Collaborative Sequential-based Detection in Wireless Sensor Networks

Sabina Zejinilović*, Joao Pedro Gomes†, Bruno Sinopoli‡

*†Department of Electrical and Computer Engineering
Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213

†Institute for Systems and Robotics — Instituto Superior Técnico (ISR/IST)
Technical University of Lisbon (UTL), Lisbon, Portugal

Abstract—Due to limited power resources, energy efficiency is an important aspect of detection in wireless sensor networks. We propose a collaborative detection scheme, based on sequential hypothesis testing, where a randomly chosen node may initiate a collaboration, collecting observations from neighboring nodes to test the hypotheses. Our simulation results show that for large networks and high SNR, the proposed scheme leads to lower communication cost and similar performance when compared to a standard detection setup, where all the observations are collected in a fusion node. We also examine how the energy efficiency of this scheme evolves as a function of the network structure.

I. INTRODUCTION

Wireless sensor networks (WSN) have found application in various fields due to their collaborative sensing power, adaptability, low cost, and rapid deployment [1]. Many applications, such as environmental, medical, or industrial monitoring, include the task of detecting events. An event of interest could be the presence of a chemical leak in an industrial environment, the presence of a contaminating source in a water reservoir or appearance of an intruder in a secure location. Sometimes, a WSN’s main objective is detection itself, as in a surveillance application. Other times, detection is a step that must precede estimation; for example, once the chemical leak has been detected, its location can be estimated.

Sensor nodes are battery powered and hence have strict energy restrictions. A wireless node’s radio significantly contributes to energy spending and the ratio of the energy spent in sending one bit of information to the energy spent in executing one instruction can be in the order of thousands [2]. Therefore, most applications tend to decrease the communication load at the expense of performing more local computation. The collaboration of sensor nodes offers opportunity to perform energy optimization, not only at a node, but at the network level.

In a standard multisensor centralized scenario, observations collected by all sensor nodes are transferred to a fusion center where hypothesis testing is performed. Although the decision is based on all the available information, this approach has several disadvantages, especially prominent in large scale networks: abundant (and costly) communication, large processing power required at the fusion node, and increased vulnerability of the network to fusion node failure. In decentralized detection, nodes perform local processing and instead of sending the raw observations they send their summaries. The fusion node now performs detection with only partial observations, so there is a loss of performance compared to a centralized detection. This loss can be made small by optimal local processing of observations. A survey of fundamental results in this area can be found in [3].

In sequential hypothesis testing, the objective is to reach a decision while minimizing not only the probability of error but also the number of observations needed for the decision. This is especially significant in the circumstances where gathering observations carries a cost, in terms of incurred delay in reaching a decision or incurred energy cost of transmitting observations, as it is the case in wireless sensor networks. Initially formulated as a centralized procedure, sequential detection was also explored in decentralized settings, and a review of results in this area is given in [4]. In general, sequential detection requires considerably fewer observations than fixed detection while achieving the same probability of error [5].

With this motivation, we propose a detection scheme where the observations are collected sequentially, not over time, but over a collaborative subset of sensors. The goal is to decrease the communication cost of the detection process, while attaining an acceptable level of performance. Energy-efficient detection was also considered in [6], where the cost of taking and transmitting observations was added to the detection problem, but not in the context of sequential testing. Optimal rules were derived for censoring the sensors with non-informative observations. Optimal policies for sequential testing in decentralized settings were considered in [7], for the case where each node chooses when to send its own decision to another node which, in turn, chooses when and how to incorporate it into its own decision. In our approach we consider that the nodes are sending observations to collaborating nodes, not their local decisions. This is justified by the fact that the cost of transmitting a single observation is almost the same as the cost of transmitting a local decision.
current wireless systems, while sending the observations is more informative than just their summaries. Additionally, to increase the robustness of the network and ensure that batteries are evenly drained from all nodes, the node that initiates collaboration and collects observations from other nodes is chosen at random. In traditional settings, nodes that are closer to the fusion center may spend their power supply much faster than the other nodes, as they aggregate messages from further parts of network, which may result in early network partitioning.

We analyze the energy efficiency of this approach and explore under which conditions concerning the network size, topology and signal to noise (SNR) ratio this simple scheme results in smaller energy spending when compared to standard centralized detection. The remainder of the paper is organized as follows. In Section II we describe the proposed collaborative detection scheme. In Section III we evaluate the performance based on computer simulations. Section IV gives the conclusion.

II. COLLABORATIVE DETECTION

We assume a network of wireless sensors that are distributed over a monitoring area with the purpose of deciding in favor of absence or presence of a known signal, corrupted by noise, represented by hypotheses $H_0$ and $H_1$, respectively. All sensors periodically gather observations described by

$$
H_0 : \quad y_{ij} = n_{ij}
$$

$$
H_1 : \quad y_{ij} = s_i + n_{ij},
$$

where $y_{ij}$ is the $j$-th observation of the $i$-th sensor, $n_{ij}$ is the corresponding measurement noise, and $s_i$ is the signal received by the $i$-th sensor. We assume noise is Gaussian, uncorrelated both in time and across sensors. This model can be used for detecting a signal point source, for example, the source of leakage of some contaminating chemical, or a distributed signal source, where the signal is perceived at relatively larger regions of the monitored area [8].

Each node has a clock that ticks at the times of a rate $\mu$ Poisson process. Once a node’s clock ticks, the node makes a decision based on its own observations and, if needed, the observations of neighboring nodes. We will refer to this node as the initiating node. Inter-tick times are assumed longer than the total time needed for the initiating node to query its neighbors and receive their observations. Therefore, only one collaborating group is formed and a decision is reached between any two successive clock ticks. From a network’s perspective, nodes tick uniformly at random, resulting in uniform surveillance of the whole monitored area.

Based on its own observations, the initiating node, for given bounds on probability of false alarm $F_F$ and probability of a miss $P_M$, performs the standard sequential hypothesis test by comparing the likelihood ratio $\Lambda (y)$ to two thresholds, $\eta_0$ and $\eta_1$

$$
\Lambda (y) < \eta_0 \quad \text{choose } H_0
$$

$$
\eta_1 < \Lambda (y) < \eta_0 \quad \text{more data needed}
$$

$$
\eta_1 < \Lambda (y) \quad \text{choose } H_1.
$$

The following approximation is often used in determining the thresholds which minimize the number of observations needed to make a decision for given bounds on $P_F$ and $P_M$ [5]

$$
\eta_0 = \frac{P_M}{1 - P_F}, \quad \eta_1 = 1 - \frac{P_M}{P_F}.
$$

If the initiating node’s likelihood ratio is below the lower threshold, the node decides in favor of $H_0$. Otherwise, the node seeks collaboration to refine the test and decrease the possibility of false alarm, even if its likelihood ratio is already above threshold $\eta_1$. Often, the events that a sensor network monitors are rare, hence the initiating node requires confirmation of its neighbors before choosing $H_1$. The initiating node determines the expected number of observations, $L$, needed to complete the test under each of the hypotheses, which is given by

$$
E[L \mid H_i] = \frac{E[\log \Lambda_k (y) \mid H_i]}{E[\log L_k (y) \mid H_i]},
$$

where $i = 0, 1$, and $\Lambda_k (X)$ denotes the likelihood ratio obtained for observations $1, \ldots, k \leq L$.

To accommodate for the scale of the network and possible node and link failures, we assume that nodes do not possess knowledge of the whole network, but only of their immediate neighbors. An initiating node cannot determine exactly what size of neighborhood it must query in order to gather the required number of observations that, on average, would result in reaching a decision. It assumes neighborhoods are of the same size, equal to the number of its immediate neighbors $d$. This assumption is based on the fact that, for large networks, the geometric random graph is a regular graph with high probability [9]. With this assumption, the size of the neighborhood $k$ that gets queried is determined by

$$
\text{minimize} \quad k
$$

subject to $\sum_{j=1}^k d (d-1)^{j-1} \geq L - 1$.

Nodes that get queried by an initiating node respond by sending their observations. To decrease the number of transmissions, each node aggregates the information from its “children” nodes before forwarding its observation to the parent node. Once the initiating node receives the requested observations, it again performs sequential hypothesis testing. If after this step the value of the likelihood ratio is still between the two thresholds, the size of the neighborhood that gets queried increases by one. The initiating node repeats the test, now including the observations from nodes that are $k + 1$ hops away from it. This process is repeated until either a decision is reached or the neighborhood cannot grow any further. In case the likelihood ratio fails to cross either $\eta_0$ or $\eta_1$ even after the initiating node has gathered all the available observations, then the rule for a truncated test is applied. The hypothesis
neighbors of initiating node are queried, the total number of neighbors themselves. This means that if neighbors of all of nodes are clustered, i.e. sensor’s neighbors can also be sensor nodes have the same number of neighbors, sensor as the network does not have a tree topology. Even if all is given by
\[ C_L \]
The cost
\[ L \]
total number of observations
\[ k \]
reached, the algorithm proceeds with additional iterations. Let
\[ L' \]
The cost
\[ C_{L'} \]
of collecting
\[ L' \]
observations equals
\[ C_{L'} = 2L' - 1 - d(d - 1)^{k-1}. \]
If, after querying
\[ k \]
hop neighborhood, a decision cannot be reached, the algorithm proceeds with additional iterations. Let
\[ p \]
denote the number of additional iterations of algorithm after
\[ k \]
hop neighborhood has been queried for observations. The total number of observations
\[ L'' \]
collected then corresponds to
\[ L'' = L' + \sum_{j=k+1}^{k+p} d(d - 1)^{j-1}. \]
The cost
\[ C_{L''} \]
of querying and collecting of
\[ L'' \]
observations is given by
\[ C_{L''} = (2L' - 1)(p + 1) - d(d - 1)^{k-1} + \sum_{i=0}^{p-1} (2p - 1 - 2i) d(d - 1)^{k+i}. \]
(2)
The actual cost for collecting
\[ L'' \]
observations can be higher, as the network does not have a tree topology. Even if all the sensor nodes have the same number of neighbors, sensor nodes are clustered, i.e. sensor’s neighbors can also be neighbors themselves. This means that if neighbors of all of
\[ d \]
neighbors of initiating node are queried, the total number of queried sensors might be much smaller than
\[ d(d - 1), \]
as sensor nodes might have overlapping neighborhoods. However, expression (2) gives a lower bound on the cost of collection
\[ L'' \]
observations for the proposed collaborative algorithm in a network where the sensor nodes have
\[ d \]
nighbors. Assuming that there are
\[ N \]
 nodes in a network and that the observations are aggregated across the sensor nodes, the cost of transmitting the observations of all the sensors to a central location equals
\[ N. \]
In the cases where the proposed collaborative scheme is more energy efficient than the centralized scenario and the decision is reached with
\[ L'' \]
observations, then the number of neighbors
\[ d \]
and the number of additional iterations of the algorithm
\[ p \]
have to be such that
\[ C_{L''} < N. \]

III. PERFORMANCE EVALUATION
To evaluate the proposed collaborative scheme, we compare its energy efficiency and performance to standard centralized detection with a fixed fusion node. The metric we chose for evaluating the energy spent in reaching a decision is the number of transmissions between sensor nodes. This is because the communication cost dominates over all other costs for wireless sensor nodes. Transmissions that were counted to approximate energy spent include queries for observations and sending of observations. The comparison was done for different values of SNR and for different network sizes in order to determine the conditions under which sequential collaboration can achieve comparable performance to the standard scheme, but on a smaller energy budget. Simulations were performed for networks of 30 nodes that were positioned uniformly at random. The transmission radius of sensor nodes was chosen to ensure that the network is connected. The fusion node present in a centralized setting was placed in the center of the network. The distributions of sensor observations, \( y \), under the two hypotheses were
\[ H_i : y \sim \mathcal{N}(m_i, \sigma^2 I), \quad i = 0, 1, \]
with known \( m_i, \sigma^2. \) Averaging of results was performed over all sources of randomness (observations and node placements). Overall, for
For each value of SNR and each network size, we calculated the percentage of 50,000 runs where the collaborative scheme had smaller transmission cost than the standard centralized scheme. Figure 1a shows this percentage increasing as expected with increasing values of SNR. With higher SNR, observations of each node are more reliable, thus we need less of them to perform detection. A collaborative detection group thus comprises fewer nodes, and the total number of transmissions stays small. By contrast, all nodes participate in the centralized setting, sending redundant information and resulting in higher energy spending.

The corresponding probabilities of error of the two schemes are shown on Figure 1b, on logarithmic scale. In the centralized setting, detection was based on the Neyman-Pearson criterion, where the probability of detection is maximized for a given bound on the probability of false alarm. Standard sequential detection involves setting the bounds on both probability of false alarm and probability of miss. The achieved error probability can differ from the given bounds due to approximations (1) for determining thresholds and due to the effect of truncation — when there is lack of observations needed to complete the test. In the centralized scheme decisions are based on all available information, therefore its performance is, as expected, superior to that of a collaborative scheme which includes only a limited number of observations. However, our results for both approaches are comparable.

Figure 2a illustrates the effect of growing network size on the percentage of simulations where collaboration led to lower energy spending. For fixed SNR, increasing the size of the network increases the energy efficiency of collaboration. The graph shows phase transition behavior where, for networks up to 25 nodes, the centralized scheme is dominantly more energy efficient, while for networks above 35 nodes, the opposite holds. Thus which scheme is more energy-efficient completely changes by adding about ten nodes to the network. Figure 2b shows the corresponding probabilities of error for both schemes. Similar conclusions can be drawn as in the case of increasing SNR, where the attained level of error of the collaborative scheme is higher, but still comparable to that of a centralized detector.

For instances of SNR and network size where the two schemes have comparable number of transmissions, and no scheme
clearly dominates, we explore the effect of network topology on energy efficiency. Figure 3a shows the effect of average degree of the network and figure 3b the effect of average path length, for a network of 30 nodes and for $\sigma^2/m_1^2 = 0.6$. The average degree denotes the average number of neighbors that a node has. If each node is highly connected, then the size of the collaboration group increases substantially with each hop. With just the query of its immediate neighbors, the initiating node might end up with more observations than needed for a decision. More energy is spent than in a centralized scheme, as collaboration includes querying for observations, while observations are just periodically sent to the fusion node of a centralized detector. Therefore, the collaborative scheme is more favorable in topologies with smaller average degree.

The average path length represents the average number of hops between nodes in the network; the smaller it is, the more interconnected nodes are. Topologies with lower average path length result in fast growth of collaborating groups, as explained above, resulting in lower energy efficiency when compared to standard centralized detection.

IV. CONCLUSION

We presented a collaborative detection scheme in wireless sensor networks where observations are sequentially collected across sensor nodes, including hop-by-hop neighborhoods, until a reliable decision can be reached. Simulation results show that, in large networks and for high SNR values, the detection performance of sequential collaboration is comparable to that of centralized detection, while requiring fewer transmission and thus consuming less energy. As for networks where the number of nodes and the value of SNR equally favor collaborative and the centralized scheme, collaborative scheme is more energy-efficient in networks with lower average degree and higher average path length.

REFERENCES