Energy Efficient Data Gathering with Network Coding for Cluster-Based Wireless Sensor Networks

Nyan Lin, Rithea NGETH, Yasuo TAN, and Yuto LIM
School of Information Science, Japan Advanced Institute of Science and Technology (JAIST)
{nyanlin, ngethrithea, ytan, ylim} @jaist.ac.jp

Abstract

This paper presents the introduction of network coding technique into the data gathering process of cluster-based wireless sensor networks (WSNs), which possess the limited energy and computing resources. The network coding can indirectly reduce the amount of energy consumed for the transmissions from the cluster heads to the base station. An energy-efficient 2-hop path selection protocol is also applied to further reduce the energy consumption for data transmission from sensor nodes to a cluster head node. We set up a computer simulation in MATLAB to see the performance of our proposed scheme. Energy consumption and the number of transmissions are used as the performance evaluation metrics. The numerical results show that our proposed scheme can reduce the energy consumption for data transmissions from sensor nodes to the sink and improves the lifetime of the sensor network.

1. Introduction

Wireless sensor networks (WSNs) can be defined as a network of spatially distributed autonomous sensors (often hundreds) that sense the physical or environmental properties such as temperature, pressure, humidity, lighting, sound intensity, vibration and etc. to monitor the desired physical process in the environment. The acquired information is wirelessly communicated to a base station (BS), which propagates the information to remote devices for storage, analysis and processing. For many scenarios, direct communication between two peers is limited due to the distance or the obstacles between them. Multihop communication is a desirable property for a WSN and participants in a WSN works in ad hoc fashion. As a participant in an ad hoc network, a node has to involve in many tasks such as finding a next hop to send the data, routing decision and forwarding the data received from a neighbor node. Moreover, all the nodes in a WSN except the BS have their primary job, sensing the physical environment and transmitting the acquired information. All these jobs use the available energy in the battery of each sensor node.

For the durable battery life of individual sensor node or the longevity of the whole WSN, energy-efficient operations are needed. There are many works done in the literature for this purpose. These works can be classified into three main groups. The first group focuses on the in-network data aggregation. Providing the summary of the data collected by computing the statistical metrics (e.g., average, sum, max) can reduce the amount of data to be transmitted from a large number of sensors to a distant sink [1-2]. Another approach of this same group is the data compression based on the compressive sensing theory [3], in which the sink receives only a few encoded data of all the readings from sensors, being able to recover the original data, and reduces the global communication cost. These two approaches are called in-network aggregation with size reduction and in-network aggregation without size reduction in the survey paper of [4] respectively.

The second classification of energy-saving operations for a WSN is the design of routing protocols to route the packets from different sources directing towards the same destination. They are different from the traditional ad hoc routing protocols in the metrics they use to select the paths. For example, in data centric routing, sensor nodes should route packets to the next hop based on the packet content in order to promote in-network aggregation [4]. Shortest path routing is also a technique in which path with the least hop counts is selected in order to reduce, for example, the delay. Routing protocols in WSNs differ from each other depending on the application and the network architecture. Furthermore, they can also be classified into multipath-based, query-based, and QoS-based.

The third group of classification is the structure of nodes working together as a tree, or a cluster, or as a centralized approach. It may be considered as a part of the routing protocol designs because they are directly related to each other. Tree-based routing usually bases on hierarchical organization of the nodes in the network. A spanning tree is constructed rooting at the sink and aggregation is performed level by level from its leaves to the sink by answering the queries generated by the sink [5]. The drawback of this approach is due to the fact that wireless channels are not reliable. Loss at a given level of the tree means loss all data from that subtree. On the other side, cluster-based approach has many advantages.

A number of clustering algorithms have been designed for the WSNs in the literature depending on the network architecture or the characteristics of the cluster head nodes [6-7]. A cluster head may be elected by the sensors in a cluster or pre-assigned by the network designer. The cluster head can be a member of other sensors or a special sensor with high energy and computing resources. It can also be a member of a second-tier network or just a neighbor of the base station. Clustering can help save the communication bandwidth as the member sensors of a cluster only communicate to the cluster head and avoid interactions with other cluster. As the cluster head is defined the topology maintenance overhead are reduced and sensors will only care for connecting with their cluster heads. Although many data aggregation techniques and routing protocols are designed for a number of performance metrics such as energy efficiency, reliability, quality of service, etc., there is an important issue relating to the energy efficiency of the entire process of data gathering in WSNs. It is the error and loss in wireless communication due to the dynamic channel conditions.

A larger percentage of energy is especially consumed for the transmission than for sensing and computation in sensor nodes. Therefore, it is important to provide reliable communication to improve network lifetime of wireless sensor networks. The traditional way to provide the reliability is to use the feedback messages to report the received or lost packets. However, these feedback messages consume bandwidth. The main objective of retransmission scheme is to improve the throughput but at the expense of energy consumption. Energy efficient error control techniques to prolong network lifetime in resource limited network like WSNs and wireless communication remains a challenge [8]. Proper choice of error control code to provide the reliable communication and making the system energy efficient is another smart way. Network coding [9], which is initially introduced in the paper of Ahlswede et al., has recently become popular due to its advantages such as throughput increase and improvement in reliability and robustness of the wireless networks. The broadcast nature of wireless networks makes network coding more attractive to them because there is no restriction to the information flow in a path. Every node in the network can perform as a relay, coding the incoming packets together and sending the coded packets on outgoing links. In our work, we exploit network coding in order to increase the reliability of transmissions from the cluster heads to the sink. The transmissions from the sensor nodes to the cluster head is routed by an energy-efficient 2-hop path selection protocol to further reduce the required energy. We study the number of transmissions and energy consumption for the data gathering from sensor nodes to the sink.

The rest of the paper is organized as follow. Section 2 describes our proposed scheme, which comprises the network architecture, energy-efficient 2-hop path selection protocol, and the construction of random linear network coding technique. In section 3, we will show how the simulation is set up and run. Discussion on the performance of the proposed system will follow in section 4 based on the results obtained in section 3. Finally, section 5 concludes the paper and discusses the future work.

2. Proposed Energy Efficient Data Gathering Scheme

The proposed scheme is for the data gathering process at a sink node, which is the destination of all the other sensor nodes in the network. Especially our work focuses on the data transmission process from the sensor nodes to the sink. We do not consider the data aggregation functions on the acquired information in a cluster head or at the sink. The transmitted data packets may be the raw data packets or the aggregated ones. We are interested in the reliable communication between the sink and the cluster heads, which we assume as a leader node with the potential to perform the random linear network coding function over the incoming data packets from the member sensor nodes of its own cluster.

We assume that the network has been divided into some clusters with its own cluster head each. Clustering a network can be performed as described in the first phase of LEACH [10]. Firstly, nodes organize themselves into clusters and a node is elected as the cluster head within each cluster based on the probability distribution decision. If a node has been assigned as a cluster head, it sends advertisements to its neighbors. The neighbor nodes decide which cluster to join based on the signal strength of these messages.

2.1. Network Architecture for the Proposed Scheme

Our network architecture consists of a base station (sink), which is connected to an infrastructure network with high resources, and one or more cluster heads representing their own clusters. The number of clusters will depend on the clustering process of the clustering algorithm like [10] and the number of sensor nodes inside the network. The cluster heads can directly communicate to the sink. They are relay nodes that forward the sensor data from the sensor nodes inside the clusters to the sink. Another role of cluster heads is that they perform the network coding to merge the packets coming from different sources into the same packet so that a packet of the same size can carry much more physical information. These coded data packets are transmitted to the sink. According to the technique of random linear network coding, more redundancy is included in the transmitted coded packets, which will improve the reliability of communication, and packet error and loss are covered. The data transmissions from sensor nodes to cluster head inside a cluster are performed in the energy-efficient 2-hop path selection protocol.

2.2. Energy-efficient 2-hop Path Selection Protocol (2-PSP)

The 2-PSP [11] is proposed for a set of nodes, in which data can be sent faster using adaptive rate control capability of IEEE 802.11a/b/g MAC protocol via a relaying concept. The main objective is to build upon the opportunistic rate adaptation to assist a sender, a relay and a receiver to reach a higher rate data transmission. We modified 2-PSP in order for the decision of the path selection is made based on the lower energy consumption while in their work, the decision is only based on the available higher rate. We found that selecting the higher rate path cannot always guarantee the lower energy. Therefore, we modify it to fit into the desire mission of our job, which is to choose the energy efficient path selection.

We define three new control messages called relay RTS (RRTS), relay CTS (RCTS) and Ready to Relay (RTR) in our work. When a sender wants to send a data packet, it first waits for the DIFS+CW time period before it can transmit the data. After the sender can access the idle channel, it transmits a RRTS message. If the receiver receives the RRTS message correctly, it replies a RCTS message to the sender. A relay node that hears these control messages waits for a random short back off internal (SBI) period.

Within this period, the relay node decides whether to help the sender or not by estimating the energy consumption of direct transmission and the transmission with its help. Firstly, the relay node determines a suitable pair of higher data rates based on the signal strength of the receiving RRTS and RCTS messages. Then it calculates the energy consumption with one-hop transmission and that with two-hop transmission. If two-hop transmission saves more energy, the relay broadcasts a Ready to Relay (RTR) message. This message contains the information about the selected rates that the sender and receiver should use when they send their data packets. If the sender cannot hear any response after SBI interval times out, the sender will transmit the data packet according to the standard Distributed Coordination Function (DCF) procedure. If the sender can correctly decode the RTR message, it will transmit its data packet with the new data rate defined in the RTR message. The message handshaking procedure of 2-PSP is shown in Figure 1.

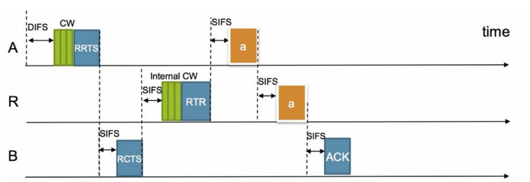

Figure 1. Message handshaking of 2-PSP

2.3. Construction of Random Linear Network Coding (RLNC) Technique

Network coding can be classified as either inter or intra-session. Inter-session network coding allows the packets from different sessions (sources) to be mixed before forwarding into the next hop. In contrast, intra-session network coding can be used to mix the packets from the same sessions to address the packet loss problem. RLNC [12] is a well-known coding technique of intra-session network coding. In RLNC, when a node receives a packet, the node stores the packet in its memory. When a packet injection occurs on an outgoing link of a node, the node forms the packet from a random linear combination of the packets in its memory. Suppose the node has L packets $\{u_1, u_2, \ldots, u_L\}$ in its memory. Then the packet formed is

$$u_0(t) = \sum_{i=1}^{L} \alpha_i u_i \quad (1)$$

where α_i is chosen according to a uniform distribution over the elements of F_q. The packet's global encoding vector γ that is placed in its header and which satisfies

$$u_0(t) = \sum_{k=1}^{K} \gamma_k w_k \quad (2)$$

where w_k is the message packets from the source. For the decoding process, sink node performs Gaussian elimination on the set of global encoding vectors from the packets in its memory to find the original messages.

The reason for choosing a linear framework is that the algorithms for coding and decoding are well understood. The size of the encoded packet is the same as the size of an incoming packet, but an encoded packet generally carries information about several original packets. We work on the combination of many incoming packets from different sensors in a cluster. The number of packets being coded together will be different from time to time depending on the packets received at a cluster head of the sensor nodes.

3. Numerical Simulations

To investigate the performance of the proposed scheme, we write a simulation in MATLAB. We use the IEEE 802.11a hardware specifications for the simulation assuming the multihop wireless network environment. We separate our simulation into two parts. The first part is for the evaluation of 2-PSP protocol in a cluster with 50 sensor nodes. For this work, nodes are randomly placed in a 2-D area with 100 x 100 m^2 size where the cluster head is placed at the upper right corner of the grid. We assume that all nodes have same communication range. The nodes communicate with the cluster head via the neighbor nodes in multihop fashion. Random sources are selected and data flows from these sources are routed along the paths discovered by the AODV routing protocol. Then in the data transmission to the cluster head, the 2-

PSP protocol is used to find a path in low energy consumption with the help of a relay as described in section 2.2. In our simulation, the transmission rates are approximated according to the distance between two nodes based on the transmission ranges as described in Table 1.

The second part of simulation is to see the performance of network coding and decoding at each cluster head and at the sink respectively. We create two clusters in our program, each with a cluster head that encodes 6 incoming data packets together at a time. These linear coded packets are transmitted to the sink until the sink can recover all the original data packets sent from the sensor nodes inside each cluster. We assume the link quality between the cluster heads and the sink to be at least 0.5. We also allow a cluster head to overhear a packet from the neighbor sensor node that is belonged to another cluster. This may be helpful to the sink to get more information for the decoding process. The simulation parameters are described in Table 1.

Table 1. Simulation parameters

Parameter	Value
Antenna type	Omni Antenna
RCTS, RTR size	15 bytes
RRTS size	20 bytes
ACK size	14 bytes
MAC header	34 bytes
Transmission power	100 mW
Transmission range	50, 45, 39, 33, 26, 19, 15, 13 m
Transmission rate	6, 9, 12, 18, 24, 36, 48, 54 Mbps
MAC protocol	CSMA/CA, 2-PSP
Payload size	1000 bytes

4. Results and Discussion

In this section, we will discuss the performance of the proposed scheme based on the results obtained from the simulation. For the transmissions from the sensor nodes to the related cluster heads, we study throughput and energy consumption of the modified 2-PSP protocol compared with the conventional CSMA/CA protocol. The number of data flows in a cluster is varied from 10 flows to 50 by increasing 10 flows each time. In the second part of simulation, we test the performance of transmission with the help of RLNC network coding compared to the transmission with LEACH of [10].

4.1. Comparison of Energy Consumption

As the Figure 2 depicts, the total energy consumption increases as the number of data flows increases inside the cluster. The energy consumption is the energy used for the transmission of control messages and the successfully transmitted data packets. The amount of energy consumption by the modified 2-PSP protocol for 10 data flows is 4.2302 mJ and 20.913 mJ for 50 flows. The amount of energy saving due to the modified 2-PSP protocol is about 15.2% in average. In our work we chose a relay based on the energy consumption rather than the rate by comparing two consumptions, one with direct transmission and another with two-hop transmission. This is different from the work of [11], where a candidate relay is selected only based on the achievable higher rate that the relay will apply in forwarding.

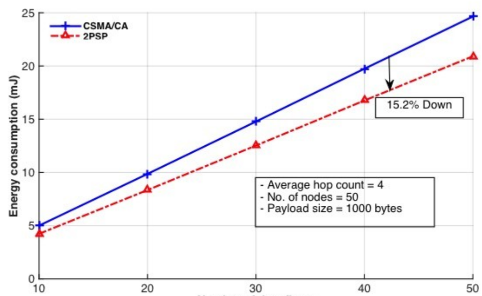

Figure 2. Energy consumption of 2-PSP

4.2. Comparison of Throughput

Figure 3 shows how throughput varies with the increase in the number of data flows from the sensors. Throughput is defined as the amount of payload data that can be received at the destination (here the cluster head) for all the unicast data flows transmitted by the sources (sensor nodes). The 2-PSP outperforms the CSMA/CA protocol by up to 14.9% improvement in average. The throughput decreases as the number of data flows increase. The reason for this behavior is because of the high latency when data flows increases. High latency means low throughput. When the number of flows in the network increases, there is high possibility of competing the media access, which leads to the delay and affects the throughput performance.

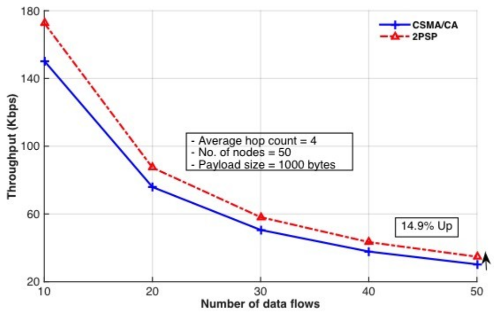

Figure 3. Throughput as a function of flow increase

4.3. Comparison of Number of Transmissions

The performance of transmission with the RLNC network coding and LEACH of [10] are compared for the cluster-based data gathering network architecture. Each cluster head node performs network coding function on the receiving data packets which include data from the own cluster and the overheard data from other cluster. We chose 6 data packets to be coded together each time to reduce the overhead. Transmission with RLNC coding scheme needed 25 transmissions with the link quality probability equal to 0.5 for the successful receiving of 6 original data packets. It costs 5.2 µJ for receiving 6 data packets. For the total 50 data flows from a cluster, the energy consumption is 0.26 mJ from the cluster head to the sink by the network coded transmission scheme and 20.913 mJ from 50 sensor nodes to the cluster head by the 2-PSP protocol. It means that it will totally cost 42.346 mJ for all the 100 data flows from the sensor nodes in two clusters while the number of transmissions by LEACH is 100 transmissions for energy 500 mJ of sensor nodes as described in their paper.

5. Conclusion

In this work, we have applied RLNC network coding technique for the data transmission between a sink and the cluster heads of each cluster inside the WSNs. We also applied modified 2-PSP protocol for the data transmissions from sensor nodes to the corresponding cluster heads. The simulation results showed that 2-PSP can help reduce the energy consumption in the data propagation process of sensors nodes up to 15.2%. The proposed transmission scheme with RLNC network coding can also reduce the number transmission by coding many packets together in each transmission. The 2-PSP protocol can also be applied while the cluster heads transmit the network coded data packets to the sink for further increase the throughput and to reduce the energy consumption for the case of many cluster heads away from the sink. This will be our future work.

References

[1] T.W. Kuo, and M. J. Tsai, “On the Construction of Data Aggregation Tree with Minimum Energy Cost in Wireless Sensor Networks: NP-completeness ad Approximation Algorithms”, *INFOCOM Proceedings*, IEEE, Orlando, 2012, pp. 2591-2595.

[2] Gao, Jie, et al., “Sparse Data Aggregation in Sensor Networks”, *Proceedings of the 6th International Conference on Information Processing in Sensor Networks*, ACM, 2007, pp. 430-439.

[3] Chong Luo, Feng Wu, Jun Sun, and Chang Wen Chen, “Compressive data gathering for large-scale wireless sensor networks”, *In Proceedings of the 15th annual international conference on Mobile computing and networking (MobiCom '09)*, ACM, New York, NY, USA, 2009, 145-156.

[4] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi,

"In-network Aggregation Techniques for Wireless Sensor Networks: A Survey", *Wireless Communications*, IEEE, Vol.14, Issue.2, 2007, pp. 70-87.

[5] S.Lindsey, C. Raghavendra, and K. M. Sivalingam, "Data Gathering Algorithms in Sensor Networks using Energy Metrics", *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, no.9, September 2002, pp. 924-935.

[6] G. Gupta, M. Younis, "Load-balanced clustering in Wireless Sensor Networks", *Proceedings of the International Conference on Communication*, Anchorage, Alaska, May 2003.

[7] O. Younis and S. Fahmy, "HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks", *IEEE Transaction on Mobile Computing*, vol. 3, no. 4, 2004, pp. 366-379.

[8] P. Salija and B. Yamuna, "Optimum Energy Efficient Error Control Techniques in Wireless Systems: a Survey", *Journal of Communications Technology and Electronics*, Pleiades Publishing, Inc., vol. 60, no. 11, 2015, pp. 1257-1263.

[9] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, "Network Information Flow", *IEEE Transactions on Information Theory*, vol. 46, no. 4, 2000, pp. 1204-1216.

[10] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan. "An Application-specific Protocol Architecture for Wireless Micro sensor Networks", *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, 2002, pp. 660-670.

[11] A. O. Lim and S. Yoshida, "A 2-hop Path Selection Protocol (2-PSP) in Multi-rate Ad Hoc Wireless Networks", *IEICE Transaction on Communication*, vol. E90-B, no.1, Jan. 2007.

[12] C. Fragouli, J. Y. L. Boudec, and J. Widmer, "Network Coding: an Instant Primer", *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 1, 2006, pp. 63-68.