Context-aware cluster-based hierarchical protocol for Wireless Sensor Networks

Md Enamul Haque*, Noriko Matsumoto and Norihiko Yoshida

Graduate School of Science and Engineering,
Saitama University, Saitama, Japan
E-mail: enamul@ss.ics.saitama-u.ac.jp
E-mail: noriko@mail.saitama-u.ac.jp
E-mail: yoshida@mail.saitama-u.ac.jp
*Corresponding author

Abstract: Wireless Sensor Network (WSN) has been a key research topic for its multi-dimensional applications. Its applications require accurate information collecting as well as, longer service life. In WSN, employing an efficient routing protocol plays a significant role in attaining such service requirements. In this paper, an energy efficient routing protocol, Context-Aware Clustering Hierarchy (CACH) has been proposed where cluster formation is entirely based on the context of the environment. Moreover, an efficient technique has been utilised to avoid similar data traffic across the network and cluster head role has been equally distributed among the nodes. The performance in the simulation shows substantial amount of energy optimisation.

Keywords: wireless sensor networks; hierarchical routing protocol; context-awareness.


Biographical notes: Md Enamul Haque is currently a PhD student of the Graduate School of Science and Engineering at Saitama University, Japan. His research interests include wireless sensor network, wireless ad hoc network.

Noriko Matsumoto is an Assistant Professor of Computer Systems in Saitama University, Japan. She received her MEng in 2004 from Nagasaki University, Japan, and DEng in 2006 from Saitama University, Japan. Her research interests include adaptive networks.

Norihiko Yoshida is a Professor of Computer Systems in Saitama University, Japan. He completed his Doctoral Degree in 1990 from Kyushu University, Japan. From 1993 to 1994, he was a visiting scholar in Stanford University, USA. His research interests include system design methodologies, and parallel and distributed computation.

1 Introduction

Application of WSN has been expanded from industrial operation to daily common use. The key elements of a WSN are typically small-size nodes, which have some inherent constraints such as limited energy, less communication and computation capability (Chandrakasan et al., 1999; Clare et al., 1999). Along with the conventional wireless network research themes like medium access control, routing, bandwidth, security etc., energy efficiency has been a crucial focus of research on WSN for a while. There are different research perspectives and trade-off to prolong the network lifetime. For example, energy efficiency can be obtained through circuitry enhancements of sensor nodes or optimising operational activity like routing, or through load balancing.

Lifetime of a WSN can be defined as the elapsed time between the activation of the first sensor and the deactivation of the last node due to energy depletion. Duration of lifetime of nodes has some critical significance in some application. For example, sensors deployed inside a volcano for monitoring must be active without any maintenance for a long time. The energy consuming operations of sensor nodes are usually sending and receiving data, processing query request, forwarding data to neighbours and finally sensing the environment (Younis and Fahmy, 2004). At the same time there are some activities, which actually lead to wastage of energy such as idle listening to the media, retransmitting packets due to collision, handling of too many control packets. A great deal of research effort has been invested to reduce such
activities as much as possible and optimise the regular operations of WSN. Considering these phenomena, it is obvious that energy efficiency can be greatly achieved if a well-constructed energy-aware routing protocol is being used in WSN.

There have been a good number of state-of-the-art routing protocols proposed (Al-Karaki and Kamal, 2004) for WSN. Overall, these routing protocols can be classified into three categories based on underlying network structure: flat, hierarchical and location based. Among such categories, hierarchical or cluster-based protocol technique is originally derived from the wired network to wireless network because of its scalability and efficient communication. In a hierarchical structure, higher residual energy nodes are assigned with the task of a cluster head, which is processing and long-range communication and other nodes are used to sense the environmental phenomenon. Formation of such hierarchical structure significantly increases the overall system scalability, system life and energy efficiency. Furthermore, the cluster head performs data aggregation, which decreases the number of packets transmitted to the base station. The activities of such protocols can be layered into two phases – first selection of cluster heads and second, the routing.

Considering different aspects of hierarchical routing technique, many protocols have been proposed (Heinzelman et al., 2000; Lindsey and Raghavendra, 2002; Manjeshwar and Agarwal, 2002; Li et al., 2001). Some of these produce significant improvement compared with the previously proposed protocols. But until recently, a vital perspective has been disregarded in the design of these routing protocols – i.e., context of the environment or sensed data. In the previous cases, the cluster head selection is done solely on sensor characteristics or some statistical assumption or even randomly. Considering the sensed data, for example, temperature or humidity value, it can be understood that such values naturally create some clusters of the environment. Taking into account this tendency, a routing protocol’s clustering mechanism, cluster head selection and regular operation can be greatly improved. In some recent publications (Jin and Park, 2006; Zhou and Hou, 2007), new routing protocols are being proposed considering such events. In this paper, we proposed a context-aware routing protocol, CACH, which is based on the sensed data of the environment, more specifically the context of the environment. The cluster heads will be carefully selected to disseminate energy consumptions among the nodes. Moreover, packet contains the same value as previous are not sent, which reduces the traffic between cluster heads and nodes significantly.

The paper is organised as follows: in Section 2, problematic issues of current routing models are discussed with points of improvement. In Section 3, routing protocols related to CACH are discussed with their shortcomings. In Section 4, possible and expected improvements are mentioned. In Section 5, the CACH protocol is presented in detail. In Section 6, the simulation result of CACH is presented with comparison and Section 7 concludes the paper with future direction.

2 Problem statement

Characteristics of sensors in WSN have put some constrains to protocol designers. As a result, designers want to keep their objective as simple as possible such as keeping shortest paths between nodes, cluster head, base station and consuming less energy. One of the basic routing models is the one-hop model. But it is quite impractical for some reasons. For example, the distant nodes will be out of energy very quickly due to longer communication distance with the base station. Even if the sensors are nearer to the base station, in a WSN comprised of thousands of sensors, the network density will lead to frequent collision and ultimately degrade the network efficiency.

A number of protocols have been proposed based on the multi-hop model (Lindsey and Raghavendra, 2002; Sadagopan et al., 2003; Intanagonwiwat et al., 2003). This model has impressive advancement such as larger coverage area and less collision. It uses data aggregation technique to keep traffic low. But this design criterion also leads to drawbacks like high latency time especially in a network of good number of sensors. Targeted to the base station, data encounters serious delay in every hop due to processing. A more serious drawback is the self-induced black hole effect (Ibriq and Mahgoub, 2004). Nodes adjacent to the base station must always forward data to the base station as intermediaries. As a result of such repetitive actions, these nodes are the first in the network to become out of energy creating a black hole around the base station for incoming traffic.

The cluster-based hierarchical model offers significant improvement despite some critical issues. Developing a protocol in this category must consider the following issues. First, the cluster boundary needs to be meticulously computed to keep the distance between the base station and the node minimal. Again, such accurate computation will keep clusters non-overlapping decreasing the possibility of collision and intra cluster interference. Second, the role of the cluster head must be effectively distributed among the nodes so that the energy consumption will be shared by all nodes. Thus, it is expected that the service duration of the nodes will be almost equal. Third, in most of the protocols nodes are organised according to the mutual distance and residual energy. Thus process of data collection is not related to the environmental phenomenon. However, in some major applications of WSN, data gathered from adjacent sensors are similar. This trend is needed to be properly utilised for node’s role selection. Fourth, the phenomenon of the environment usually changes quite steadily by the law of nature. However, most protocols are unaware of this fact and sending the sensed data in a continuous manner. Such natural phenomenon should be taken into consideration. Finally, the lifetime of a node must be effectively exploited by maximising the sleep time between radio communications, because it has been observed that passively listening to the radio signal consumes significant amount of energy. There are also some other issues like heterogeneity of nodes, data security that can be addressed as the secondary level of requirement.
3 Related studies

Among the hierarchical protocols, some protocols have the ability to identify the cluster heads with the help of a central controller during the node deployment (Tillet et al., 2002). In some other cases (Chen et al., 2003), the cluster heads are being chosen by the network planner at the time of actual physical deployment. Moreover, prior location information has been utilised in some protocols (Kalpakis et al., 2002; Dasgupta et al., 2003). Although these options increase the efficiency of network operation, they do not match the characteristics of a typical WSN.

Low-Energy Adaptive Clustering Hierarchy (LEACH) (Heinzelman et al., 2000) has been considered as the benchmark for many protocols proposed later. It has some distinctive characteristics like self-reconfiguration, adjustment of communication range according to distance, schedule of data transmission of individual nodes, etc. It has some assumptions like fixed-base station location; energy constrained homogeneous nodes and predetermined ratio of cluster heads among all nodes. The operation of LEACH is separated into a series of equal length time spans. In each of these time spans, cluster head selection, the cluster formation and scheduling procedures are completed, respectively, at the very beginning. Cluster heads are selected based on a probabilistic value satisfying the condition that those nodes have not played that role previously. Upon receiving broadcasted advertisement messages from a single or multiple cluster heads, a node sends joining declaration to the nearest cluster head. The cluster head then create a TDMA schedule and notify its member nodes. The following data transmission phase has the larger chunk of each span, which is also divided into a number of equal frames. Despite of significant advantages, LEACH also has to deal with some drawbacks like probability of selected cluster heads falling on a small geographical portion of the total area due to the dependency of probabilistic value. To overcome these, LEACH-C (Heinzelman et al., 2002) was proposed imposing centralised control. Nevertheless, none of these two versions is concerned about the context of the environment.

In the Information Technology paradigm, the term ‘context-awareness’ was first introduced by Schilit et al. (1994). But it has not been utilised on the WSN research area until recently. In some works, the ‘context-aware’ concept has been utilised as a technique for energy optimisation. Chong et al. (2005) have proposed a software-based framework to process the surrounding contextual data and trigger power saving functionalities, such as adjustment of different sensing criteria. Discovery of appropriate context and use of such context as trigger are termed as ‘context discovery’ and ‘context trigger’, respectively. To support these activities, two centralised databases are used which are located at the base station end. The centralised method of observation of this frame work may inflict the claimed improvement of using the context-based approach. Moreover, ‘off the site’ context processing limits its applicability to simple, non-real time sensing only.

A context-aware clustering has been proposed for efficient data aggregation by Jin and Park (2006). Authors have proposed a clustering mechanism that strives to form clusters of sensors with similar output data within the bound of given tolerance parameter. The operational objective behind this is to use a simple data aggregation technique without introducing large errors. Thus, the expected benefit is to reduce energy consumption and to prolong network service life. This algorithm has shown impressive performance on the environment where the change of the surrounding context is quite gradual. On the contrary, the cluster head role has not been distributed, which might cause energy exhaustion of some nodes early. Again, there is no distinctness between the set-up and data transmission phase. Besides, if the environmental phenomenon is changing rapidly, the algorithm might not be suitable. Thus, it limits this algorithm’s applicability in certain fields only.

In another context-aware publication (Wu et al., 2008), authors have proposed a protocol, which organises the sensors based on both distance information and data distribution. A multi-granularity query processing method has also been included that obtains more accuracy and efficiency in querying compared with a random access system. However, a closer look at this protocol reveals some unattended issues. In the selection of high residual cluster heads, the energy information must be shared among every node. Without having an active routing protocol, dissemination of such information is quite impossible unless being broadcasted. Broadcasting this information might significantly increase the traffic and ultimately will decrease the energy efficiency. Moreover, the multilevel clustering hierarchy is not optimised in selecting cluster heads. As a result, in certain situations the cluster head role might be played by the same node in multiple levels, which may cause faster depletion of the residual energy.

4 Expected improvements

The main objective of designing the proposed protocol is to attain maximum energy efficiency. On that effort, a number of key issues of WSN has been identified and attempted to be resolved. Such vital improvements are illustrated here:

Distributed cluster formation is considered as an important feature of WSN. In our proposed protocol, the cluster formation is fully context-aware and distributed. Because of the law of nature, environmental phenomenon tends to be similar in different small geographical areas. Thus, when cluster are formed based on the context, obviously cluster regions become non-overlapping and disjoint. Moreover, there are no possibilities of selected cluster head falling into a smaller part of the whole coverage area. Thus cluster heads are positioned in a distributed manner throughout the deployed area.

When a node becomes a cluster head, there is an increase of energy consumption due to long-range communication with base station and computation of
data packets. Thus, playing such role can exhaust the residual energy of a node pretty quickly. To cope with this situation, the cluster head role is rotated among the member nodes per every round in our proposed protocol. At the beginning of a new round, the cluster head checks whether all of its member nodes have become cluster head before or not. If so, it just informs them the equal opportunity to become cluster head in next round. So nodes will be playing cluster head role equal number of times and no node will be out of energy sooner than the others. In this way, the proposed protocol offers better quality service and longer network operation.

On the steady data transmission phase, it is fairly common that environmental phenomenon can be unchanged within certain slot time or even within round time. But conventional protocols will send data packets to the cluster heads as usual. This trend has been carefully utilised in our proposed protocol. Rather than sending data packets blindly, every node verifies the data changes. If there is no change in content compared with the previous data, it just stops sending data packets until further content changes. The cluster head continues to send aggregated data to base station without the data from that node. However, the base station always checks any such lapse and if so, it proceeds with the previous data. Such technique reduces data traffic significantly and ultimately contributes on significant energy efficiency of the total network.

5 Context-aware cluster-based hierarchy

This section describes the prior assumptions, algorithm of the proposed protocol and its operation criteria.

5.1 Assumptions

In this proposition, we have assumed some of the parameters of the protocol and characteristics of the environment as a priori.

- Presence of the regional property of the environmental data. This means that there are distinguishable differences among the data of every region (cluster). More specifically, data of a single region are similar and different than other regions.
- Predetermined ratio of the cluster heads among all nodes. For example, 5% of the total nodes will be cluster head at any time. For such feature, data within every cluster will have a certain range, which will be computed from this ratio of the cluster heads.
- Nodes are expected to be quasi-stationary, which is a characteristic of a typical WSN. However, smooth movement will not hamper the network operation at all.

5.2 Algorithm

In the proposed protocol, the lifetime of the network is divided into a series of equal length time spans called round, as shown in Figure 1. Within every round, the initial portion is allocated for the set-up phase. In this phase, the cluster head selection, member listing and schedule dissemination procedures are completed. The later and longer portion is reserved for the steady operation phase in which data are transmitted from member node to cluster heads and from cluster heads to base station. This phase has also been divided into some equal length frames. Again, each of these frames is divided into a number of slots where a single separate slot is reserved for every member node. Thus, every node of a certain cluster will have a separate time slot to send data to the cluster head and these slots are mutually exclusive within the cluster. The details of the proposed algorithm are described here in a number of steps:

- **Step 1:** After the initial deployment, nodes are activated in asynchronous manner and enter into the set-up phase. At this point, there are no cluster boundaries nor any node’s role is been defined. Each node senses the context of the environment and checks whether any cluster head advertisement with same sensed value is received or not. If not, it broadcasts a message containing its address and the sensed values as a cluster head announcement. Upon sending the cluster head announcement, the node waits for join request message. On the other hand, if a cluster head advertisement with same sensed value is received, the node classifies its role as a member node and selects the advertising node as a cluster head. Then it sends a join request message to that cluster head. In some rare cases, the cluster head might not receive any join request message. In such situation, the node forms the cluster with only one node that is itself. However, with a network of hundreds of sensors, such case is supposed to be quite rare. In this manner, clusters are formed with non-overlapping boundaries based on the context of the environment. Figure 2 shows the flow diagram of this context-aware cluster formation in the set-up phase.

- **Step 2:** After receiving joint request messages within certain time period, the cluster head proceeds with creating a TDMA schedule for every member node. Then, this schedule is sent to the member nodes in a unidirectional manner. Upon receiving such schedule, every member node sets its own slot time in the frame for this current round. However, this schedule is valid for this round only.
After every round, this schedule is being created by the cluster head without any knowledge of the previous one. These two steps actually constitute the set-up phase of the protocol.

- **Step 3**: Now the network enters into the steady operation phase. Each node sends its own data to the cluster head according to the slot time. But this time, the transmission range has been reduced based on the signal strength received in set-up phase from cluster head. Each member node also stores their data temporarily. In the consecutive slots, upon sensing the environment, the node compares its current sensed value with the previous one. If there is no change in value, the node does not send the data. When cluster head receives data packets from its member nodes, it aggregates those packets into a single packet in sequential style. Thus, when a member node does not send data within designated slot time; the cluster head just ignores it and aggregates the remaining packets. This aggregated packet is sent to the base station directly. On this end, base station keeps track of each member node of every cluster. So, an aggregated packet is being extracted and checked for any missing member node data. If so, the base station assumes the current data content is same as the previous one and continues its operation with the previous data. As the base station has virtually unlimited energy, faster computability and more storage, storing the temporary data will not decline the network performance.

Figure 2  Distributed cluster formation

After finishing Step 3, nodes restart the cycle by executing the Step 1 again. In the later rounds, whenever nodes enter into the set-up phase in Step 1, a node checks whether it has been cluster head in any previous round or not. If so, it just waits for cluster head announcement. Moreover, upon forming a cluster, a cluster head checks every member nodes role history. If all nodes within a cluster have already been a cluster head in previous rounds, this role history is been erased. Thus from next round, every node have the possibility to become a cluster head just like the initial stage. This technique ensures the cluster head role rotation.

In a publication of Heinzelman et al. (2000), communication interference between clusters has been handled using different CDMA code by each cluster. Each cluster head select a random CDMA code during set-up phase, which is informed to the member nodes also. Thus, communication interference between clusters can be avoided. Our protocol also utilises this technique to filter neighbour’s signal and reduce the possibility of corrupted signal within the cluster.

### 6 Performance evaluation

To check the performance of CACH, a simulation environment has been created to monitor the temperature of the surrounding. The simulation is implemented on the sensor network extension of J-Sim (Miller et al., 1997). For performance comparison, LEACH protocol has been selected, which is treated as the benchmark in many related publications. As the energy efficiency is one of the key objectives, energy consumption is meticulously measured and analysed by generating necessary graphs.

#### 6.1 Network and radio model

The network and radio models used in the simulation are similar to the model discussed in Al-Karaki and Kamal (2004) Heinzelman et al. (2002). Total 100 sensor nodes are deployed in a random manner into a \(100 \times 100\) simulation area. All the nodes are of same type and each node starts with 0.25 unit energy. These nodes have the mobility capability and some nodes can randomly move. The base station is located in (0, 0) location and assumed to be have virtually unlimited energy and longer communication capability. The context has been created by target nodes, which in this case are broadcasting the temperature value. This temperature value is changing as time passes. However, this changing rate has been re-adjusted to create different environmental circumstances. The radio model used here is a simple one. Nodes of 914 megahertz frequency have the bandwidth of 1 megabyte. Energy consumption is calculated using Friis equation (Friis, 1946) based on the distance between transmitter and receiver.

#### 6.2 Simulation results

In Figure 3, the energy consumption of CACH and LEACH protocols have been compared with each other with respect to the environmental data change interval. In a single simulation, after every certain period of time (e.g., 15 s) temperature value is changed and the total system energy consumption has been measured at the end. This period of time is gradually increased in consecutive simulations and
the system energy consumption is measured, respectively, for these two protocols. In this way, the data for this graph has been accumulated. For CACH, the energy consumption commences from 17 units for very frequent environmental data change interval and later sharply drops down to around 7 units when this interval is 15 s. With further increment of this interval, the energy consumption steadily decreased up to over 6 units. For LEACH, this data change interval has almost no effect on energy consumption and it remains on 15 units energy with negligible fluctuation. For the graph in Figure 4, similar pattern has been followed to gather the simulation data. This time, the total number of packets sent to the base station has been compared with respect to the environmental data change interval for both protocols. For CACH, total number of packets sent is just below 44000 for very frequent data change interval and it drops to over 8000 packets when the data change interval is almost 8 s. Total number of sent packets reaches to exactly 500 with 120 s interval. Again for LEACH, the total number of packets sent always remains near 42000 mark disregarding the data change interval. In the Figure 8, a bar chart shows the number of packets sent and dropped in LEACH and CACH in 2000 s simulation duration with environmental data change interval of 15 s. In LEACH, total 41676 packets are sent and 59 packets are dropped. Whereas, in CACH total 8160 packets are sent and only one is dropped. In Figures 5 and 6, energy depletion patterns for both CACH and LEACH have been portrayed. In both cases, the simulation has been continued for 2000 s. In the case of LEACH, the depletion pattern is a bit scattered. However, in the case of CACH, the energy depletion pattern is quite coherent and at 2000th s, most of the nodes in the system still have around 50% energy. In Figure 7, a comparison between CACH and LEACH for number of nodes remaining alive has been shown. For CACH, all of the 100 nodes are alive. However, for LEACH, the first node dies at around 1100th s and at the end of the simulation 63 nodes are out of service. Thus, almost one-third of the nodes are remaining alive.

Figure 3 Total system energy consumption of CACH vs. LEACH (see online version for colours)

Figure 4 Total number packet sent from all nodes (see online version for colours)

Figure 5 Energy dissipation in LEACH (see online version for colours)

Figure 6 Energy dissipation in CACH (see online version for colours)
6.3 Analysis and discussion

The key performance issue of CACH is energy efficiency achievement. As shown in Figure 3, CACH consumes almost half energy compared with LEACH when the data change interval is around 8 s. With further lengthy data change interval, the energy consumption is slightly reduced. Thus CACH clearly outperform LEACH in terms of energy efficiency when there is a steady environmental data change. Moreover, the technique followed in CACH for handling the same data as previous rounds reduces significant amount of traffic within the network system. Figure 4 shows that the total number of packets sent in CACH is five times less than LEACH when environmental data change interval is between 8 s and 120 s. This is one of the expected improvements and the result shows tremendous performance of CACH over LEACH.

In Figure 5, the energy dissipation pattern of LEACH reflects uneven consumption of energy by the nodes. Some nodes are out of energy quite early and remaining nodes are about to be out of service. After approximately 1100 s, the service quality starts degrading and ultimately all nodes are expected to be exhausted in around 2400th s. Conversely, CACH shows fur more equal energy consumption across the network. At 2000th s, most of the nodes of the network have 50% energy and still continuing to give full-scale service. The concurrent state of the network is clearly reflected in Figure 7 that shows number of active nodes as 37 and 100, respectively, in LEACH and CACH. Moreover, CACH shows reasonably low ratio of dropped packets according to Figure 8. This performance achievement is possible for CACH mainly because of its efficient cluster head role rotation and distributed cluster formation. All this performance measurement signifies the suitability of CACH in situation where the environmental context has discrete geographical phenomenon changing in a gradual manner.

7 Conclusions and future work

In this paper, a context-aware protocol, CACH has been proposed for WSN. This cluster-based protocol forms clusters according to the context of the geographical regions of the deployed area. Effective technique has been utilised to avoid similar data traffic between member nodes, cluster heads and base station. In consecutive cluster formations, cluster heads are selected in such a way that the cluster head role is rotated between the nodes of same type of context. As energy efficiency is one of the foremost requirements in many WSN applications, CACH aimed to resolve this issue without compromising other benefits. Performance comparison with LEACH protocol clearly point out the energy efficiency of CACH protocol in gradual changing environment, which is quite common in real world. Moreover, comparison shows reduced data traffic in CACH compared with LEACH, which ensures more idle time and longevity of the network.

The future work of this protocol includes the implementation of multi-hop communication between the cluster head and the base station, which can contribute more in energy efficiency. Moreover, for large number of sensor nodes, super-clustering might be an effective technique that is worth of further research work.

Acknowledgement

This research was supported in part by MEXT in Japan under Grants-in-Aid for Scientific Research on Priority Area 19024012, and by JSPS in Japan under Grants-in-Aid for Scientific Research (B) 20300024.
References


