhaustive search can be used to compute the maximum IQ values for each cost in the small network in Figure 1. The solid lines (labeled as *baseline*) in Figure 2 illustrate baseline curves for cost vs maximum IQ, for each upstream node of  $v_0$ . As highlighted previously, such a baseline representation embeds detailed network topology and IQ distribution information, at the cost of excessive computational and communication overheads. To reduce information content and overheads of the representations, quantization levels are introduced.

1) *Quantization*: Let  $\phi$  be the number of quantization levels in the baseline representation. This results in  $\phi$  histograms, each of width  $\frac{I_T}{\phi}$ , where  $I_T$  is the IQ threshold required for PoI detection. The range of IQ values represented by the  $i^{th}$  block ( $i = 1, \dots, \phi$ ) is  $(i - 1) \cdot \frac{I_T}{\phi}$  to  $i \cdot \frac{I_T}{\phi}$  (inclusive).

The number of data points that fall within the range of IQ values governed by the  $i^{th}$  block of the histogram is denoted as  $n_i$ . The corresponding cost  $c_i$  for the  $i^{th}$  block is  $c_i = \sum_{j=1}^{j=i} n_j$ . It can be deduced that with a cost of  $c_i + 1$ , the maximum IQ obtainable is at least  $i \cdot \frac{I_T}{\phi}$ . The dotted lines (labeled as *histogram*) in Figure 2 illustrate the relationship between the baseline and histogram representations from the perspectives of nodes  $v_1$ ,  $v_2$  and  $v_3$  in Figure 1.

2) *IQ Estimation*: After quantization, the (original) baseline representation of the IQ that is known to an arbitrary node  $v_e$  is substituted by a (compact) histogram representation, to be transmitted to its downstream node  $v_d$ . We now discuss how  $v_d$  estimates the IQ of the subtree rooted at  $v_e$ , based on the histogram representation it receives from  $v_e$ .

Each of the  $n_i$  points in the  $i^{th}$  block of the histogram is associated with cost  $c_i^k = c_i - n_i + k$ , where  $1 \leq k \leq n_i$ . Assuming that each of these points is uniformly distributed

TABLE I  
ESTIMATED AND ACTUAL IQ GAIN PER INCREMENTAL COST  $c$ .

| Cost $c$ | $v_1$    |                | $v_2$    |                | $v_3$    |                |
|----------|----------|----------------|----------|----------------|----------|----------------|
|          | $q_1(c)$ | $\hat{q}_1(c)$ | $q_2(c)$ | $\hat{q}_2(c)$ | $q_3(c)$ | $\hat{q}_3(c)$ |
| 1        | 0.3      | 0.3            | 0.4      | 0.4            | 1.2      | 1.2            |
| 2        | 0.4      | 0.65           | 1.4      | 1.4            | 1.8      | 1.6            |
| 3        | 2.1      | 2.1            | 1.9      | 1.9            | 2.2      | 2.2            |
| 4        | -        | -              | 2.7      | 2.7            | -        | -              |
| 5        | -        | -              | 3        | 2.9            | -        | -              |
| 6        | -        | -              | 3.2      | 3.2            | -        | -              |

within the block range, the corresponding estimated IQ  $\hat{q}_e(c_i^k)$  obtained with a cost of  $c_i^k$  (from the perspective of  $v_d$ ) is:

$$\hat{q}_e(c_i^k) = \frac{I_T}{\phi} \left[ (i - 1) + \frac{k}{n_i + 1} \right]. \quad (12)$$

Given the IQ  $q_e$  obtainable from the upstream node  $v_e$  and the maximum IQ  $q_e^M$  that can be obtained using the subtree rooted at  $v_e$ , the estimated IQ in Equation (12) can be further tightened by using  $q_e$  and  $q_e^M$  as the lower and upper bounds of the histogram. Note that  $c_i$  and  $\hat{q}_e(c_i^k)$  are undefined  $\forall i = \lceil q_e^M \cdot \frac{\phi}{I_T} \rceil$  as these are not regions of interest.

The dotted lines with points (labeled *estimation*) in Figure 2 plot the values of estimated IQ  $\hat{q}_e(c)$  for each additional edge from the subtree rooted at  $v_e$ , from the perspective of its downstream node  $v_d$ . The *estimation* plot can be interpreted in this way: For a cost of  $c$ , it is likely that at least an IQ of  $\hat{q}_e(c)$  can be obtained. Table I lists the baseline  $q_e(c)$  and estimated  $\hat{q}_e(c)$  values for each of the subtrees rooted at the upstream nodes of  $v_0$  in Figure 1. Note that  $q_e(c)$  can be larger or smaller than  $\hat{q}_e(c)$ .

Finally, we discuss the selection of the parameter  $\phi$ . In our IQ estimations, each of the  $n_i$  points in the  $i^{th}$  block is assumed to be uniformly distributed; the corresponding IQ estimations for each of the points is also uniformly distributed. If this assumption is valid, then the value of  $\phi$  can be small. However, if IQ values vary significantly among nodes, then a larger  $\phi$  value is required to increase the accuracy of the piecewise linear approximation. We can now describe our proposed IQ-aware routing protocol in the following section.



**Algorithm 1** IQ Approximation Algorithm in Phase 2

1: **Input:**  $I_T, \hat{q}_j(c_j) \forall v_j \in V_i^u, 1 \leq c_j \leq C_j^M$   
 2: **Variable:**  $I_{total} = 0, \bar{c}_j = 0 \forall v_j$   
 3: **Output:**  $\hat{q}_i(c_i) 1 \leq c_i \leq C_i^M$   
 4: **while**  $\sum \bar{c}_j < \sum C_j^M$  **OR**  $I_{total} < I_T$  **do**  
 5:    $k \leftarrow \operatorname{argmax}_j [\hat{q}_j(\bar{c}_j + 1) - \hat{q}_j(\bar{c}_j)]$   
 6:    $\bar{c}_k \leftarrow \bar{c}_k + 1$   
 7:    $I_{total} \leftarrow I_{total} + [\hat{q}_j(\bar{c}_j + 1) - \hat{q}_j(\bar{c}_j)]$   
 8:    $\hat{q}_i(\sum \bar{c}_j) \leftarrow I_{total}$   
 9: **end while**

V. IQ-AWARE ROUTING PROTOCOL

In an event-driven WSN, data generated by a sensor provides information about the likelihood that a PoI has occurred. Section III describes how this data is mapped to the IQ  $q_a$  provided by an activated node  $v_a$ . To minimize the cost of data transmission, activated nodes first generate hints that are aggregated towards the fusion center  $v_0$  so that a minimum cost detection tree can be constructed. The minimum IQ  $I_T$  required to detect the PoI is assumed to be known.

Upon activation, each node  $v_j$  transmits hints to its downstream node  $v_i$  in the form of a quadruple comprising:

- 1) information quality  $q_j$  of  $v_j$ ;
- 2) maximum information quality  $q_j^M$  that can be obtained using the subtree rooted at  $v_j$ ;
- 3) maximum cost  $C_j^M$  of the subtree rooted at  $v_j$ ; and
- 4) histogram  $\{n_j^1, n_j^2, \dots, n_j^\phi\}$  representing the topology-aware IQ obtainable using the subtree rooted at  $v_j$ .

A. Initialization

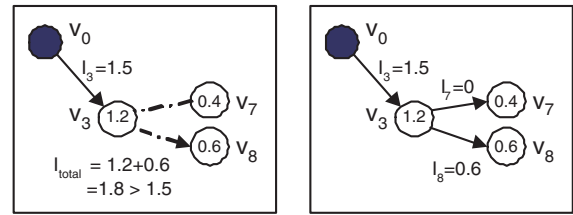
A distance-based aggregation tree is constructed using a shortest-path algorithm during network initialization. To maximize aggregation opportunities, data generated by an activated node is transmitted only after a delay that is inversely proportional to its distance from  $v_0$ . As an activated leaf node (without activated upstream nodes) has only one data point, building the histogram is trivial using the method in Section IV-B. Non-leaf nodes can have multiple upstream nodes.

B. Aggregation and Update

The histogram at node  $v_i$  is updated in three main phases.

- In Phase 1,  $v_i$  estimates the IQ  $\hat{q}_j(c_j)$  that can be obtained for each cost  $1 \leq c_j \leq C_j^M$ , using the subtree rooted at each upstream node  $v_j \in V_i^u$ .
- Phase 2 is triggered if  $v_i$  has multiple (activated) upstream nodes. A greedy heuristic is used to approximate the maximum IQ obtainable for each given cost  $1 \leq c_i \leq \sum C_j^M$ .
- In Phase 3,  $v_i$  incorporates its cost and IQ  $q_i$  into the IQ estimations obtained in earlier phase(s), and translates these estimations back into a (new) histogram for transmission to its downstream node.

We now detail each of these phases, using numerical examples from the topology in Figure 1, with  $I_T = 5$  and  $\phi = 5$ .



(a)  $v_3$  includes  $v_8$  into the pruned tree and updates  $I_{total}$ . (b)  $v_3$  assigns the appropriate IQ thresholds to its upstream nodes.

Fig. 3. Sequence of pruning activities for subtree rooted at  $v_3$ .

1) *Phase 1:* Estimation of the IQ  $\hat{q}_j(c_j)$  for the subtree rooted at each upstream node  $v_j$  is done by utilizing the histogram which is part of the quadruple transmitted from  $v_j$  to  $v_i$ . Considering the subtree rooted at  $v_3$ , the quadruples transmitted by  $v_7$  and  $v_8$  to  $v_3$  are  $\{0.4, 0.4, 1, \{1, \emptyset, \emptyset, \emptyset, \emptyset\}\}$  and  $\{0.6, 0.6, 1, \{1, \emptyset, \emptyset, \emptyset, \emptyset\}\}$  respectively.  $v_3$  then estimates the information quality  $\hat{q}_7(c_7) \forall 1 \leq c_7 \leq C_7^M$  and  $\hat{q}_8(c_8) \forall 1 \leq c_8 \leq C_8^M$  using Equation (12). Since  $v_7$  and  $v_8$  are both activated leaf nodes,  $v_3$  can easily and accurately estimate  $\hat{q}_7(1) = 0.4$  and  $\hat{q}_8(1) = 0.6$ .

2) *Phase 2:* From a global perspective, it is trivial to see that with a cost of 1, only  $v_3$  is included in the routing path. Similarly, with a cost of 3, all the three nodes ( $v_3, v_7$  and  $v_8$ ) in the subtree rooted at  $v_3$  are included. However, with a cost of 2, either  $v_7$  or  $v_8$  is included, with the latter yielding a higher cumulative IQ. The general complexity of computing the highest IQ for each cost  $1 \leq c_i \leq \sum C_j^M$  using an exhaustive brute-force search is  $\prod_{v_j \in V_i^u} C_j^M$ .

We approximate the maximum IQ for each cost using a greedy heuristic that significantly reduces computational complexity while maintaining reasonable IQ accuracy. Let  $I_{total}$  be the estimated cumulative IQ of all sensors that are included in the minimum cost tree  $M_i$  (initially empty). Recall that  $v_i$  has computed  $\hat{q}_j(c_j)$  for each incremental cost  $1 \leq c_j \leq C_j^M$  along each subtree rooted at  $v_j \in V_i^u$  in Phase 1. Let  $\bar{c}_j$  denote the current cost of the subtree rooted at  $v_j$ , that has been included in  $M_i$ ; initially,  $\bar{c}_j = 0 \forall v_j$ .

At each iterative step, the estimated IQ  $\hat{q}_k(\bar{c}_k + 1)$  that provides the maximum IQ gain to  $I_{total}$  is included in  $M_i$ . The IQ gain is computed by  $\hat{q}_k(\bar{c}_k + 1) - \hat{q}_k(\bar{c}_k)$ . This process repeats until: (i) all the subtrees rooted at the upstream nodes have been added to  $M_i$ ; or (ii)  $I_{total}$  exceeds the IQ threshold  $I_T$ . Based on the subtree rooted at  $v_3$  (and excluding itself) in Figure 1, the estimated maximum IQ values for the different costs using this greedy heuristic, are  $\hat{q}_3(1) = 0.6$  and  $\hat{q}_3(2) = 0.6 + 0.4 = 1.0$ . Algorithm 1 summarizes the IQ approximation procedure that is executed by  $v_i$  in Phase 2.

3) *Phase 3:* In the final step of the algorithm, the estimated maximum IQ obtained for the subtree rooted at  $v_i$  is updated to include its own cost and IQ  $q_i$ . The values of  $\hat{q}_3(c_i)$ , where  $1 \leq c_i \leq C_i^M$  are updated such that  $\hat{q}_3(1) = 1.2$ ,  $\hat{q}_3(2) = 1.2 + 0.6 = 1.8$  and  $\hat{q}_3(3) = 1.2 + 1.0 = 2.2$ . Based on this updated set of IQ estimations, a new quantized histogram is constructed and forwarded to the downstream node  $v_0$ .

### C. Pruning

The pruning phase commences after  $v_0$  receives data from its upstream nodes  $v_j \in V_0^u$ . Its objective is to prune off as many nodes as possible from the initial distance-based aggregation tree, such that: (i) IQ constraint  $I_T$  is still satisfied; and (ii) total transmission cost of collecting data from the resulting pruned tree is minimized. Hence,  $v_0$  has to allocate an IQ threshold  $I_j$  (with corresponding estimated cost  $\hat{c}_j \leq C_j^M$ ) to each upstream node  $v_j$  such that: (i)  $\sum I_j \geq I_T$ ; (ii)  $\hat{q}_j(\hat{c}_j) \geq I_j$ ; and (iii)  $\sum \hat{c}_j$  is minimized.

The pruning algorithm adopts a greedy approach similar to that in Phase 2 of the data aggregation algorithm. Based on previous computations, each node  $v_i$  has the estimated maximum IQ  $\hat{q}_j(c_j)$  of all its upstream nodes  $v_j \in V_i^u$  for each cost  $1 \leq c_j \leq C_j^M$ . The pruned routing tree is initially empty with total IQ  $I_{total} = 0$ .  $v_i$  iteratively includes into its pruned tree, the value of  $\hat{q}_j(c_j)$  that provides maximum IQ increment. This process repeats at  $v_i$  until: (i) subtree rooted at each upstream node  $v_j$  is included in the pruned tree; or (ii) aggregated IQ  $I_{total}$  of the pruned tree exceeds the IQ threshold  $I_i$  at  $v_i$ . The expected output of the pruning algorithm at  $v_i$  is the assignment of  $I_j$  to each upstream node  $v_j$ . The pruning algorithm is executed recursively at each upstream node  $v_j$  along the initial aggregation tree using  $I_j$ .

An activated node  $v_j$  with assigned IQ threshold  $I_j = 0$  is considered to be **pruned** off and not required to be part of the resulting pruned tree. Pruned nodes suppress their data for a time epoch before resuming the forwarding of data towards  $v_0$ . The temporary suppression of data enables the aggregation routing path to be adaptive towards dynamic changes in the network and PoI, while reducing transmission costs.

Assuming that  $v_0$  has assigned  $I_3 = 1.5$  in Figure 1, we look at how  $v_3$  assigns  $I_7$  and  $I_8$  to its upstream nodes  $v_7$  and  $v_8$ . Since  $\hat{q}_8(1) > \hat{q}_7(1)$ , the former is included into the pruned tree of  $v_3$  and  $I_{total} = q_3 + \hat{q}_8(1) = 1.2 + 0.6 = 1.8 > 1.5 = I_3$ , as in Figure 3(a). Since the IQ threshold at  $v_3$  is met, the pruning algorithm at  $v_3$  terminates with  $I_7 = 0$  and  $I_8 = 0.6$ , as shown in Figure 3(b).

### D. Discussion

The advantage of the proposed aggregation scheme is that while the fusion center sees a highly aggregated summary of how IQ is distributed in the WSN, the accuracy improves in the pruning process when specific subtrees are selected. Hence, one can think of the aggregation process as building a distributed structure that allows IQ distribution information to be selective refined as needed by the pruning process.

Finally, each sensor node has a low but non-zero probability of false positive event detection. If the network is sufficiently large, aggregation of a large number of sensors that falsely detect an event may be sufficient to trigger event detection. Such false alarms can be handled by the fusion center using simple heuristics: Since sensors with false positives are randomly distributed, the cost for event detection is very large as compared to that of normal event detection; hence, a simple cost threshold may be used to suppress such false positives.

## VI. PERFORMANCE EVALUATION

We evaluate the performance of IQAR in Qualnet 4.0 [28] and compare against the following routing protocols:

- 1) *aggTree*: Distance-based aggregation tree that collects data from all activated sensors, to be processed at  $v_0$ .
- 2) *walk*: IQ-aware routing protocol that routes data greedily towards next hop with highest IQ. When no IQ can be further gained from neighbors of the transmitter, or when aggregated data has sufficient IQ, data is routed back to  $v_0$  using a shortest path algorithm. Routing process is initiated from node with highest global IQ among all activated nodes, and there is only one ongoing transmission (and routing path) at any time.
- 3) *brute-force*: IQ-aware routing protocol that is similar in operations to IQAR, but uses an exhaustive brute-force search to compute the maximum IQ for each cost function during data aggregation and pruning.

The fusion center  $v_0$  is located near the bottom left hand corner of the terrain of size  $\{100m \times 100m\}$ . All the other sensors are uniform-randomly distributed in the network. An exponential sensing model [29] is used for function  $f(r_i)$  in Equation (1). The performance result illustrated is averaged over the sensing interval (1 second), and 20 seed runs. The target detection and false alarm probabilities are  $P_d = 0.9$  and  $P_f = 0.001$  respectively, and the transmission range is  $\approx 8m$ .

### A. Varying Local Information Quality

The per-sample false alarm probability  $p_0$  is increased in Figure 4, which leads to: (i) increase in number of activated nodes and detection region; and (ii) decrease in local IQ of each node. The PoI occurs at a fixed location  $\{80m \times 80m\}$  and the network has 250 nodes.

The aggregation cost in Figure 4(a) measures number of transmissions required to collect data from fused nodes, and is highly correlated to the number of fused nodes. Due to the lowered per-node IQ as  $p_0$  increases, more sensory data have to be aggregated to meet the IQ threshold  $I_T$ , resulting in increased aggregation cost for all protocols. *aggTree* incurs the highest aggregation cost as it fuses data from all the activated nodes. In contrast, the three IQ-aware protocols - *walk*, *brute-force* and *IQAR* - aggregate data from only a subset of the activated nodes and incur less aggregation costs.

The forwarding cost in Figure 4(b) measures number of transmissions required to forward data from the last aggregated node to  $v_0$ , and is dependent on activated node locations as well as PoI location. The enlarged detection region resulting from the increase in  $p_0$  leads to a corresponding increase in forwarding cost, especially for *aggTree* as it collects data from all the activated nodes. *IQAR* incurs higher costs than *brute-force* as the former adopts a greedy approach that may not yield the best routing path for a given IQ threshold. The gradual decrease in forwarding cost observed for all the protocols as  $p_0$  increases is due to the activation of *more* sensor nodes that are nearer to  $v_0$ , which decreases the distance between  $v_0$  and first aggregated node along each routing path.

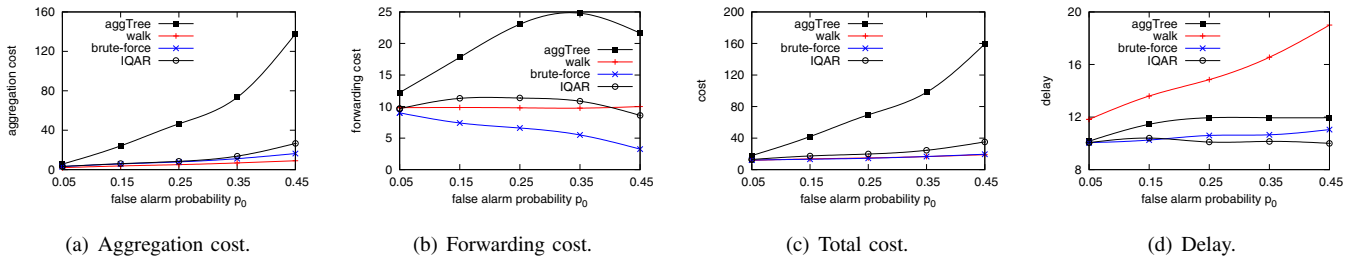


Fig. 4. Performance with increasing per-sample false alarm probability  $p_0$ .

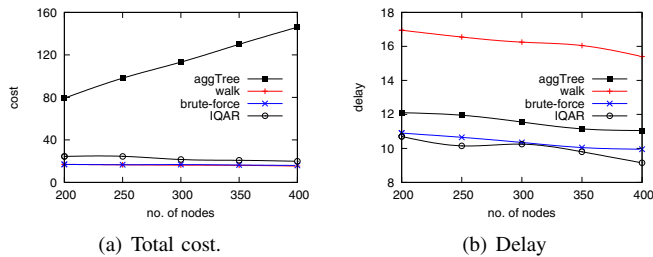


Fig. 5. Performance with increasing network density.

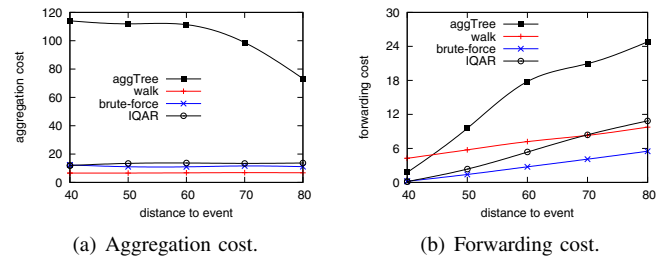


Fig. 6. Performance with increasing distance to event (PoI).

Figure 4(c) illustrates the total cost of aggregating and forwarding data packets to  $v_0$ . Since aggregation cost dominates over forwarding cost in a network with a small network diameter, the total cost has a similar trend to aggregation cost.

The delay incurred in Figure 4(d) is measured in terms of number of (sequential) transmissions. Despite the low transmission cost incurred by *walk*, it incurs the highest delay as transmissions occur sequentially along a single path, where aggregation occurs strictly before forwarding. Due to the presence of multiple paths in *brute-force* and *IQAR*, multiple aggregation and forwarding of data can take place simultaneously, thus reducing the overall delays.

### B. Varying Network Density

In Figure 5, network size increases from 200 to 400 nodes and PoI is located at  $\{80m \times 80m\}$  with  $p_0 = 0.35$ . Due to the increase in network size (and density), the number of activated nodes increase. The total cost incurred by *aggTree* increases correspondingly in Figure 5(a) as it collects data from all activated nodes. The remaining three IQ-aware protocols do not collect data from all the activated nodes and can achieve lower costs. The excessive delays incurred by *walk* in Figure 5(b) highlights the caveat of having only a single routing path that limits parallelism of data aggregation and forwarding.

### C. Varying Distance between Event (PoI) and Fusion Center

We vary the distance between the PoI and fusion center  $v_0$  in a network of 250 nodes with  $p_0 = 0.35$  in Figure 6. The  $x/y$  coordinates of the PoI are varied from 40m to 80m<sup>1</sup>.

As distance to event increases, it is located nearer to edges of the terrain, leading to decrease in number of activated nodes. Consequently, the number of fused nodes and aggregation

cost using *aggTree* decreases in Figure 6(a). Forwarding cost generally increases with increasing distance to PoI in Figure 6(b), as more transmissions are required to forward data from activated sensors to  $v_0$ . *IQAR* uses a greedy heuristic to estimate IQ contributions of nodes; hence it incurs higher forwarding cost than *brute-force*. Despite the low aggregation cost incurred by *walk*, it incurs relatively higher forwarding costs as the activated sensor with the highest global IQ may be further away from  $v_0$  than other activated nodes.

### D. Varying Suppression Interval

In the above scenarios, the PoI is statically located throughout the monitoring period. Both *brute-force* and *IQAR* can achieve significant cost and delay savings over *aggTree* as they aggregate data from only a subset of activated nodes to satisfy the IQ threshold. Data from the remaining activated nodes are suppressed for a **suppression interval** to reduce transmission costs<sup>2</sup>. However, one main concern with such protocols (which utilize data suppression to reduce costs) is whether a mobile PoI can be detected with sufficient IQ.

In Figure 7, the speed of the PoI is fixed at  $2.5 \text{ ms}^{-1}$  in a network of 250 nodes with  $p_0 = 0.35$ . As suppression interval increases from 2s to 10s, the amount of aggregated data decreases as nodes are suppressed for longer periods of time. Subsequently, there is a decrease in total cost and delay of *brute-force* and *IQAR* in Figures 7(a) and 7(b). However, it should be noted that due to the suppression of data, the detection accuracies achieved by these two protocols deteriorate by 5% to 10% with mobile PoIs. The total cost and delay incurred by *aggTree* and *walk* remain constant as they do not suppress data.

<sup>1</sup>A distance of 60 to the PoI implies that the PoI is located at  $\{60m \times 60m\}$ .

<sup>2</sup>With a suppression interval of  $x$  seconds, an activated node suppresses its data for at least  $x$  seconds after its last transmission.



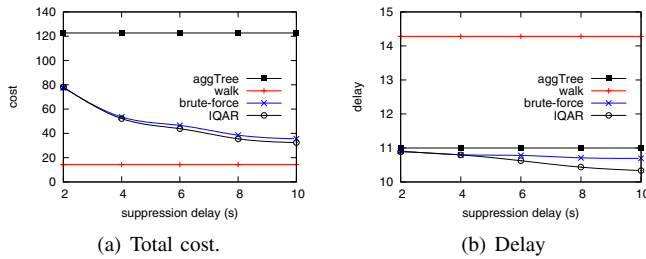


Fig. 7. Performance with varying suppression interval (delay).

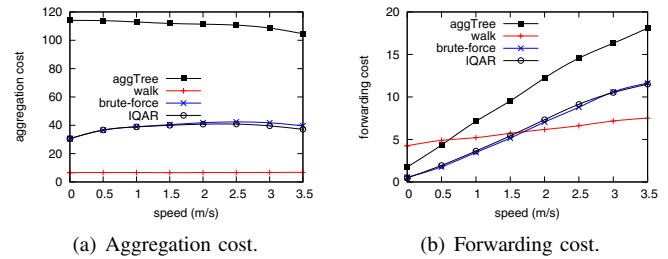


Fig. 8. Performance with varying event (PoI) mobility.

### E. Varying Event Mobility

In Figure 8, the PoI moves with varying speeds from  $0 \text{ ms}^{-1}$  to  $3.5 \text{ ms}^{-1}$  diagonally across the network with 250 nodes and  $p_0 = 0.35$ . The suppression interval is fixed at 5 seconds.

As event mobility increases, forwarding cost in Figure 8(b) increases as the PoI is located increasingly further away from  $v_0$ . As the PoI also exits the suppression region more quickly with higher event mobility, more sensors are activated, leading to the increased aggregation cost incurred by *brute-force* and *IQAR* in Figure 8(a). Since the PoI moves diagonally across the network, it is much closer to the network edge at higher mobilities, which limits the number of activated nodes. Hence, the aggregation costs incurred by *aggTree*, *IQAR* and *brute-force* drop slightly when speed exceeds  $2 \text{ ms}^{-1}$ .

## VII. CONCLUSION

In this work, we propose IQAR - an Information Quality Aware Routing protocol for event-driven sensor networks. IQAR considers the individual IQ contribution of each sensory data, and collects only sufficient data for a phenomenon of interest (PoI) to be detected reliably. Redundant data is suppressed for a time interval to reduce traffic load and alleviate medium access contention. This allows IQAR to achieve significant energy and delay savings while maintaining information quality in event detection.

## ACKNOWLEDGMENT

This work is done under the USCAM-CQ project which is part of the UWB-Sentient Computing Research Programme funded by SERC, A\*STAR Singapore.

## REFERENCES

- [1] T. Q. S. Quek, D. Dardari, and M. Z. Win, "Energy Efficiency of Dense Wireless Sensor Networks: To Cooperate or Not to Cooperate," *JSAC*, vol. 25, no. 2, 2007.
- [2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless Sensor Networks: A Survey," *Computer Networks*, vol. 38, no. 4, 2002.
- [3] H. Zhang, A. Arora, Y. Choi, and M. G. Gouda, "Reliable Bursty Convergecast in Wireless Sensor Networks," in *MobiHoc*, 2005.
- [4] B. Krishnamachari, D. Estrin, and S. Wicker, "Modeling Data-Centric Routing in Wireless Sensor Networks," in *INFOCOM*, 2002.
- [5] L. Yu and A. Ephremides, "Detection Performance and Energy Efficiency Trade-off in a Sensor Network," in *Allerton Conference on Communication, Control, and Computing*, 2003.
- [6] R. M. Karp, "Reducibility Among Combinatorial Problems," in *Symposium on the Complexity of Computer Computations*, 1972.
- [7] K. W. Fan, S. Liu, and P. Sinha, "Scalable Data Aggregation for Dynamic Events in Sensor Networks," in *SensSys*, 2006.

- [8] M. C. Vuran, O. B. Akan, and I. F. Akyildiz, "Spatio-Temporal Correlation: Theory and Applications for Wireless Sensor Networks," *Computer Networks*, vol. 45, no. 3, 2004.
- [9] K. W. Fan, S. Liu, and P. Sinha, "On the Potential of Structure-Free Data Aggregation in Sensor Networks," in *INFOCOM*, 2006.
- [10] H. Luo, Y. Liu, and S. K. Das, "Routing Correlated Data with Fusion Cost in Wireless Sensor Networks," *TMC*, vol. 5, no. 11, 2006.
- [11] C. Intanagonwiwat, D. Estrin, R. Govindan, and J. Heidemann, "Impact of Network Density on Data Aggregation in Wireless Sensor Networks," in *ICDCS*, 2002.
- [12] A. Goel and D. Estrin, "Simultaneous Optimization for Concave Costs: Single Sink Aggregation or Single Source Buy-at-Bulk," in *Annual ACM-SIAM Symposium on Discrete Algorithms*, 2004.
- [13] W. Zhang and G. Cao, "Optimizing Tree Reconfiguration for Mobile Target Tracking in Sensor Networks," in *INFOCOM*, 2004.
- [14] —, "DCTC: Dynamic Convoy Tree-based Collaboration for Target Tracking in Sensor Networks," *Transactions on Wireless Communications*, vol. 3, no. 5, 2004.
- [15] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks," *Transactions in Wireless Communications*, vol. 1, no. 4, 2002.
- [16] S. Lindsey and C. S. Raghavendra, "PEGASIS: Power-Efficient Gathering in Sensor Information Systems," in *Aerospace Conference*, 2002.
- [17] S. Pattem, B. Krishnamachari, and R. Govindan, "The Impact of Spatial Correlation on Routing with Compression in Wireless Sensor Networks," in *IPSN*, 2004.
- [18] M. Chu, H. Houssecker, and F. Zhao, "Scalable Information-Driven Sensor Querying and Routing for Ad Hoc Heterogenous Sensor Networks," *IJHPCA*, vol. 16, no. 3, 2002.
- [19] F. Zhao, J. Liu, J. Liu, L. Guibas, and J. Reich, "Collaborative Signal and Information Processing: An Information-Directed Approach," *Proceedings of the IEEE*, vol. 91, no. 8, 2003.
- [20] J. Liu, F. Zhao, and D. Petrovic, "Information-Directed Routing in Ad Hoc Sensor Networks," *JSAC*, vol. 23, no. 4, 2005.
- [21] L. Yu, L. Yuan, G. Qu, and A. Ephremides, "Energy-Driven Detection Scheme with Guaranteed Accuracy," in *IPSN*, 2006.
- [22] L. Yu and A. Ephremides, "Detection Performance and Energy Efficiency of Sequential Detection in a Sensor Network," in *HICSS*, 2006.
- [23] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Detection in Dense Wireless Sensor Networks," in *WCNC*, 2007.
- [24] A. Wald, "Sequential Tests of Statistical Hypotheses," *Annals of Mathematical Statistics*, vol. 16, no. 2, 1945.
- [25] R. Viswanathan and P. K. Varshney, "Distributed Detection With Multiple Sensors: Part I - Fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, 1997.
- [26] S. C. A. Thomopoulos, R. Viswanathan, and D. K. Bougoulas, "Optimal Distributed Decision Fusion," *Transactions on Aerospace and Electronic Systems*, vol. 25, no. 5, 1989.
- [27] H. X. Tan, M. C. Chan, W. Xiao, P. Y. Kong, and C. K. Tham, "Information Quality Aware Routing in Event-Driven Sensor Networks," School of Computing, National University of Singapore, Tech. Rep., 2009, [http://www.comp.nus.edu.sg/~tanhweeX/research/iqar\\_2009.pdf](http://www.comp.nus.edu.sg/~tanhweeX/research/iqar_2009.pdf).
- [28] Qualnet Network Simulator, Scalable Network Technologies, Inc. [Online]. Available: [www.qualnet.com](http://www.qualnet.com)
- [29] K. Chakrabarty and Y. Zou, "A Distributed Coverage- and Connectivity-Centric Technique for Selecting Active Nodes in Wireless Sensor Networks," *Transactions on Computers*, vol. 54, no. 8, 2005.