Computer Communications 40 (2014) 65–75

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Adaptive data aggregation and energy efficiency using network coding in a clustered wireless sensor network: An analytical approach

Rashmi Ranjan Rout [a], Soumya K. Ghosh [b,*]

[a] Computer Science and Engineering, National Institute of Technology, Warangal, Warangal 506 004, India
[b] School of Information Technology, Indian Institute of Technology, Kharagpur, Kharagpur 721 302, India

ARTICLE INFO

Article history:
Received 10 November 2012
Received in revised form 13 November 2013
Accepted 18 November 2013
Available online 6 December 2013

Keywords:
Wireless sensor network
Network coding
Network lifetime
Duty cycle
Data aggregation

ABSTRACT

In an energy-constraint wireless sensor network (WSN), the sensor nodes near the *Sink* deplete their energy quickly due to heavy traffic and create a bottleneck zone. In a large monitoring area, the sensor nodes in a cluster may communicate with the cluster head using multi-hop mode. A bottleneck effect can also be created near the cluster heads in a cluster based WSN with several clusters. The clustering technique (along with data aggregation) exploits the data correlation among the sensors' data and provides opportunities for reduction of energy consumption in WSN. Integrating an energy efficient scheme, namely, network coding, with clustering and duty cycling may facilitate the design of a new cluster based data collection scheme. This work proposes an energy efficient adaptive data aggregation strategy using network coding (ADANC) which improves the energy efficiency in a cluster based duty-cycled WSN. A set of nodes (one-hop away from the cluster head) act as network coder nodes and rest of the nodes act as simple relay nodes in a cluster. The network coder nodes also act as aggregation points, opportunistically, based on the level of data correlation. The proposed ADANC strategy provides reduction of traffic inside a cluster and thus improves the energy efficiency of the bottleneck zone. It has been shown that the packet delivery ratio improves with the proposed scheme in a cluster. The improvement in energy efficiency with ADANC scheme has been analyzed and the upper bound of network lifetime has been estimated. A detailed theoretical analysis has been performed for a cluster based duty cycled WSN.

1. Introduction

The technological advancements in the area of micro-electronics and low power wireless communication methods facilitate the application developments of wireless sensor networks. The wireless sensor networks (WSNs) have a broad range of applications, such as, environmental monitoring, military surveillance and object tracking, disaster area relief, industrial control and seismic monitoring. Typically, a sensor network consists of a number of autonomous sensor nodes self-organize among themselves to perform a particular task. The WSNs are generally energy constraint, as sensors may not be recharged and replaced. The processing power of the tiny computational devices have been significantly improved in the last few years where as the battery powers of such devices have not enhanced significantly. Therefore, the design of data gathering protocols for WSN needs extra care in terms of energy consumption.

Usually, the energy dissipation of the nodes mostly depend upon the architecture of the WSN. Based on network architecture, broadly, a WSN is divided into flat ad-hoc and cluster based networks. A hierarchical design of a dense network facilitates the reduction of traffic in the network [1–6]. Hierarchical design creates opportunity for data aggregation locally [7]. Recently, the issues of modeling data aggregation and routing have received considerable attention. Moreover, a reduced volume of aggregated local data from a deployment area are propagated towards a central *Sink*. Data aggregation is one of the energy efficient technique used in WSN. The network lifetime can be enhanced with duty cycle based sensor nodes in a dense network. Furthermore, the density of sensor nodes (nodes per unit area) may be kept high during deployment for monitoring a large and remote geographical area with better coverage and connectivity. However in a dense WSN, a lot of energy is consumed due to redundant transmissions and receptions. The duty-cycle based sensor nodes turn-off their radios most of the times for reduction of energy penalty [8]. The traditional data aggregation techniques with highly correlated data can reduce network traffic without providing any reliability for data communication. Also, in a sparse WSN [9], the nodes are not in close proximity and data correlation is very low. Thus, the traditional data aggregation methods give very little improvement in energy efficiency. Moreover, with sporadic correlated (i.e. the level

* Corresponding author. Tel.: +91 3222 282332; fax: +91 3222 282206.
E-mail addresses: rashmirsays@gmail.com (R.R. Rout), skg@iitkgp.ac.in (S.K. Ghosh).

0140-3664/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comcom.2013.11.003

of correlation varies with time) sensor data, the WSN needs energy efficient and robust mechanisms for communication in addition to normal data aggregation.

Bandyopadhyay and Coyle [10] have considered a sensor network where sensors are randomly and uniformly distributed and the *Sink* is situated at the center of the monitoring area. The total area is divided into a number of clusters. Each cluster has a cluster head. The sensor nodes use *multi-hop* mode to communicate with the cluster head (CH) [10]. The CH aggregates data and forwards it to the *Sink* using the *multi-hop* mode. The energy burden on the sensors near the *Sink* has not been considered by Bandyopadhyay and Coyle [10]. The network lifetime through bottleneck zone analysis with *all-node-active* condition for a cluster based WSN (with single hop communication between sensor nodes) has been studied in [11,12]. However, there are needs for energy efficient and reliable data communication techniques in the bottleneck zone to improve the performance of the WSN. In the scenario of *multi-hop* mode of communication between the sensor nodes and the cluster head, there exist a bottleneck zone near the CH in each cluster. By reducing the data at the cluster level, the bottleneck zone effect near the *Sink* can be reduced in a clustered WSN.

Using *network coding*, the intermediate (relay) nodes can appropriately encode the incoming received packets before transmission in a network [13–15]. In this work, the network coding technique has been used to improve the energy efficiency in a cluster based duty-cycled WSN. The multi-hop mode of communication inside a cluster has been considered. The major contributions are as follows:

- Development of an adaptive data aggregation strategy using network coding (ADANC) to improve the energy efficiency in a cluster based WSN.
- Analysis of energy efficiency and estimation of lifetime upper-bound for a cluster based WSN with ADANC strategy.
- Estimation of the network lifetime upper-bound of a cluster based WSN with duty cycle and ADANC strategy.

1.1. Network coding

Network coding is a technique by which the intermediate nodes of a network can appropriately encode the information received from the neighboring nodes. Let, $(L_1, L_2, \ldots, L_n)$ be a set of packets generated and transmitted by various sensor nodes in a network. Let, these packets, $(L_1, L_2, \ldots, L_n)$, are received at an intermediate sensor node. Suppose, the intermediate node can encode at most v packets. The encoder intermediate node will generate $\frac{n}{v}$ number of packets and forward the same in the network. Generalized encoding and decoding procedures of linear network coding have been given below [16,17].

The coded packets using linear network coding are viewed as elements in finite field $GF(2^s)$. Here, 2^s is the size of the GF. A node chooses a sequence of coefficients $q = (q_1, q_2, \ldots, q_n)$ when the node transmits network coded data. The term q is called the encoding vector, from $GF(2^s)$. A set of n number of packets, G_i $(i = 1, 2, 3, 4, \ldots, n)$, are linearly encoded into a single output encoded packet at an intermediate node. The output encoded packet is formed from the following linear combination.

$$Y = \sum_{i=1}^{n} q_i G_i, \quad q_i \in GF(2^s) \tag{1}$$

Every intermediate node, which wants to transmit encoded data, computes a linear combination as given in Eq. (1). The network coded packets in the network are transmitted with the n coefficients. The receiver node can decode the encoded packets using the coefficients. The decoding procedure at the receiver side is described below.

At the receiver side, the encoding vector q is received with the network coded data. Suppose a node has received successfully a set of data $(q^1, Y^1), \ldots, (q^m, Y^m)$. The terms Y^j and q^j stands for the information symbol and the coding vector for the jth received data packet, respectively. The receiver node has to solve the following linear combinations (m equations and n unknowns):

$$Y^j = \sum_{i=1}^{n} q_i^j G_i, \quad j = 1, m \tag{2}$$

From Eq. (2), it can be seen that at least n linearly independent packets must be received at the recipient node for proper decoding of the network coded packets. In Eq. (2), the unknown term G_i contains the original packets forwarded in the network. The XOR network coding [13], which is a special case of linear network coding, has been used in this work. All the packets that are transmitted in the network are elements of $GF(2) = \{0, 1\}$.

1.2. Data aggregation in WSN

In this section, the traditional data aggregation methods have been mentioned. Also, the motivation behind the development of a hybrid technique by combining network coding and data aggregation has been discussed.

Data aggregation is one of the most important method for reduction of redundant data flow in WSN [18]. Data aggregation reduces the number of transmissions. In a typical dense WSN, the measurements of adjacent nodes are usually correlated [19]. Data aggregation is aimed at obtaining the inputs from some or all of the nodes in a WSN, accompanied by a common aggregate operator function such as *Sum*, *Average*, *Max*, *Min* and any other function [19,18]. Data aggregation can be done by a cluster head (CH) as well as at the node level in a clustered WSN [4–6,10]. A CH collects data from other cluster members and aggregates the data before forwarding to the *Sink*. Each member node also can locally reduce the sensed and generated data. In a multi-hop cluster based WSN, the data may be aggregated at the intermediate nodes.

Both network coding and data aggregation strategies may improve the performance of a typical WSN. However, the traditional data aggregation method performance will be limited if the data correlation is very low. If correlation is less, then the traffic generation will be more from the data aggregation points (nodes) in a WSN. For example, in a sparse network with distance based data correlation approach, the common data aggregation methods can not reduce the transmitted traffic towards a *Sink*. Furthermore, the traditional data aggregation approaches provide limited reliability (in terms of protection of data against link or node failure) in WSN. If an aggregate packet lost in the transmitted path then the packet can not be retrieved at the *Sink*. Moreover, the network coding approach (as discussed in Section 1.1) provide reliability and energy efficiency in a WSN [20,15]. However, in a clustered network where correlation of data among neighbor sensor nodes is very high, the traditional data aggregation also improves the energy efficiency in a WSN. For example, in a dense WSN, the data correlation is very high. So, in this case (with high data correlation), the encoding procedures needed for a sensor node (for network coding) can be avoided by using traditional data aggregation technique. Thus, the focus of this work is to develop an adaptive network coding based data aggregation strategy for a clustered WSN.

The rest of the paper is organized as follows. In Section 2, the related work has been presented. Section 3 presents the system model of the proposed technique. An energy consumption model has been discussed in Section 4. The proposed strategy has been presented in Section 5. In Section 6, the analysis of energy efficiency has been presented. The performance evaluations of