A Semantic Clustering Routing Protocol for Wireless Sensor Networks

F. Bouhafs, M. Merabti, and H. Mokhtar
Computing and Mathematical Sciences School, Liverpool John Moores University, Liverpool, UK
{cmpfbouh, M.Merabti, H.M.Mokhtar}@livjm.ac.uk

Abstract

The paper presents a new clustering scheme for wireless sensor networks based on semantic properties. Unlike current clustering schemes where signal strength and neighbouring are the most used criteria in the cluster formation, in this work, semantic properties are also taken in consideration in the process of cluster formation. Nodes inside a same cluster are organized like a hyper tree where the cluster-head is the root. This hyper tree organization allows a layered data aggregation, avoids the clusterhead overload and energy dissipation, and makes the network more robust against links and nodes failures. We evaluate this work by simulations and show that our approach is more efficient and more energy conservative than other clustering schemes.

1. Introduction

Sensor networks are a promising technology for applications ranging from environmental monitoring to industrial asset management. This type of networks is expected to change our life in many ways, in schools, hospitals, houses, and many other places. Sensor networks are, generally, wireless and usually formed of tiny, low-power, low-cost, multifunctional, sensor nodes that could be mobile and may be deployed in unfamiliar environments. These tiny sensor nodes consist of sensing, data processing and communicating components, and communicate unthithered in short distances. Researchers, riding on advances in micro-electro-mechanical systems technology, expect that this kind of networks will become smaller, cheaper and thus deployed in large number. However, despite these advances wireless communication will continue to dominate the energy consumption of embedded networked systems for the foreseeable future, and thus, there is a real need to design communication techniques that could minimize the amount and range of communication as much as possible, in order to prolong the life time of the sensor network. In this paper we address this challenge by developing a new clustering scheme that involves only nodes that are relevant to a given query or task, and groups them in a cluster, which offers a possibility to minimize the communication energy cost through local collaboration and data aggregation. The remainder of the paper is structured as follows. Section 2 will review the major communications techniques found in the literature, and will outline their drawbacks. Section 3 will introduce the semantic routing paradigm and will describe its properties. Section 4 will give an overview on our protocol and describe its different phases. In section 5 we will evaluate our protocol by simulations. Section 6, will outline open research issues and future work related to semantic clustering routing.

2. Background

Due to the sensor networks characteristics, designing efficient communication architecture brings many new challenges. Routing is one of the most important challenging tasks in sensor networks. It has attracted a lot of attention in recent years. Several routing mechanisms have been proposed and can be classified into two major types: Data-centric routing, and Hierarchical clustering routing, although there are few distinct ones based on sensor nodes location, network flow or quality of service (QoS) awareness.

Hierarchical clustering routing approaches maintain the energy consumption of sensor nodes by involving them in multi-hop communication within a particular cluster. Cluster formation is based, generally, on the energy reserve of sensors and sensor’s proximity to the cluster head. The protocol LEACH developed in [1] is one of most popular clustering algorithms in sensor
networks. The idea is to form clusters of sensor nodes based on the received signal strength and use local cluster heads as routers to the sink. Data aggregation and fusion are local to the cluster, this will save energy since the transmissions will only be done by such cluster heads rather than all sensor nodes. Cluster heads change randomly over time in order to balance the energy dissipation of nodes. The idea proposed in [1] has been an inspiration for many hierarchical routing protocols [2, 3]. While all these works focus on saving energy by aggregating the collected data, they present several drawbacks, the most important are:

- **Cluster formation is not energy-efficient:** Since each elected cluster-head has to broadcast an advertisement message to all nodes and regulate its transceiver power such as all sensor nodes could hear this message.

- **Cluster-head overload:** Since all cluster members send their data to the cluster-head, in a dense network where clusters contains thousands of nodes, the cluster-head will dissipate a great part of its energy in this operation, and thus a cluster-head re-election is needed too often.

In data-centric routing, the sink sends queries to certain regions and waits for data from the sensors located in the selected regions. Since data is being requested through queries, attribute-based naming is necessary to specify the properties of data. SPIN (Sensor Protocols for Information via Negotiation) [4] is the first data-centric routing protocol which considers data negotiation between sensor nodes in order to eliminate redundant data and save energy. In SPIN a high-level descriptors or meta-data are exchanged among sensors via a data advertisement mechanism before data transmission. Directed diffusion [5] is another data-centric communication paradigm where a sink sends out a request for data by broadcasting an interest to its neighboring nodes. Each neighbour subsequently broadcasts the interest to its neighbours until it reaches a node capable of servicing the request. As interests diffuse throughout the network, a node that receives an interest from a neighbouring node forms a gradient pointing to the sending node that indicates the direction in which data from a source node will eventually flow. The source node then generates data messages using its sensors which propagate back to the sink following the gradients formed along the paths through which the interest originally traversed. Every sink that receives data messages from more than one neighbour, reinforces a particular neighbour so that subsequent data messages arrive only from the chosen neighbour. This chosen neighbour also performs the same procedure on its neighbouring nodes it received a data message from. This process is repeated until data messages propagate only along the reinforced path from source to sink. If the quality of data transmission from a certain neighbour deteriorates, a node can opt to negatively reinforce another better-performing neighbour instead, in order to cope with varying network dynamics. Many other protocols have been proposed either based on Directed Diffusion or following a similar concept [6, 7]. However, although there is generally some correlation between sensor nodes that respond to the query, surprisingly, this correlation is not exploited, and no coordination is established between these sources.

3. Semantic Routing

In many sensor networks applications the user’s query or the application task may inherently specify a limited logical scope and defines the nodes involved in this task or which nodes should answer the user’s query. Rather than flooding the entire network, the querying system and the networking layer might instead coordinate to provide efficient data dissemination and semantically scope floods.

A query system may define some policies so query messages are delivered only to nodes that satisfy a particular application’s condition or user’ query. An example of semantic routing is when the query processor influences network topology formation to optimize in-network processing. It allows the application to achieve timeliness, support multiple concurrent queries, and trade off quality for greater energy savings. However, these benefits are in contradiction with current routing approaches where paths between sources and sinks are optimized for reliable and shortest-path delivery.

In order to achieve a more data aggregation and correlation opportunities, it may be better for the query processing system to choose less reliable paths, as such paths can reduce the overall transmission load on the network. Recall that sensor networks are applications specific, and information collected by nodes are highly correlated, thus, an ideal routing protocol would be able to exploit in-network processing as much as possible while still delivering the end result to their destinations. Such routing protocol would consider both the semantic information from the query or the task and link-layer reliability and connectivity properties learned from neighbouring nodes.
4. A semantic clustering architecture

The main idea of the clustering scheme proposed in this work is to group sensor nodes in clusters such as the clustering policy considers both semantic information and connectivity properties. A user query is disseminated through the sensor network looking for a specific group of sensor nodes. Unlike Directed Diffusion where each node has to propagate the query to its neighbours whether it satisfies this query or not, in our approach, upon the query reaching a node satisfying query, this node will elect itself as cluster-head and start forming a cluster that contains all nodes in its region that satisfy the same query, by sending an advertisement message that contains the query to its neighbours. In this section we describe with details the elements of our protocol.

4.1. Interest propagation phase

This phase starts by sending the query from node to node. Each node upon receiving this query checks whether it satisfies this query or not as described in figure 1.

In this phase we distinguish two cases:

1. The sensor node does not satisfy the query: In this case the node must send the query to all its neighbours.
2. The sensor node satisfies the query: In this case the sensor node launch a cluster formation phase.

4.2. Cluster Formation Phase

The cluster formation is started by the first node that satisfies the user query. Unlike LEACH protocol which broadcasts advertisement messages to all sensor nodes in the network, in our approach a cluster-head sends an advertisement message to just its neighbours.

This advertisement message will contain the cluster-head’s ID, a header that indicates the type of the message, and the query. Each neighbour receiving the cluster-head advertisement message will check if its data satisfies this query and decides whether it will join the cluster or not. Afterward, each cluster-head’s neighbour that has decided to join the cluster must inform the cluster-head that it wants to be a member of the cluster by sending a join-request message back to the cluster-head. This message will contain the node’s ID, the cluster-head’s ID and a header that indicate the type of the message. This operation is considered to be the first step in the construction of the cluster as well as the semantic hyper tree, where the cluster-head is the root of the tree and the neighbours nodes that decide to join the cluster will be its direct children’s nodes.

4.3. Data Dissemination Phase

This phase can be considered as the final stage in this clustering scheme. Each node inside the cluster...
gathers its data and pushes it toward its parents in the semantic hyper tree. Each parent node waits until it receives gathered data from all its children nodes, apply an aggregation operator on it and send the result to its parent. When the cluster-head receives all aggregated data, it applies its aggregation operator on it and sends the result toward the sink following the shortest multi-hop path like in Directed Diffusion. We present an example of the data formation in figure3.

**Fig.3. Example of data dissemination**

### 4.4. Cluster Maintenance

Although the cluster is built on a constant queried event, some cluster maintenance must occur. In particular, new nodes may appear, link qualities may change, and existing nodes may not satisfy the user query anymore. We define a sensing interval $SNI$: the maximum time in which each child node must inform its parent about its sensed information. If a node finds that its data does not satisfy any more the query $q$ it stops sending data to its parent. If a parent node does not hear from its child node for some number of sensing intervals, it removes its tree entry. For new nodes appearance, we assume that former nodes neighbours that did not satisfy the query could check this condition at each time interval $t$. When query is satisfied, a node will ask the neighbour from where it hears the last advertisement message to join the cluster and this neighbour will be its parent in the tree.

### 5. Evaluation

To evaluate our protocol we have used the Georgia Tech Network Simulator (GTNetS)[8] and we used the same radio model discussed in [1]. For our experiments we simulated a 100-node network where nodes were randomly distributed between $(x=0,y=0)$ and $(x=100,y=100)$ with the base station (BS) at location $(x=50,y=175)$. Although our protocol is a query based clustering protocol where just nodes that satisfy the user query communicate with the user, we chose to involve the most possible number of nodes in the communication, in order to assesse the clustering scheme efficiency. For that, we added to the simulated nodes temperature sensing capability and we fixed the temperature in the study field between 50 degrees and 52 degrees. A query was broadcasted from the BS looking for the position of nodes detecting a temperature exceeding 50 degrees. Each node detecting this event sends a position message to the user. The results obtained in this work were compared to the results obtained by LEACH with the same simulations parameters which are presented in [1]. Figure 4 shows the total number of nodes that remain alive over time for our semantic clustering protocol and LEACH.

**Fig.4. Number of nodes alive over time**

From this figure, we could observe that in our protocol nodes stay alive much longer than in LEACH. This is because our scheme performs aggregation of data at each level of the hyper tree, thus, the amount of data to be sent is reduced greatly at the expense of more delay. Figure 5 shows the total number of data messages received at the BS over time. From this figure, we can observe that our protocol sends much less data to the base station than LEACH. These results are due mainly to two reasons:

1. The cluster members in LEACH send their data to the cluster head by one hop transmission while in our protocol the cluster members send their data by multihop transmission which takes longer time to reach the cluster head, thus our protocol is not suitable for time-sensitive applications.

2. As the data traveling from the cluster member to the cluster head is aggregated at each level of the tree, the amount of data received by the cluster head is significantly reduced which helps the cluster head to save its energy and, thus, avoiding to elect a new cluster head.

To compare the distribution of energy consumption in both protocols we measured the number of nodes alive according to the consumed energy in the network.
Figure 6 shows the number of nodes alive per amount of energy consumed by the overall network. This figure shows that, in our protocol, despite the fact that 95% of the initial network energy has been consumed (190 Joule), more than 90% of network nodes were still alive, while in LEACH, for the same amount of energy consumed, just 35% of network nodes were still alive. These results prove that our protocol achieves a better energy consumption distribution among the cluster members, and thus, maintains the network lifetime much longer, and avoid a new cluster head election.

6. Conclusions and future work

In this paper, we have introduced a new clustering scheme for wireless sensor networks that uses semantic properties as its foundation. We described a new cluster formation model that considers relevancy of nodes to a specific query and organizes nodes within a cluster in a tree, and showed using simulation that our new clustering scheme is more energy efficient and achieves better energy. We are currently working on mobility and inter-cluster routing as we believe that both are two important issues to take in consideration. Once solutions for these issues have been fully designed we plan to integrate all these developments in the same framework.

7. References


