Budget-Based Clustering with Context-awareness for Sensor Networks

Jiaxi You, Dominik Lieckfeldt, Matthias Handy, and Dirk Timmermann

Institute of Applied Microelectronics and Computer Engineering

University of Rostock, 18119 Rostock, Germany

{jiaxi.you|dominik.lieckfeldt|matthias.handy|dirk.timmermann}@uni-rostock.de

Abstract

As the scale of modern sensor networks continues to grow, energy consumption, scalability and routing efficiency are becoming key design challenges. Network management plays an important role in achieving these goals. By decomposing a sensor network into smaller groups, clustering and its variants have been presented as efficient ways in network management. In this paper, we propose a dynamic, localized clustering approach derived from generic budget-based clustering techniques. The approach generates dynamic cluster sizes for a hierarchy of cluster heads, with respect to network context such as residual energy and activity rates of sensor nodes. We further refine the local estimated cluster sizes by using additional feedback during clustering process. Simulation results of stochastic deployment are used to demonstrate the performance of our algorithm, as well as the impact of context information as clustering parameters.

1. Introduction

A Wireless Sensor Network (WSN) is typically composed of a number of sensor nodes, which are capable of sensing, signal processing and wireless communi-cation. Sensor nodes measure physical conditions from terrains of interest, like temperature, humidity, entering of object etc., and send gathered data to a data sink where information is processed. Nowadays, WSNs are becoming widely used in commercial and military applications such as environmental monitoring, target tracking, and system control. Recent advances of electronic technology yield extremely small and inexpensive sensor nodes [1] [2], and motivate development of even larger networks.

Due to the unattended nature of their deployment, sensor nodes are usually battery powered. Therefore, energy consumption remains critical for the lifetime of WSNs. One of the main components of power consumption is wireless communication that is usually carried out by on-board radio transmitters. During radio signal propagation the transmission power decreases proportionally to the square of the distance or worse. Therefore, transmission range and data rate are highly energy constrained in WSNs. Scalability is another major design attribute of WSNs. In a direct-connected network, the data sink becomes overloaded with increasing number of sensor nodes [3]. Furthermore, as the network covers larger areas, sensor nodes that are placed relatively far away are possible to lose their contact with data sink. In networks of large physical dimension, multi-hop communication is usually preferred to direct communication between sensor nodes and data sinks, which can also improve the scalability of WSNs.

By dividing a geographical region into a number of smaller zones [4], clustering techniques aid effectively to organize large-scale WSNs, as well as to prolong their lifetimes. A generic cluster has a cluster head and several cluster members. Sensor nodes in individual clusters transmit data only to the local cluster heads which aggregate received data and route them to data sink. Because of the relatively small size of clusters, cluster heads and associated cluster members are normally located near to each other, where intra-cluster communications are energy-efficient via single-hops [5]. Furthermore, since the cost of transmitting a bit is higher than processing a bit in WSNs [1], data aggregation and fusion by cluster heads can decrease the size of messages and save energy significantly [6] [3]. Finally, aggregated data is routed to data sink through an overlay of cluster heads [5], while leaving the cluster members released from the global routing processes.

For some delay sensitive WSN applications such as battlefield surveillance or forest fire detection, time for a detected event to arrive at data sink is of significant importance. In WSNs, efficient routing paths between sensor nodes and data sink can help to achieve this goal. Routing paths can be formed statically using specific network structure [7] [9], or are built dynamically on demand [6] [10] [11]. In a clustered WSN, careful selection of cluster heads and optimal formation of clusters help to decrease the end-to-end delay of the network.

Context is any information that can be used to characterize the situation of an entity [12]. In WSNs, context information mainly refers to power profiles of sensor nodes, their storage and processing capability, network topology, as well as communication traffic. In pervasive computing, user context and application context [13] are also utilized in management of WSNs [14] [15]. Among all kinds of context information, one important network attribute is the event distribution in the area, namely the sensing activity rates of sensor nodes. The amount and location of events correspond to the sensing activities of sensor nodes, as well as data transmission follows from that.

In this paper, we present a novel hierarchical clustering approach, which targets not only energy conservation and scalability of WSN architectures, but also efficient routing between sensor nodes and data sink. Our approach is derived from the idea of budget-based clustering [16] [17]. The idea of budget-based clustering is to decompose a WSN into clusters of bounded size. To the best of our knowledge, prior budget-based clustering algorithms consider only geometrical properties of sensor nodes during cluster head election and cluster formation. In contrast, our clustering idea considers various aspects of WSNs. Besides positions of sensor nodes, different kinds of context information are used as parameters for the clustering processes, such as energy levels of sensor nodes and event distribution in the network. In order to further balance the traffic flows, dynamic cluster sizes are estimated locally based on context information. However, we keep the message complexity of clustering low, since most of the clustering parameters are estimated locally on individual sensor nodes. In contrast to most of the existing clustering approaches, the proposed clustering algorithms yield a connected dominating set of cluster heads. The objective of this work is to form hierarchical WSN architectures by using intelligent clustering algorithms, with consideration to network load balance, energy conservation, as well as routing efficiency between sensor nodes and data sink.

2. Related work

Various clustering algorithms of WSNs have been proposed in recent years, most of them are developed in conjunction with hierarchical routing protocols. The objective of hierarchical routing protocols is to conserve energy by restricting most of communication within clusters [3]. Global routing of data packets is carried out by cluster heads. In order to perform the extra routing tasks, cluster heads are generally the nodes with less energy constraints than their cluster members.

The Low-Energy Adaptive Clustering Hierarchy (LEACH) [7] is one of the fundamental hierarchical routing protocols for WSNs. In LEACH, cluster heads are selected randomly. After the cluster heads broadcast advertisements to neighbor sensor nodes, clusters are formed according to the received signal strength on sensor nodes. Cluster members are able to transmit to local cluster heads directly, where data aggregation and global routing are performed. The clustering process is fully distributed and does not need global knowledge of network. Additionally, dynamic rotation of cluster heads is used to balance energy dissipation on sensor nodes. Handy et al. [8] proposed the eXtended LEACH (XLEACH) algorithm in order to highlight the extra energy consumption on cluster heads. XLEACH considers the residual energy level for cluster head selection, and increases the lifetimes of LEACH networks by 30 %.

Another way to form bounded clusters is to constrain cluster sizes with a predefined budget value. In budget-based clustering algorithms, selected cluster heads are assigned with budget values that indicate the maximal size of their clusters. During cluster formation, cluster heads distribute the budget values among their neighbors, which decrement and forward the received budget values until exhausted or no further growing is possible [16] [17]. Two distributed budget-based clustering algorithms are proposed in [16], where distributed cluster formations with a predefined size are performed. The control message overhead is decreased with randomized technique. Additionally, feedback mechanism is used to collect exceeding budget values, thus to build clusters closer to the desired size. The algorithm is further developed in [17], where the distribution of budget values is directed to avoid overlapping of clusters.

3. Budget-based clustering algorithms with context-awareness

In this section, we describe two novel budget-based clustering algorithms with context-awareness: the Clustering with Dynamic Budget (CDB) algorithm and the Interactive Clustering with Dynamic Budget (ICDB) algorithm. We also demonstrate the impact of various clustering parameters derived from context information on our clustering algorithms.

In our scenario, we assume a randomly deployed WSN, where sensor nodes of the same type are static or at least quasi-static after deployment. Since sensor nodes are normally left unattended, their energy level may be not uniform due to different activity rates. To ensure a general use of our algorithms, we also keep the sensor nodes as simple as possible. The sensor nodes only have one transmission power that defines their transmission range. As in common WSN applications, sensor nodes are named by their attributes, normally their positions.

As mentioned earlier, prior budget-based clustering algorithms use static budget values with the objective of forming clusters with a specific average size. Cluster heads distribute budget values blindly to their neighbors [16] or in certain geometrical directions to avoid overlapping of clusters [17]. Considering real cases, a static budget size can be inflexible due to the in general inhomogeneous deployments of WSNs. Sensor nodes in a network normally have different residual energy or activity rates [18], which contributes dramatically to the complexity of the network. Therefore, our algorithms utilize such context information as important parameters during clustering processes. For instance, cluster heads with more residual energy should run bigger clusters to improve load balancing. And sensor nodes with higher activity rates should be placed nearer (in terms of hops) to data sink, so minimal hops will be needed for frequent messages coming from the active nodes.

Both clustering algorithms are triggered by cross-layer parameters, such as start of new query batches, change of network topology, and break down of sensor nodes. Periodical re-clustering is eliminated to conserve even more energy.

3.1. Clustering with Dynamic Budget (CDB)

In CDB, we consider various kinds of context information as clustering parameters. Context information is used in estimation of dynamic budget values, election of cluster heads, as well as selection of cluster members. In this paper, we demonstrate the impact of two clustering parameters: the residual energy and activity rates of sensor nodes. During budget estimation, cluster heads with more residual energy are assigned bigger budget values. But the budget values will be reduced when the cluster heads have high active rates, since more energy will be needed for their own sensing events. The activity rates of neighboring sensor nodes are compared when a cluster head starts to create its own cluster. Sensor nodes with higher activity rates are selected first as cluster members, in order to place active nodes nearer (in terms of hops) to data sink. A cluster head stops the cluster formation when its budget value is exhausted or there are no more free sensor nodes in the range. Among new cluster members, sensor nodes with higher residual energy but lower activity rates are elected as cluster heads of next level in the spanning tree.

The generic process of the CDB algorithm is described in the following steps:

Step1 (Initialization): The data sink announces itself as the first initiator (node that initiates a cluster formation process). A reference budget value is estimated at data sink, with respect to the trade-off between the depth of network hierarchy and the average cluster size. A HELLO message is initialized including the reference budget value. Depending on the context information needed as clustering parameters, the HELLO messages are attached with requests of various aspects of sensor nodes, such as their positions, energy levels, activity rates, etc. Rules for dynamic budget value estimation, selection of cluster head and cluster member are also defined here, with context information as parameters. Clustering rules are broadcasted to all sensor nodes in the network.

Step2 (Message Exchange): The initiator broadcasts the HELLO message to all sensor nodes in its transmission range. Sensor nodes that receive the HELLO message send back their RESPONSE messages to the initiator, which include the requested context information.

Step3 (Cluster Formation): The initiator node collects context information from neighboring sensor nodes, estimates their priorities using the clustering rules provided, and select new cluster members to the amount of its budget value. Additionally, part of the new cluster members are elected to be cluster heads of the next level. A CONFIRM message with such information is sent to the new cluster members.

Step4 (Clustering Propagation): The newly elected cluster heads adjust the reference budget values according to their own context information, and loop back to Step2 (Message Exchange) until all sensor nodes are clustered or there is no further growing possible.

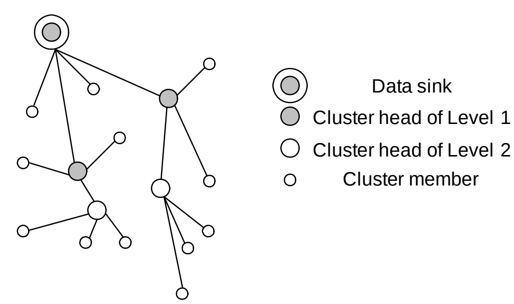

Figure 1. Network structure after clustering

Clustering processes of each level are assumed to be synchronized, so that sensor nodes of higher priority are placed nearer (in terms of hops) to the tree root. During cluster formation, multiple advertisements arriving simultaneously at a sensor node are also resolved with context comparison. In our example, cluster heads with higher residual energy or lower activity rates will win such competitions. In the end of the clustering process, the hierarchy of cluster heads forms a spanning tree structure (Figure 1), where the cluster heads being the tree nodes store a table of their cluster members for routing of queries or control messages. Gathered data on cluster heads can be routed to the data sink along with the tree structure.

3.2. Interactive Clustering with Dynamic Budget (ICDB)

For the proposed CDB algorithm, the size of each cluster is estimated locally on individual cluster heads. The lack of regional consideration may lead to poor load balance and decreased lifetimes of WSNs. Given two cluster heads with same budget values, the one with more active cluster members will suffer severer energy drain than the one with the same amount of, but less active cluster members. Therefore, context information of cluster members should be also considered during clustering.

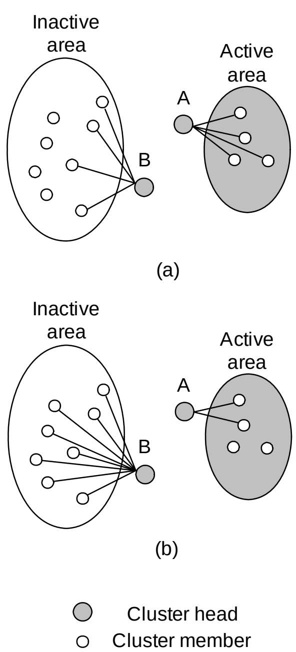

Figure 2. The feedback mechanism. Both cluster heads have budget value = 4. (a) Cluster formation with CDB, where all cluster members have the same occupation weight = 1. (b) Cluster formation with ICDB, where high active cluster members have occupation weight = 2, low active cluster members have occupation weight = 0.5.

In order to consider the impact of cluster members with different attributes, we propose the Interactive Clustering with Dynamic Budget (ICDB) algorithm as an extension to the CDB algorithm. During cluster formation, cluster heads request additional information from sensor nodes being clustered. The extra information is attached to the RESPONSE messages, so the total number of control messages remains the same. We demonstrate the idea using the activity rates of cluster members as feedback. However, various other types of context information can be also utilized.

The generic process of the ICDB algorithm is similar to CDB. In addition to the reference budget value, a reference activity rate is also estimated at data sink, and propagated to the cluster heads. During Step3 (Cluster Formation), cluster heads compare the activity rates of candidate cluster members with the reference activity rate and evaluate their occupation weights to the budget values. For instance, when a high active sensor node is selected as a new cluster member, the budget value will be decremented by 2, while a low active cluster member only occupies 0.5 (Figure 2). With such mechanism, individual cluster sizes can be further adjusted dynamically to enable better load balancing.

4. Simulation results

We simulate the proposed clustering algorithms with selected context information (residual energy and activity rate) as clustering parameters. The results are compared with a generic budget-based clustering algorithm adopted from the rapid algorithm [16], where hierarchical cluster heads (initiators) are selected randomly and assigned fixed budget values. The simulated network has a dimension of 200 m × 200 m, where 500 sensor nodes and data sink are randomly deployed on the square plane. Each sensor node is capable of sensing and wireless transmission, and can be assigned with the role of either cluster head or cluster member. The transmission range of sensor nodes is fixed to 50 m. In order to provide enough candidate nodes to choose from, our network deployment is relatively dense. We use a simple disc-communication model: sensor nodes in the transmission range can receive signal from a transmitter without loss. To be consistent with general cases, sensor nodes are initialized with random residual energy between 1000 units and 5000 units, and random sensing activity rates between 10 % and 50 %. To further simplify the network model, we assume that each transmission consumes 1 unit of energy.

The clustering algorithms CDB and ICDB are used to organize WSNs into a hierarchy of clusters. In our demonstration, equally-weighted combination of residual energy and activity rate are used as context parameters. According to the clustering rule, nodes with higher residual energy and activity rates are selected first as cluster heads, where dynamic budget values are estimated proportionally to their energy levels. During the feedback phase of ICDB, cluster members calculate their individual occupation weights and transmit them to the corresponding cluster heads. After having completed the clustering processes, sensor nodes detect random events at a certain rate which is given by their activity rates. A data packet is generated at a sensor node with each event. Data packets are first sent to local cluster heads and then routed to data sink following the hierarchy of cluster heads. Since we are focusing on the performance of the clustered network structures, data aggregation or fusion are not involved in our simulations. To enable comparison with general WSN situations, an intermediate value (30 %) is selected as the reference activity rate. The performance of the clustering algorithms is valuated with different reference cluster sizes, namely a range of reference budget values.

The objective of the simulation experiments is to demonstrate the impact of selected context information as clustering parameters, as well as the advantages of our clustering algorithms in terms of energy conservation and end-to-end delay. We use the following metrics to capture the performance of our clustering algorithms:

Average packet delay: Given a specific distribution of activity rates, the amount of data packets generated in the network are statistically stable. The average delay is defined as network throughput divided by simulation time. This metric indicates the average time it takes for an event to be received by data sink.

Average energy consumption per packet: Defined as total energy consumption of the WSN divided by network throughput. This metric gives an accurate measure of network management in terms of energy conservation.

Time for first cluster head to die: When a cluster head runs out of energy, the routing tree below the node becomes no longer valid. This metric reflects the lifetime of the clustered network structure, as well as the quality of network load balancing.

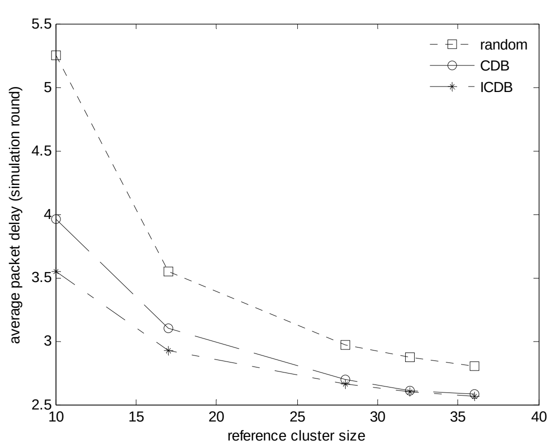

Figure 3. Average packet delay

Average delays of packets are calculated on their arrivals, as presented in Figure 3. Comparing to random cluster formation, the CDB algorithm achieves significantly advantage in terms of end-to-end delay. The result can be explained as follows: Active sensor nodes are included first in the spanning tree, which means major packet sources are placed nearer (in terms of hops) to data sink. Additionally, the CDB algorithm chooses sensor nodes with higher residual energy as cluster heads, where bigger budget values will be estimated. During clustering, the spanning tree grows implicitly with more sensor nodes added on each level. Thus, the tree depth is therefore decreased, which results averagely shorter distance (in terms of hops) between sensor nodes and data sink. The ICDB algorithm shows even better delay performance than CDB, because even more sensor nodes with high residual energy can be clustered when their active rates are low, which results in even shorter tree structures.

During simulation, data packets are generated mainly from these sensor nodes with high activity rates. When clustering processes are performed without context awareness, it is possible that highly active sensor nodes are placed far away (in terms of hops) from data sink. That means more routing efforts are need to routing the remote packets to data sink. With consideration to context information of sensor nodes during clustering, more active nodes tend to stay nearer (in terms of hops) to data sink, which leads to a higher network throughput. Given the same event distribution of a simulated WSN, the total energy consumption is supposed to be constant during a specified simulation time. Therefore, our clustering algorithms yields better performances in terms of energy consumption per packet.

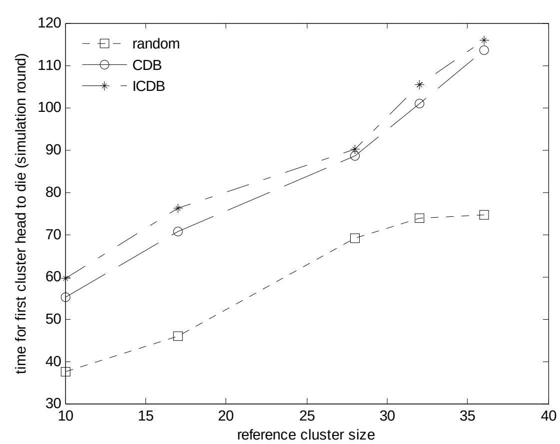

Figure 4. Time for the first cluster head to die

Figure 4 illustrates the shortest lifetime (in terms of simulation rounds) of cluster heads in the network. As explained earlier, when using the random budget-based clustering algorithm, it is also possible that sensor nodes with very little energy can be blindly selected as cluster heads and run out of batteries quickly. In contrast, the CDB algorithm avoids this situation by assigning the roles of cluster head only to sensor nodes with enough residual energy. To further balance the workloads of cluster heads, dynamic cluster sizes are estimated proportional to their residual energy. The ICDB algorithm further improves the performance using feedbacks during clustering. Cluster sizes are refined according to the activity rates of cluster members. We observe that the shortest lifetime of cluster heads increases with bigger reference cluster sizes. The reason is: A decreased depth of the network hierarchy reduces the overall energy consumption. Furthermore, when there are more sensor nodes on each level of the spanning tree, average routing loads on individual nodes will be eased.

5. Conclusions and future work

In this paper, we have proposed two context-aware budget-based clustering algorithms: the Clustering with Dynamic Budget (CDB) algorithm and the Interactive Clustering with Dynamic Budget (ICDB) algorithm. Context information of WSNs is utilized during clustering processes, where dynamic cluster sizes are estimated. Cluster heads and cluster members are also selected according to the context information of sensor nodes. Context information like residual energy and activity rates of sensor nodes are demonstrated as important clustering parameters. Simulation results demonstrate significant advances in terms of average packet delay, network throughput and energy conservation.

Currently, the clustering processes are performed hierarchically from data sink. In future, we intend to develop the algorithms into distributed versions, thus to improve their scalabilities. Depending on specific application scenarios, various kinds of context information (e.g. buffer size of sensor nodes, priority of events, network bandwidth, etc.) can be utilized in clustering processes. We plan to consider more kinds of context information in our algorithms, as well as their impacts as clustering parameters.

References

[1] G. J. Pottie and W. J. Kaiser, “Wireless Integrated Network Sensors”, Communications of the ACM, Vol. 43, No. 5, pp 51-58, May 2000.

[2] B. Warneke, M. Last, B. Liebowitz, Kristofer and S. J. Pister, “Smart Dust: Communicating with a Cubic-Millimeter Computer”, Computer Magazine, Vol. 34, No. 1, pp 44-51, Jan. 2001.

[3] K. Akkaya and M. Younis, "A survey of routing protocols for wireless sensor networks", Elsevier Ad Hoc Network Journal, 2005, Vol. 3/3, pp. 325-349.

[4] M. Joa-Ng and I.-T. Lu, “A peer-to-peer zone-based two-level link state routing for mobile ad hoc networks”, IEEE Journal on Selected Areas in Communications, August 1999, pp. 1415–1425.

[5] O. Younis and S. Fahmy, "Distributed Clustering in Ad-hoc Sensor Networks: A Hybrid, Energy-Efficient Approach", in Proc. of INFOCOM, March 2004.

[6] C. Intanagonwiwat, R. Govindan, D. Estrin, “Directed diffusion: a scalable and robust communication paradigm for sensor networks”, in Proc. of MobiCom, 2000.

[7] W. Heinzelman, A. Chandrakasan, H. Balakrishnan, “Energy efficient communication protocol for wireless sensor networks”, in Proc. of HICSS, January 2000.

[8] M. J. Handy, M. Haase, and D. Timmermann, "Low energy adaptive clustering hierarchy with deterministic cluster-head selection", in Proc. of MWCN, 2002.

[9] M. Younis, M. Youssef, and K. Arisha, "Energy-aware routing in cluster-based sensor networks", in Proc. of MASCOTS, October 2002.

[10] C. Schurgers, M.B. Srivastava, “Energy efficient routing in wireless sensor networks”, in Proc. of MILCOM, 2001.

[11] M. Chu, H. Haussecker, F. Zhao, “Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks”, The International Journal of High Performance Computing Applications 16 (3) (2002) 293–313.

[12] A. K. Dey, D. Salber and G. D. Abowd, “A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications”, Human-Computer Interaction, 16(2-4): pp. 97-166, 2001.

[13] Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., and Steggles, “Towards a Better Understanding of Context and Context-Awareness”, in Proc. of HUC, 1999.

[14] F. Michahelles, M. Samulowitz and B. Schiele, "Detecting Context in Distributed Sensor Networks by Using Smart Context-Aware Packets”, in Proc. of ARCS, 2002.

[15] WB Heinzelman, AL Murphy, HS Carvalho, and MA Perillo. “Middleware to support sensor network applications”, IEEE Network Mag. 18(1):6-14, 2004.

[16] R. Krishnan and D. Starobinski, “Efficient clustering algorithms for self-organizing wireless sensor networks”, Ad Hoc Networks, vol. 4, no. 1, pp. 36-59, January 2006.

[17] L. Tzevelekas, and I. Stavrakakis, "Directed budget-based clustering for wireless sensor networks", in Proc. of MASS, 2006, pp. 674-679.

[18] Y. J. Zhao, R. Govindan, and D. Estrin, "Residual energy scan for monitoring sensor networks", in Proc. of WCNC, March 2002.