A Flexible Cluster-Based Approach for Architecting Wireless Sensor Networks

Bruno Galvão[1], Flávia Coimbra Delicato[1] , Paulo F. Pires [1,2],
Luiz Fernando Rust da Costa Carmo[1], Luci Pirmez[1]

Núcleo de Computação Eletrônica – NCE[1] & Computer Science Department – DCC[2]
Federal University of Rio de Janeiro
P.O Box 2324, Rio de Janeiro, RJ, 20001-970, Brazil
bruno.galvao@cebim.com.br, {fdelicato, paulopires, rust, luci} @ nce.ufrj.br

Abstract

This work presents CLIN, a CLuster-based and INterest-oriented protocol for data dissemination in wireless sensor networks (WSN). The protocol proposes an energy efficient solution for transmitting application requested data in WSN. A hierarchical topology is used and an interest-driven approach is adopted. The protocol dynamically adapts the network behavior to the user requirements, aiming to assist a large range of applications. Simulation results and a comparison with existent protocols are presented to validate the proposal.

Keywords

Wireless Sensor Networks, Data Dissemination Protocols

1. Introduction

Wireless sensor networks (WSN) represent a new and important type of data acquisition system. Throughout the *in situ* installation of hundreds or thousands of sensing devices, connected by a wireless communication framework, it is possible to monitor the physical environment even in hostile places. In general, WSNs are composed of a set of sensor nodes and sink nodes. Sensor nodes are responsible for collecting environmental data and sending it to sink nodes. Sink nodes or base-stations (BS) are responsible for gathering data from sensors and delivering it to the users.

Sensors are equipped with limited battery power. Therefore, energy saving is a crucial issue for WSNs. The most significant source of energy consumption is the data transmission. Thus, most protocols for WSNs try to minimize the data transmission, performing filtering and aggregation procedures as close of the data source as possible, in order to minimize the size and number of messages to be sent. Also to save energy, the short range, hop-by-hop communication is preferred instead of the long-range communication, directly from sensor nodes to the BS (Heinzelman *et al*., 2000).

An important issue in WSN design is the node identification scheme. From WSN applications point of view, it is more useful to have nodes identified by the type of sensor device or by their geographical location. Therefore, several works (Estrin *et al*., 2000, Heidemann *et al*., 2001) suggested the use of data-centric naming schemes. Since WSN nodes are described by well-defined attributes, a *publish and subscribe* scheme can be used for querying and extracting data from the WSN. In such scheme, a sensor node performs a "*publish*" operation, which consists in publishing a set of attributes that describes the type of data sensed by the node. Applications query the WSN through "*subscribe*" operations that are performed in sink nodes. Such operations consist of advertising an interest, which is represented by a (sub) set of attributes published by sensor nodes. An interest message is disseminated throughout the WSN from sinks until it reaches sensor nodes with attributes that match the interest.

The present work proposes a new approach for data dissemination in WSNs, aiming to assist a wide range of applications in an energy efficient way. The system works on demand, using an activating policy based on a *publish* and *subscribe* scheme. Applications access the WSN by subscribing interests. Such interests trigger the sending of data either on a periodical basis, or in an event-driven fashion, depending on the application class. The proposed protocol, called CLIN (Cluster-based Interest-oriented), organizes the WSN in groups (clusters) of nodes and the interests are propagated after the cluster formation. To save energy, sensor-collected data is transmitted by using the short range, hop-by-hop communication among the leaders of clusters, until it reaches the sink. The inter-cluster communication is performed using the CSMA protocol. We propose a selective TDMA scheme for the intra-cluster communication in which a time slot is given only for sensors that are to be active in a cluster. Sensors located outside the target area or those considered redundant to meet an interest do not participate of the TDMA schedule and they can remain in a power save mode.

The remaining of this work is organized as follows. Section 2 presents the basic concepts used in our proposal. Section 3 presents related works. Section 4 contains the description of the proposed system. Section 5 describes the simulations performed to validate our proposal. Finally, Section 6 depicts some conclusions.

2. Wireless Sensor Networks (WSN)

A typical WSN is composed of two main physical components: sensor nodes and sink nodes. A sensor node can contain one or more specialized sensing devices. Furthermore, it can have routing and data aggregation capabilities. Sink nodes, or base-stations (BS) provide an interface through which external systems can access the information collected by the network. BSs are able to aggregate data, but they often do not have sensor devices. They are more powerful nodes, considering energy, processing and communication capabilities.

2.1 Data Dissemination Protocols

The basic goal of a WSN application is to gather environmental information through sensor nodes and send it to a BS for further processing and storing. Assuming that the energy consumption is directly proportional to the transverse distance, in WSNs the short-range hop-by-hop communication model is preferred over the long-range direct communication model. In the hop-by-hop communication, sensors closer to the BS usually have their energy resources depleted before other sensors in the WSN. A possible solution for evenly distributing the energy consumption is to adopt a policy in which nodes have their responsibilities alternated between sensing and routing functions. Examples of well-known data dissemination protocols for WSNs are Directed Diffusion (DD) (Estrin *et al*., 2000) and LEACH (Heinzelman *et al*., 2000).

2.2 Cluster-based Protocols

An energy efficient way for organizing the network is to create groups of nodes (Heinzelman *et al*., 2000). These groups are called clusters and are created based on the geographical proximity of nodes. Each cluster has a leader, called cluster-head (CH). After gathering data sent by all its cluster members, the CH can perform processing in such data (a fusion, for example) and send it to the BS. The communication inside the cluster is accomplished by adopting a TDMA schedule, which is defined by the CH and is broadcasted to all cluster members.

One reason for adopting clustering is that the distance among cluster members and the respective CH is, in general, smaller than the distance between these sensors and the BS. Therefore, sensors in a cluster save energy since they transmit only to the CH and this, in its turn, forwards the information to the BS. Such forwarding to the BS results in a high-energy consumption by cluster-heads. A solution for this problem is to make a periodic rotation of the leader among the sensors in a cluster (Heinzelman *et al*., 2000). This solution provides all sensors a chance to save energy when they are not acting as a CH. An example of a cluster-based data dissemination protocol is LEACH.

3. Related Works

Although there is an expressive number of previous works dedicated to mobile ad-hoc networks (Ogunshola, 2002), most approaches are not suitable for WSNs, as they do not consider their specific constraints like the scarce energy sources.

Recently, a large number of data dissemination protocols specifically designed to WSN has been proposed, such as Directed Diffusion (Estrin *et al*., 2000) and LEACH (Heinzelman *et al*., 2000). Directed Diffusion (DD) is an efficient data dissemination protocol for WSNs. Individual nodes reduce the sampled waveform generated by a target into a relatively coarse-grained "event" description. Such description contains a set of attributes. Applications requesting data send interests through some BS in the network. These interests are also represented as a set of attributes and they are distributed on the network through a flooding method. If the attributes of the data generated by source nodes match these interests, a gradient is setup within the network and data is pulled toward the BS. Therefore, it is essentially a reactive, receiver-initiated and interest-oriented routing protocol. In reactive WSNs, nodes react immediately to drastic changes in the value of a sensed attribute. DD assumes a flat topology for the WSN. Thus, all nodes participate in the interest dissemination, even if the target area possibly comprises only a sub-set of nodes. CLIN adopts an interest-oriented approach but it assumes a hierarchical topology aiming to send interests and activate sensors only in the sub-region of interest, thereby increasing the network lifetime.

LEACH is a protocol suitable for proactive or periodical applications, which assumes the network has always data to send. In this protocol, the process of cluster formation is distributed among sensors. Any sensor may become a leader with probability X. The CH generates a TDMA schedule and transmits it in broadcast for all sensors in the cluster. After some time, each sensor enters again in the phase of cluster formation and this cycle is repeated until the complete energy depletion of all WSN sensors. An event-driven approach for LEACH is presented in (Cheng *et al*., 2002), but its main goal is to address the mobility of nodes, not the intrinsic features of reactive applications. If the application needs the data only in the occurrence of specific events of interest or if the target geographical area does not comprise the whole area of the network, LEACH-like protocols spend an unnecessary amount of energy. CLIN adopts the cluster-based approach and takes advantage of several features of LEACH, however it tries to embrace a wider range of target applications and to minimize the wasted energy.

TTDD (Luo *et al*., 2003) adopts a two-level hierarchy to disseminate data in a WSN. It uses a grid structure so that only sensors located at grid points need to acquire the forwarding information. Each data source proactively builds the grid and sets up the forwarding information at the sensors closest to grid points (called dissemination nodes). An analogy can be done between CLIN and TTDD, in the sense of both protocol organize the sensor network in groups of nodes, called cells in TTDD. Dissemination nodes are analogous to cluster-heads. A main drawback of the TTDD is the overhead for the grid maintenance. Grids are built for each data source in the network, even if there is no active interest. Therefore, TTDD is not well suitable for monitoring an environment where few events occur with high frequency but the application interests are sporadic.

4. System Description

Our work proposes a hybrid, interest-oriented and cluster-based solution to the data dissemination in WSNs. In the hybrid approach, WSN can work both as proactive and reactive. Interests that represent user queries are sent to the WSN, activating the data sensing and transmission. CLIN does not assume that the WSN always has data to send. Instead, the system works on demand, based on the user interest. The proposal does not require the direct communication between CHs and the BS. To save energy, the data is transmitted by using the hop-by-hop communication among the CHs towards the BS. CLIN adopts the same mechanism of the LEACH-C (Heinzelman, 2000) for cluster formation, which is centrally performed by the BS. CLIN also includes a distributed stage for CHs rotation that is similar to LEACH (Heinzelman *et al*., 2000) and has the goal of distributing the energy consumption evenly among the sensor nodes.

4.1 Cluster Formation and Generation of the Multi-Hop Topology

Let us consider the deployment of a WSN for habitat monitoring in a remote place. After the deployment of sensors in the target area, a configuration phase takes place, in which all sensors send to the BS a set of information, including the node geographical location (lat. and long.), its residual energy, and its types of sensing devices. The geographical information can be obtained by nodes using a triangulation method (Nasipuri and Li) to avoid GPS utilization in every node.

The BS runs the Simulated Annealing algorithm (Callum, 1990) to define the network clustering organization. The next step is the formation of the multi-hop topology, aiming to guarantee that all clusters reach the BS. In that stage, CHs should be connected to the BS through a topology represented as a graph where CHs and the BS are the nodes and the edges are the logical connections among such devices. To decrease the number of edges while meeting the requirement that all nodes are connected to its potential neighbors, the Delaunay Triangulation algorithm was adopted (Aurenhammer, F., 1991). After this stage, the BS applies the Dijkstra algorithm in the generated graph. At the end of the execution, the BS has established an optimal multi-hop topology for the network. This graph generation process is very scalable since both algorithms used are linear-time with complexity O(n * log n) (Efrat *et al*, 1998). This is an important point since during this stage all nodes are active and isolated among them, thus consuming energy.

The next step of the protocol is to send the generated structural information to the nodes. This is accomplished by transmitting a broadcast message in high power so that all nodes are able to "listen" to the message. The energy spent in such transmission is not a problem as the BS is not energy constraint. The message contains the network structure and it is formatted in such way that the receiving nodes are capable, after a small processing, to discover if they are common nodes or cluster-heads. A common node discovers who is its CH. A CH is able, to enumerate (i) its group of child nodes, (ii) the next CH in its route towards the BS and (iii) the CH(s) to which itself is the next hop towards the BS. A topology is created in which a given sensor always sends information to its CH. The CH forwards its data directly to the BS or, if it is a distant CH, forwards data to its next hop (a CH closer to the BS). Such forwarding follows until the information reaches the BS.

4.2 Periods of Cluster Attention

In the next stage of the initialization phase, each CH sends a message to its respective cluster members. In such message, the CH tasks its child sensors to turn off and periodically wake up, in pre-defined time intervals, remaining wakened up during a short time (called *period of cluster attention*). Periods of attention provide sensors a time for receiving "orders" from their CHs. This stage is called *economic wait* and its goal is to save energy by keeping nodes in a power save mode, while being able to eventually receive orders from the CH.

4.3 TDMA Scheduling and Interest Reception

During the stage of economic wait, an application eventually performs a subscribe operation, which consists of sending an interest to the BS. When receiving an interest, the BS computes the distance between itself and the most distant CH that meets the interest. The BS forwards the interest by sending a message with power enough to reach all target CHs. The BS keeps a control based on ACKS to guarantee that all CHs indeed received the interest message. When a CH receives an interest message, it verifies in its table of child sensors if some of them match to the interest. Thus, the CH waits the next period of cluster attention and broadcasts a message containing a list with the sensors that match the received interest and the respective slots of time in which each sensor must send its data (TDMA schedule).

When receiving such message, each sensor checks if its own ID is in the list and, if so, the sensor leaves the period of cluster attention. From this moment, the sensor should start sensing the environment and sending its data to the CH.

Transmissions inside clusters are accomplished by using a spreading code (CDMA (Tanenbaum, 2003)) defined by the CH. This code is different from those used in neighboring clusters to avoid inter-cluster interference. Transmissions among CHs are accomplished using CSMA through an auxiliary CDMA channel avoiding intra-cluster interference.

5. System Evaluation

To evaluate the performance of CLIN, we performed simulations by using the network simulator version 2 (NS-2) and the extensions developed by the MIT´s uamps project (Heinzelman *et al.*). The evaluation process comprised two different stages. In the first stage, we compared CLIN protocol to the "Directed Diffusion" protocol (Estrin *et al*., 2000). The goal of this stage was to evaluate if CLIN is an energy efficient solution for WSNs that adopts an interest-oriented approach. In the second stage, we compared CLIN to LEACH to evaluate CLIN in WSNs that adopt a periodical approach. In both stages, our interest was to evaluate CLIN in terms of energy efficiency. An important parameter in both evaluation stages is the interval between CH rotations. Through simulations, we found out that the ideal rotation interval value to CLIN should be between 10 and 12s.

5.1 CLIN x Directed Diffusion

The first metric considered in the performed simulations is the average delay of the information delivery. This metric means the average amount of time that information takes to leave the sensor nodes and to be delivered to the BS. The second metric is the average dissipated energy and it measures the average amount of energy dissipated to deliver each piece of information to the BS. This metric is directly related to the WSN lifetime.

Initially, the simulation rounds were executed in a 160m x 160m square environment with 50 randomly distributed nodes throughout the observation field. To evaluate the protocol with larger networks, we created scenarios with 100,150, and 200 nodes having the total area of the field proportionally increased according to the number of nodes. The value of 5% of the total number of sensors is adopted as the number of clusters to be generated (Heinzelman, W. *et al*, 2000). The rate of event generation was 2 events per second so that all sensor nodes, when powered on, detect two distinct events per second. In order to create a similar scenario to the one used in Directed Diffusion, we submitted an interest to the BS capable of activating nodes that cover an area about 43% of the total field. The adopted energy dissipation model is the same as described in (Estrin *et al*., 2000), in which a node spends 35mW in the idle periods, 395mW while receiving information through its radio circuitry and 660mW while transmitting.

The graph in Figure 1 shows the average dissipated energy over 5 network sizes, comparing CLIN to the Directed Diffusion. The error bars shown in the figure represent a confidence interval of 95%. According to this graph, CLIN can save up to 80% of the dissipated energy when using DD. We can associate this improvement to the different routing methods adopted by both protocols. CLIN only sends interest information to those nodes that are located around the interest region thus, saving energy that would be spent by the DD protocol in which interests are flooded throughout the network. It is worth noting that with CLIN, nodes that are distant from the target area are kept in a power-save mode.

A higher average delay is experienced when comparing CLIN to the Directed Diffusion protocol. Such fact can be explained by the use of a TDMA scheme in the communication between CHs. The adoption of TDMA forces each CH to wait until its time slot before transmitting. However, a new version of the protocol, which adopts an auxiliary channel for the communication between CHS, is under development. In addition, we used only one sink node (BS) in our simulations, while in DD the simulations were accomplished over networks containing 5 sink nodes. It is easy to realize that having only one sink node, all the information collected must converge to the same exit point, creating a higher total average delay. Considering this, it is acceptable to assume that a lower total average delay may be achieved in CLIN if there would be more than one BS in the WSN.

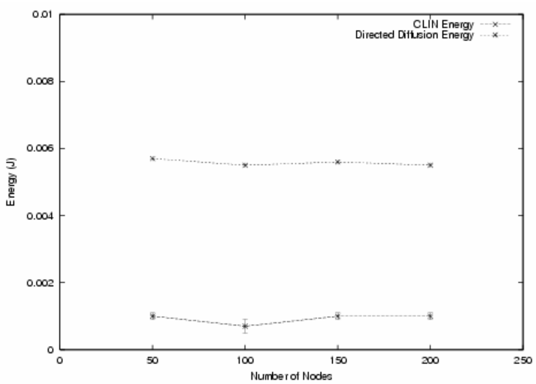

Figure 1 – Average dissipated energy – Number of Nodes x Energy (J) CLIN x Directed Diffusion

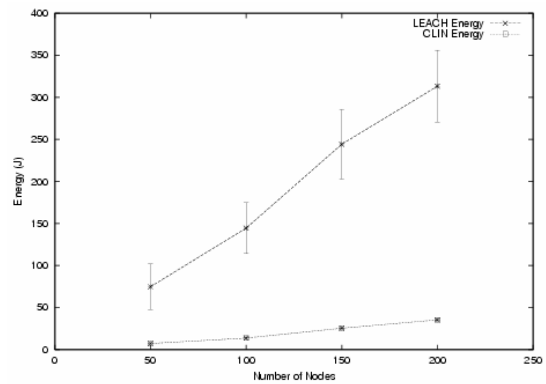

Figure 2 – Total dissipated energy – Number of Nodes X Energy (J) CLIN x LEACH.

5.2 CLIN X LEACH

The second evaluation compares CLIN to LEACH. The same network scenarios used on the previous simulation were applied. To compare CLIN to LEACH, we submitted to CLIN an interest capable of selecting all the network nodes for activation. After that, the simulations were run and the results are shown in Figure 2. This graph display the total energy (J) dissipation running each protocol in a 100s simulation interval.

The energy dissipation model used in this simulation was the same as in (Heinzelman *et al*., 2000). In this comparison, we realized that the WSN nodes dissipated up to 5 times less energy when running CLIN than when running LEACH. This gain can be explained by considering that with CLIN, CHs send information to their neighbor CHs instead of sending it directly to the BS, as done in LEACH. It is important to note that the scenario adopted in this simulation does not represent an optimal situation for CLIN since the whole WSN is active. We used this scenario to evaluate the specific energy gains by using multi-hop communication among CHs instead of directly to the BS.

6. Conclusions

This paper presents a new data dissemination protocol for WSNs, called CLIN. The main goal of CLIN is to attain a wide range of WSN application classes with energy efficiency. To provide such flexibility, an interest-oriented approach is adopted. According to the target application, an interest is defined to trigger the data sending either in a periodical fashion or in an event-driven way. The WSN becomes a flexible and powerful system for extracting environmental information, in which the data gathering and delivering are driven by the current user needs.

The adoption of a cluster-based approach organizes the WSN in sub-areas, in order to activate only sensors that are localized in the target region. With such approach, the communication occurs only when there is the explicit need of monitoring a given phenomenon. Sensors located out of the target area are able to save their scarce resources. The network cycle of duty is configured such that a minimum set of nodes is on active mode, in order to keep the network connectivity and to meet the application needs. In addition, a hierarchy of CHs was adopted to overcome two drawbacks in WSNs. First, it is not a realistic assumption that all nodes in a WSN (potentially CHs) are able to reach the BS. Second, in very large networks, the direct communication between distant CHs and the BS spends a lot of energy.

Simulations were carried out to demonstrate that the proposed protocol has reached its goal of providing an energy efficient solution for data dissemination in WSNs. CLIN performed well when compared to Directed Diffusion and LEACH protocols. Future works will be developed to demonstrate the flexibility of CLIN behavior in face of different classes of applications, in order to test the protocol response to different user demands.

References

Aurenhammer, F. (1991), "Voronoi Diagrams - A Survey Of A Fundamental Geometric Data Structure", *ACM Computing Surveys* 23, pp.345-405.

Callum, B. (1990) "*Blind deconvolution by simulated annealing*". Optics Communications, 75(2), pp.101-105.

Cheng, Z., Perillo, M., Tavli, B., Heinzelman, W., Tilak, S. and N. Abu-Ghazaleh (2002), "Protocols for Local Data Delivery in Wireless Microsensor Networks," *45th IEEE Midwest Symp. on Circuits and Systems* (MWSCAS '02), Tulsa, OK, Aug. 2002.

Estrin, D., Intanagonwiwat, C., Govindan, R. (2000), "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks". *Procs. of the Sixth Annual International Conference on Mobile Computing and Networking*, Boston, Massachussetts.

Efrat, A. *et al*. (1998), "Fly Cheaply: On the Minimum Fuel-Consumption Problem", *Procs. of the Symposium on Computational Geometry*, pp.143-145

Heidemann, J. *et al*.(2001), "Building Efficient Wireless Sensor Networks with Low-Level Naming". *Procs. of the Symposium on Operating Systems Principles*, pp. 146-159, Alberta, Canada, ACM.

Heinzelman, W. (2000), "Application specific protocol architectures for wireless networks", PhD thesis, MIT. Available in: http://citeseer.nj.nec.com/context/ 1674064/431346.

Heinzelman, W., Chandrakasan, A., Balakrishnan, H. (2000), "Energy Efficient Communication Protocol For Wireless Microsensor NetWorks ". *Procs. of the 33rd HICSS*. Vol. 8, Jan. 2000.

Heinzelman, W., Chandrakasan, A., Balakrishnan, H., The MIT uAMPS ns Code Extensions. Available in: http://www-mtl.mit.edu/research/icsystems/uamps/leach/leach_code.html.

Luo, H. *et al* (2003), "TTDD: A Two-tier Data Dissemination Model for Large-scale Wireless Sensor Networks". *Procs. of ACM MONET2003*.

Nasipuri, A. and Li, K., "A Directionality based Location Discovery Scheme for Wireless Sensor Networks". Available in http://citeseer.nj.nec.com/552224.html.

NS-2 (The Network Simulator version 2). Available in: http://www.isi.edu/nsnam/ns/.

Ogunshola, A. "Mobile Ad Hoc Networks: Survey and Criticisms". Available in http://www.columbia.edu/itc/ee/e6951/2002spring/Projects/CVN/report11.pdf

Tanenbaum, A.S. (2003), *Computer Networks, Fourth Edition*. Prentice Hall, ISBN: 0130661023.