Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Computer Communications 29 (2006) 162–172
computer communications
www.elsevier.com/locate/comcom

An integrated energy aware wireless transmission system for QoS provisioning in wireless sensor network

Yong Yuan[a,*], Zongkai Yang[a], Zhihai He[b], Jianhua He[a]

[a]Department of Electronics and Information, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China
[b]Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65203, USA

Received 24 May 2005; accepted 24 May 2005
Available online 11 July 2005

Abstract

A significant amount of research efforts have been done to improve the energy efficiency of wireless sensor network (WSN). However, not much has been done to optimize the Quality of Services (QoS) of WSN. Actually, energy efficiency should be jointly optimized with QoS to make WSN useful. In this paper, we develop an integrated adaptive wireless transmission system for WSN to meet the QoS requirements in an energy-aware fashion. The QoS requirements in the application layer, the modulation and transmission schemes in the data link layer and physical layer are jointly analyzed and integrated into a single framework. In this framework, the cluster-based protocol is used, where the cluster heads form a multi-hop backbone. The intra-cluster QoS performance and inter-cluster QoS performance are both analyzed. Based on the results, the overall energy consumption and QoS in terms of BER performance, end-to-end transmission latency and packet loss ratio are modelled by the communication parameters of the cluster heads including the modulation level (bits per symbol) and transmit power. According to the model, a centralized off-line protocol is designed in the integrated framework to adjust the communication parameters of the cluster heads for fixed QoS provisioning. And a distributed online protocol is also designed for dynamic QoS provisioning. Extensive simulation results under various experimental settings demonstrate the energy-saving performance of the proposed integrated system. The ability to adapt to time-varying delivery quality requirements is also tested.

Keywords: Wireless sensor network; QoS provisioning; Cross-layer design; Modulation scaling

1. Introduction

A primary factor in determining the lifetime of a battery-powered wireless sensor network (WSN) is the energy supply. To maximize the operational lifetime of WSN, it is crucial to develop energy-efficient algorithms and mechanisms that optimize the energy consumption. Recently, a significant amount of research efforts have been focusing on the power optimization design for WSNs [1–6]. Since wireless transmission is the dominant power-consuming operation in WSN, special care has been taken in designing energy-aware wireless transmission schemes [1–5].

On the other hand, while much of the existing research on WSN has been focusing on energy- minimization and lifetime maximization, not much has been done to optimize the Quality of Service (QoS) of the wireless communication system. However, in many WSN applications, especially those in battlefield intelligence, the data gathering is often required to be timely and reliable [7]. In addition, different applications have different transmission quality requirements on the end-to-end latency and packet loss ratio. Even, some applications may have dynamic QoS requirements. For example, in the object tracking application, the end-to-end latency requirement will be dynamic since the moving speed of the object is time-varying. Therefore, QoS provisioning is important for WSN to be useful.

Traditionally, the problems of QoS provisioning and power optimization are considered separately at different layers of the OSI protocol stack, which is often not efficient in energy utilization. Having this in mind, in this paper, we propose to develop an integrated wireless transmission system for WSN. In the system, the QoS requirements from the application layer, the modulation and transmission at the link and physical layers are jointly designed. The cluster

* Corresponding author. Tel.: +86 27 87554315.
E-mail address: yy_hust@hotmail.com (Y. Yuan).

0140-3664/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2005.05.015

based protocol is adopted in the system, where the cluster heads form a multi-hop backbone. The intra-cluster QoS performance and the inter-cluster QoS performance are analyzed. Based on the results, the overall energy consumption and QoS in terms of BER performance, end-to-end transmission latency and packet loss ratio are modelled by the communication parameters of the cluster heads including the modulation level (bits per symbol) and transmit power. According to the model, a centralized off-line protocol is designed in the integrated framework to adaptively adjust the communication parameters of the cluster heads for fixed QoS provisioning. And we also designed a distributed online protocol in the integrated system according to the feedback control theory to adjust the communication parameters based on local information to provide dynamic QoS guarantee in an energy-aware fashion.

Our extensive experimental results demonstrate that the proposed wireless transmission system achieves significant energy saving under the QoS constraint. Because of the adaptive and distributed nature of the proposed system, it also has the capability to adapt to the dynamic delivery requirements.

To the best of our knowledge, this is one of the first papers on the joint design of the QoS requirements at the application layer and the modulation and transmission schemes in data link and physical layers for WSN. The cross-layer cooperation between these layers is also a unique feature of the proposed wireless transmission system. The rest of the paper is organized as follows. Section 2 gives a brief review of the related work. The overall architecture of the proposed wireless transmission system is described in Section 3. The overall QoS and energy consumption model of the proposed system are developed in Section 4. The design of the centralized off-line protocol and the distributed online protocol is presented in Section 5. Section 6 presents the simulation results. Section 7 concludes the paper.

2. Related work

Our work is closely related to the research work on the modulation scaling scheme and the protocol design based on feedback control theory for QoS provisioning in WSN.

The basic idea of modulation scaling is to adjust the modulation level according to the dynamic traffic load so as to make a good trade-off between the energy consumption and the transmission quality [1]. An optimization model and a distributed protocol are proposed in [5] to select the optimal modulation level for each node in the data aggregation tree to minimize the overall energy consumption under the delay constraint. AbouGhazaleh et al. [4] have proposed a dynamic transmission rate selection scheme by modulation scaling to vary the data bit-rate at each node in a given path (route) to save energy under delay QoS constraint. The work in [4] is similar to ours, since they both focus on adjusting the data bit-rate to prolong the lifetime of the network. However, there are still some differences between our work and [4]. First, in [4], the routing scheme is not considered and only the scenario of packet transmission along a fixed route is considered. In our proposed system, the cluster based routing protocol is jointly considered, and the data bit-rate of the cluster heads are adjusted for the overall QoS provisioning of the whole network. Second, only the bit transmission time is considered in [4] to adjust the data bit-rate at each node, latency introduced by queuing is not considered. Actually, due to the limited bandwidth and storage space, latency and packet loss introduced by queuing have large impacts on the end-to-end QoS performance especially under heavy traffic load, such as in the video surveillance application. In our work, the latency and packet loss introduced by queuing are both modelled into the end-to-end QoS performance.

Due to the high density and large-scale nature of WSN, QoS provisioning in WSN should be based on a predictable overall behavior through the aggregation among nodes. Since the feedback control theory excels in converging the unpredictability into a desired performance, it seems to have promise in QoS provisioning [7]. In Stankovic et al. [8] proposed a distributed feedback control scheme to balance the real-time workload among a set of computation nodes. Tian He et al. [9] proposed a soft real time guarantee routing protocol for WSN based on the feedback control theory. In the proposed system, we also design a distributed online protocol based on feedback control theory. The idea is similar to the work in [8,9]. The differences can be stated as following. First, in [9], the QoS requirements are fixed, the protocol parameters are adapted only to the dynamic network condition. In our work, we considered the use of feedback control theory for dynamic QoS provisioning. Therefore, the protocol parameters are adapted to both the dynamic QoS requirements and dynamic network condition. Second, we focus on using the feedback control theory to adjust the low level communication parameters, rather than the high level protocol parameters as the work by Stankovic et al.

3. System architecture

In this section, we present a reference architecture for the integrated wireless transmission system. As illustrated in Fig. 1, in the proposed system architecture, the end-to-end delivery QoS requirements are transferred from the application layer. Two types of applications with fixed QoS requirements and dynamic QoS requirements are both considered in the system. To correct the bit errors, an application-level error correction scheme is used which exploits the spatial–temporal correlation in the received sensor data [10]. This scheme specifies the desired BER performance.

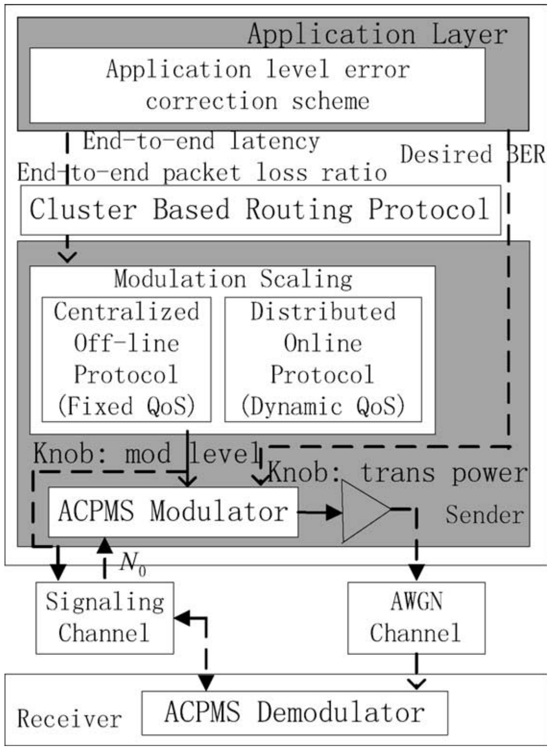

Fig. 1. System architecture.

The cluster-based routing protocol is adopted in the integrated system. In the protocol, the nodes in WSN are organized into multiple clusters. The cluster heads form a multi-hop backbone. The cluster members adopt BPSK as the modulation scheme. The cluster head adopts the adaptive code position modulation scheme (ACPMS) proposed in [11]. In ACPMS, the communication parameters including transmit power and modulation level can be adjusted. The communication parameters of the cluster heads in the network will influence the intra-cluster QoS performance, the inter-cluster QoS performance and therefore the end-to-end QoS performance, which will be modelled in Section 4. Based on the model, a centralized off-line protocol and a distributed online protocol are designed in the modulation scaling block of the integrated system. The protocols are designed to adjust the modulation level of each cluster head to guarantee the fixed QoS and dynamic QoS requirements. The design of the protocols will be discussed in Section 5.

In ACPMS, each node has access to two channels, one for data transmission and another for signaling. The data from the sender side will go through the serial to parallel conversion, modulation, CPM and wave formation. After being transmitted over the AWGN channel, the data is received and decoded by the receiver. The BER is computed by the receiver and fed back to the sender through the signaling channel. The information about BER will be used by the sender to estimate the power density of the channel noise and to control the transmit power to stabilize the BER performance at the desired value specified by the application-level error correction scheme. The power density of the channel noise will also be sent to the modulation scaling block to choose the optimal communication parameters. In [11], the BER performance of the scheme in different modulation level is analyzed. Based on the analysis, the average energy consumption per bit transmission is modelled as $E_{\mathrm{bt}}(k, BER_d, N_0)$, where BER_d is the desired BER performance, k is the modulation level and N_0 is the power density of the channel noise.

During the transmission, the operation will be divided into rounds. At the beginning of each round, the network will be organized into multiple clusters. Each cluster head will find the minimum energy consumption relaying route among other cluster heads to the sink. Then the short path tree (SPT) will be constructed by these routes. The cluster head will set up a TDMA schedule and transmit the schedule to its members. The cluster members will transmit data to the cluster head by multiple frames in each round. Within each frame, each cluster member will transmit at most one packet during its allocated transmission slot as specified by the TDMA schedule, and then sleep to save energy. The operation of the cluster members is similar to the LEACH protocol [12]. The cluster head will aggregate the data received in each frame, and transmit the aggregated data to the sink along the path in the SPT. And the modulation levels of the cluster heads will be adjusted by each frame according to the application's QoS requirements. If the application's QoS requirements are fixed, the centralized off-line protocol in the modulation scaling block will be used to find the optimal modulation level for each cluster head in the SPT. In such case, the topology information about the SPT and N_0 of the local environment of each cluster head, fed back from ACPMS, will be transmitted to the sink. The sink will find the optimal modulation level for each cluster head to guarantee the QoS requirements with minimum overall energy consumption. Then the sink will instruct the cluster heads to adjust their modulation levels. If the application's QoS requirements are dynamic, the distributed online protocol in the modulation scaling block will be used. In such case, each cluster head will adjust its modulation level locally based only on the information fed back from its parent node in SPT and N_0 fed back from ACPMS. After each cluster head changes the modulation level, ACPMS in each cluster head will adjust the transmit power to guarantee the desired BER performance. At the end of each round, the network will be clustered again and the above procedure will be repeated.

4. End-to-end QoS performance and energy consumption model of the system

As stated in Section 3, in the system, the network is organized into multiple clusters and the cluster heads form a SPT. In this section, we will model the intra-cluster QoS performance and inter-cluster QoS performance of the

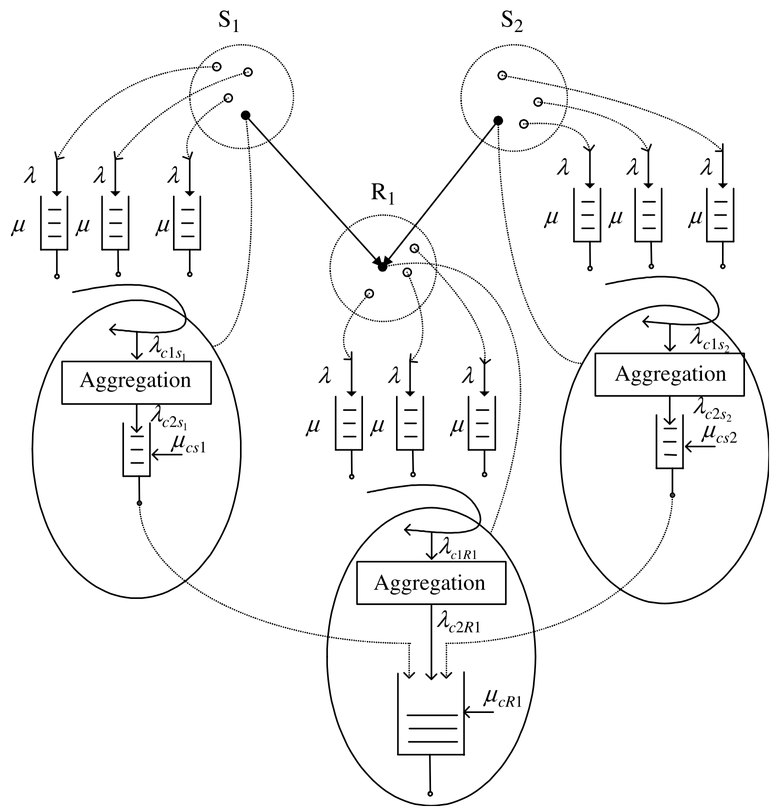

Fig. 2. Intra-cluster and Inter-cluster Queuing Process.

system at first. Then, the end-to-end QoS performance and overall energy consumption of the system will be modelled based on the result.

4.1. Intra-cluster QoS performance and inter-cluster QoS performance

As discussed in Section 3, the nodes in WSN will be organized into multiple clusters in each round. We adopt the policy of LEACH protocol for cluster head selection and cluster formation. The end-to-end QoS performance is impacted by the intra-cluster communication and the inter-cluster communication. In analyzing the QoS performance, the latency of packet transmission includes the transmission time and queuing time. And the packet loss is due to buffer overflow. In analysis, we assume the packet arrivals in each cluster member follows a Possion process, and the packet arrival rate is λ. The packets size follows a negative exponential distribution, with a mean of s, and the transmission buffer on each node can hold up to N packets. Under these assumptions, each node i can be viewed as an $M/M/1$ queuing system with limited capacity from the viewpoint of queuing theory.

According to the queuing theory, the stable queue empty probability $P_0(\mu, \lambda)$, the stable queue full probability $P_N(\mu, \lambda)$, the average queue length $L(\mu, \lambda)$, and the average sojourn time $W(\mu, \lambda)$ are functions of the service rate μ and arrival rate λ. The expression of the functions can be found in [13].

The queuing and transmission process of three clusters can be described by Fig. 2. In Fig. 2, cluster head R_1 is the parent of cluster head S_1 and S_2 in SPT.

According to Section 3, each cluster member will transmit at most one packet in each frame. Assume that there are N_{ci} members in the cluster i, the packet service rate of each member can be estimated as $\mu = \frac{R}{N_{ci}s}$, where R is the bit transmission rate. Since BPSK is used as the modulation scheme for the cluster member, R is equal to the bandwidth, denoted as B. In order to stabilize the queuing process of each cluster member, B should satisfy $B > N_{ci}\lambda s$. Therefore, the stable queue empty probability, the stable queue full probability and the average sojourn time for each cluster member can be described as $P_0\left(\frac{B}{N_{ci}s}, \lambda\right)$, $P_N\left(\frac{B}{N_{ci}s}, \lambda\right)$, and $W\left(\frac{B}{N_{ci}s}, \lambda\right)$, respectively. Since the packets arriving the cluster member will be dropped if the buffer is overflow, the packet loss ratio can be described as $P_N\left(\frac{B}{N_{ci}s}, \lambda\right)$. Therefore, the intra-cluster QoS performance for cluster i in terms of average intra-cluster latency, denoted as W_{0i}, and packet loss ratio, denoted as P_{L0i}, are $W\left(\frac{B}{N_{ci}s}, \lambda\right)$, $P_N\left(\frac{B}{N_{ci}s}, \lambda\right)$, respectively.

On the other hand, at any time the cluster head will receive packets from one of its members if the queue of the cluster member is not empty. Therefore, the packet arrival rate for the cluster head of cluster i, denoted as λ_{c1i} in Fig. 2, can be described by $\lambda_{c1i} = \frac{B}{s}\left(1 - P_0\left(\frac{B}{N_{ci}s}, \lambda\right)\right)$. And the cluster head will receive average λ_{c1i} packets per second. The cluster head will aggregate the data received before transmission. We adopt the aggregation model in [5] and use the aggregation factor to describe the correlation among data. Assuming the degree of correlation among the data before aggregation is indicated by the aggregation factor α, the number of packets per second after aggregation can be described by $\lambda_{c2i}=\frac{\lambda_{c1i}}{N_{ci}\alpha-\alpha+1}$. We assume the cluster head will use different frequency band for the intra-cluster communication and inter-cluster communication. Each pair of cluster heads in the SPT will use different frequency band. The frequency band allocation can be accomplished by the sink in constructing the SPT. Based on this assumption, the cluster head can communicate with its members and other cluster head at the same time.

Now, we are ready to model the inter-cluster QoS performance. The SPT can be represented by $T=\langle V, E\rangle$, where node set V is the set of all cluster heads in the SPT, and edge set E denotes the set of directed communication links between the cluster heads in the SPT. V can be grouped into two subsets, the set of leaf cluster heads (denoted as V_s) and the set of internal cluster heads (denoted as V_r).

As for the leaf cluster head i, such as S_1 and S_2 in Fig. 2, it only receives the packets from its members. The packet arrival rate for the queue is just λ_{c2i}. The packet service rate μ_{ci} of i is determined by its modulation level in ACPMS. If the modulation level is set to b_i, μ_i is equal to $\frac{b_iB}{s}$. Therefore, the average latency and packet loss ratio introduced by the queuing process of i, denoted as W_i and P_{Li}, are $W\left(\lambda_{c2i},\frac{b_iB}{s}\right)$ and $P_N\left(\lambda_{c2i},\frac{b_iB}{s}\right)$, respectively. And the stable queue empty probability of i, denoted as P_{0i}, is $P_0\left(\frac{b_iB}{s},\lambda_{c2i}\right)$.

As for the internal cluster head i, such as R_1 in Fig. 2, it receives not only the packets from its members but also the packets from its children cluster heads in SPT. Denote the set of children cluster heads of i as V_{si}. The total packet arrival rate of i can be described by Eq. (1)

$$\lambda_i=\sum_{j\in V_{si}}\mu_j(1-P_{0j})+\lambda_{c2i} \tag{1}$$

Therefore, the average latency and packet loss ratio introduced by the queuing process of i, denoted as W_i and P_{Li}, are $W\left(\lambda_i,\frac{b_iB}{s}\right)$ and $P_N\left(\lambda_i,\frac{b_iB}{s}\right)$, respectively. And the stable queue empty probability of i, denoted as P_{0i}, is $P_0\left(\lambda_i,\frac{b_iB}{s}\right)$.

Denote L_i as the path from cluster head i to the sink in SPT, the total mean latency and packet loss ratio of the inter-cluster communication for cluster head i can be described by $W_{CAi}=\sum W_k$ and $P_{CLAi}=1-\prod(1-P_{Lk})$.

4.2. End-to-end QoS performance and overall energy consumption model

Based on the above analysis, the end-to-end QoS performance in terms of mean latency and packet loss ratio for cluster i can be described by Eqs. (2) and (3).

$$W_{Ai}=W_{CAi}+W_{0i} \tag{2}$$

$$P_{LAi}=1-(1-P_{CLAi})(1-P_{0i}) \tag{3}$$

Therefore, the mean end-to-end latency and packet loss ratio for the whole network can be described by Eqs. (4) and (5).

$$W_A=\frac{\sum_{j\in V_s\cup V_r}N_{cj}W_{Aj}}{N_{\text{total}}} \tag{4}$$

$$P_{LA}=\frac{\sum_{j\in V_s\cup V_r}N_{cj}P_{LAj}}{N_{\text{total}}} \tag{5}$$

where N_{total} is the total number of the nodes in the network.

The overall energy consumption of the whole network can be divided into two portions, the energy consumption of the cluster heads, denoted as E_{CH}, and the energy consumption of the cluster members, denoted as $E_{\text{non-CH}}$.

As discussed in Section 3, ACPMS is used as the modulation scheme of the cluster head. The energy consumption per bit for cluster head i can be modelled as $E_{bt}(b_i,\text{BER}_d,N_{0i})=\frac{P_{ct}}{Bb_i}+\frac{N_{0i}}{A}\times 10^{0.1\cdot\sqrt{\frac{\lg(\text{BER}_d)-b_{b_i}}{a_{b_i}}}}$ [11], where P_{ct} is the circuit power consumption, A is the power loss factor, N_{0i} is the power density of noise of the local environment of cluster head i, a_{b_i} and b_{b_i} are the coefficients of the parabola approximating the curve of $\lg(BER)$ vs. $\frac{E_b}{N_0}$ when modulation level is b_i. We assume the channel is AWGN with squared power path loss, $\frac{1}{A}$ is proportional to d^2 where d is the distance between i and its parent cluster head in SPT.

The parent cluster head of i in SPT will also spend $E_{br}(b_i)$ energy to receive each bit from i. $E_{br}(b_i)$ can be described by $E_{br}(b_i)=\frac{P_{cr}}{Bb_i}$, where P_{cr} is the circuit energy consumption of the receiver.

The total energy consumption per bit in transmitting when modulation level is b_i can be described by $E_b(b_i, BER_d, N_{0i})=E_{\text{bt}}(b_i, BER_d, N_{0i})+E_{br}(b_i)$.

Therefore, E_{CH} can be described by $E_{\text{CH}}=\sum_{i\in V_s\cup V_r}E_b(b_i,\text{BER}_d,N_{0i})$.

On the other hand, $E_{\text{non-CH}}$ has been analyzed in [12], which is described by $E_{\text{non-CH}}=(N_{\text{total}}-k_c)\left(E_{\text{elec}}+\varepsilon_{f_s}\frac{M^2}{k_c}\right)$, where k_c is the number of clusters, ε_{f_s} is the energy consumption parameter in free space (*fs*) model and M is the edge length of the sensed region.

Therefore, the overall energy consumption of the whole network in terms of energy consumption per bit transmission can be described by Eq. (6).

$$E(\{b_i\},\{N_{0i}\},k_c)=E_{\text{CH}}+E_{\text{non-CH}} \tag{6}$$

5. The centralized and distributed protocols

With the end-to-end QoS performance and overall energy consumption model in the last section, we are ready to design the centralized off-line protocol and the distributed online protocol in the integrated system to provide the fixed QoS guarantee and dynamic QoS guarantee.

5.1. The centralized off-line protocol for fixed QoS provisioning

The task of the centralized off-line protocol is to adjust the modulation level of each cluster head to provide the fixed end-to-end QoS guarantee with minimum energy consumption. This protocol is used when the application's QoS requirements are fixed.

The search for the optimal modulation level for each cluster head can be formulated as an optimization problem summarized in Table 1.

In this optimization framework, if the number of clusters is small, the search space will not be too large. Then, the exhaustive search method can be used to find the optimal set of modulation levels. If the number of clusters is large, some evolutionary computation techniques, such as particle swarm optimization [14], can be used to solve the problem.

Based on the optimization model in Table 1, we can design a centralized off-line protocol to adjust the modulation level of each cluster head in each round. There are four major steps in the protocol. First, in cluster formation phase, if each cluster head has found the shortest path to sink, the shortest path will be sent to sink. Then, the topology of the SPT will be saved by sink. Second, the number of members N_{ci} and the power density of noise of local environment N_{0i} of each cluster will be sent to the sink. Third, according to the optimization model in Table 1, sink will find the optimal modulation level for each cluster head. Fourth, sink will send the optimal modulation level to each cluster head. And the cluster head will adjust its modulation level correspondingly.

It should be noted that the number of clusters, k_c can also be chosen adaptively for each round, which will be our future work. In the current work, k_c is set to be a fixed value.

Table 1
The optimization model

Definitions
The SPT of the cluster heads, $T=\langle V, E\rangle$, $V=V_s \cup V_r=\{\{S_1, S_2,\ldots,S_m\}, \{r_1, r_2,\ldots,r_n\}\}$;
For leaf cluster head $i \in V_s$
Intra-cluster mean latency, $W_{0i}=W\left(\frac{B}{N_{ci}s}, \lambda\right)$;
Intra-cluster mean packet loss ratio, $P_{L0i}=P_N\left(\frac{B}{N_{ci}s}, \lambda\right)$;
Mean latency introduced by queue of i, $W_i=W\left(\lambda_{c2i}, \frac{b_iB}{s}\right)$;
Mean packet loss ratio introduced by queue of i, $P_{Li}=P_N\left(\lambda_{c2i}, \frac{b_iB}{s}\right)$;
For internal cluster head $i \in V_r$
Intra-cluster mean latency, $W_{0i}=W\left(\frac{B}{N_{ci}s}, \lambda\right)$;
Intra-cluster mean packet loss ratio, $P_{L0i}=P_N\left(\frac{B}{N_{ci}s}, \lambda\right)$;
Total packet arrival rate for queue, $\lambda_i=\sum_{j \in V_{si}} \mu_j(1-P_{0j})+\lambda_{c2i}$, where V_{si} is the set of children cluster heads of i in SPT;
Mean latency introduced by queue of head, $W_i=W\left(\lambda_i, \frac{b_iB}{s}\right)$;
Mean packet loss ratio introduced by queue of head, $P_{Li}=P_N\left(\lambda_i, \frac{b_iB}{s}\right)$;
Inter-cluster mean latency for cluster head i, $i \in V_s \cup V_r$,
$W_{Ai}=\sum_{k:V_k \in L_i} W_k+W_{0i}$, where L_i is the path from cluster head i to sink in SPT.
Inter-cluster mean packet loss ratio for cluster head i, $i \in V_s \cup V_r$,
$P_{LAi}=1-\left(1-\left(1-\prod_{k:V_k \in L_i}(1-P_{Lk})\right)\right)(1-P_{0i})$;
Objective: min $E(\{b_i\}, \{N_{0i}\}, k_c)$. Refer to Eq. (6) for the expression of $E(\{b_i\}, \{N_{0i}\}, k_c)$.
Subject to:
The mean end-to-end latency, $W_A=\frac{\sum_{j \in V_s \cup V_r} N_{cj}W_{Aj}}{N_{total}}$, being less than T, $W_A \le T$.
The mean end-to-end packet loss ratio, $P_{LA}=\frac{\sum_{j \in V_s \cup V_r} N_{cj}P_{LAj}}{N_{total}}$, being less than P_l, $P_{LA} \le P_l$.
Expected solution: Find the optimal set of $\{b_i, i \in V_r \cup V_s\ b_i \in \{1,2,3,4\}\}$.

5.2. The distributed online protocol for dynamic QoS provisioning

When the applications have dynamic QoS requirements, a distributed online protocol is needed to adjust the communication parameters. In this paper, we designed such a distributed online protocol, which is shown in Fig. 3. During each round, after the cluster formation the protocol will adjust the modulation level of each cluster head in the SPT for dynamic QoS provisioning in the real time fashion. In the protocol, each packet is added by a head containing packet ID, node ID, a time stamp recording the generating time, denoted as *TSG* and a time stamp recording the arrival time at each cluster head, denoted as *TSR*. When packet arrives each cluster head, *TSR* will be updated to the new time. The packet ID generated by each node are continuous and unique by each round. The head is used by each cluster head and sink to calculate the QoS parameters. In the protocol, sink will calculate the mean end-to-end latency T_d and mean end-to-end packet loss ratio P_d. The method to calculate T_d is to average the time duration between the arrival time of each packet at sink and the time recorded in *TSG* for some period. Since the packet ID is continuous for each node, sink can calculate P_d by dividing the total number of lost packet IDs by the total number of packet IDs generated by each node. The task to calculate T_d and P_d is accomplished by the End-to-end QoS Calculator (EQC) module in sink. Then the difference between T and T_d, denoted as d_t, and the difference between P_l and P_d, denoted as d_p are calculated by the QoS Difference Dispatch (QDD) module of sink. Then, a packet including d_t, d_p, T_d and P_d is fed back to the QoS Difference Accumulator and Dispatch (QDAD) module of the children cluster head in the SPT.

In the internal cluster head i, the Subtree QoS Calculator (SQC) module will be used to calculate the mean latency and mean packet loss ratio introduced by the subtree rooted at i, denoted as T_{di} and P_{di}. The method to calculate T_{di} and P_{di} is similar to the method to calculate T_d and P_d by the sink. In the internal cluster head, a Node QoS Calculator (NQC) is also used to calculate the mean latency and mean

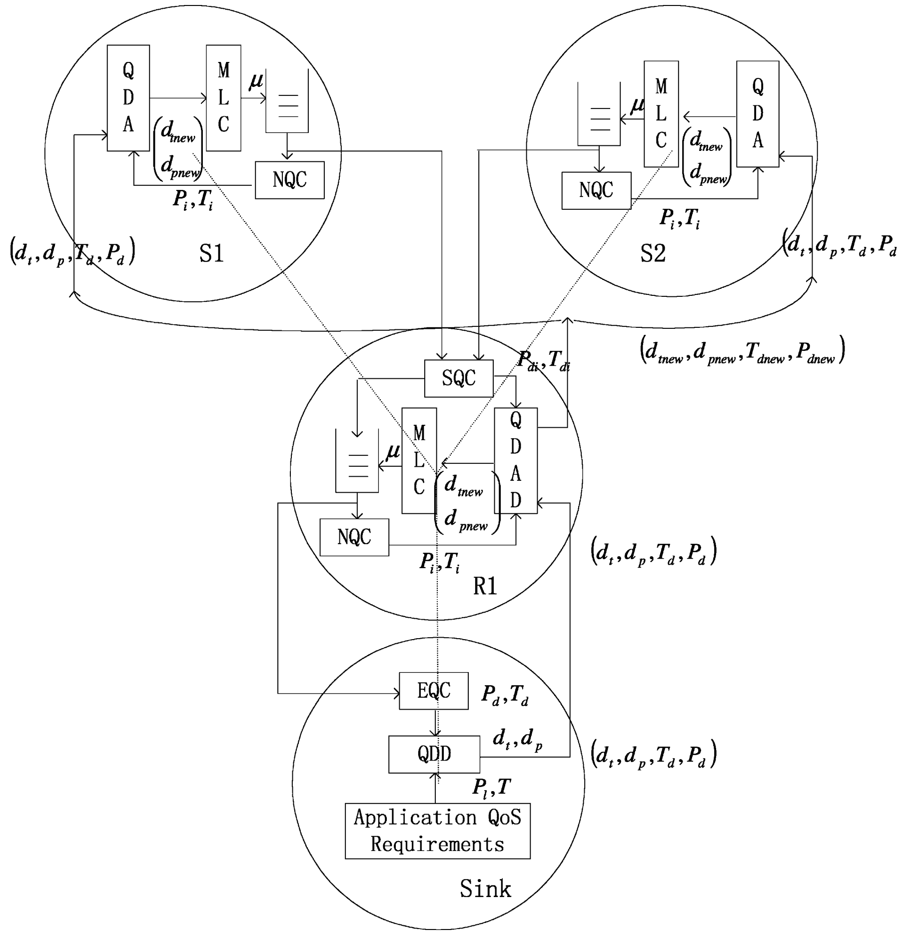

Fig. 3. Distributed online protocol.

packet loss ratio introduced by the queuing process at i, denoted as T_i and P_i. The method to calculate T_i is to average the time duration between the leaving time of each packet at i and the time recorded in *TSR* in the packet for some period. The number of packets arriving when the queue is overflow is added up, and divided by the number of total packets arriving the queue to get P$_i$. Based on P_i, T_i, P_{di}, T_{di}, the QDAD module will update d_t, d_p, T_d and P_d by setting $d_{\text{tnew}} = d_t + (T_d - T_{di} - T_i)$, $d_{\text{pnew}} = d_p + (P_d - (1-(1-P_{\text{di}})(1-P_i)))$, $T_{\text{dnew}} = T_{\text{di}}$ and $P_{\text{dnew}} = P_{\text{di}}$. Then, the modulation level of i will be adjusted by the MLC (Modulation Level Controller) module of i. The algorithm of MLC can be described by Fig. 4. where β is a parameter to trade off the speed of convergence of the protocol and the oscillation of system performance, *MAXLATENCY*, *MAXPLOSS* are the maximum possible mean end-to-end latency and packet loss ratio acquired by setting all modulation levels to 1. After adjusting the modulation level, the packet service rate of i will be $\mu = \frac{b_1 B}{s}$

In Fig. 4, the maximum available modulation level *MAXMODLEV EL*(N_0, BER_d) depends on the local power density of noise N_0 and the desired BER performance BER_d. If the N_0 is large or the required BER_d is rigorous, adopt large modulation level will cost too much transmission energy to maintain the desired *BER* performance. So *MAXMODLEV EL*(N_0, BER_d) should be inverse proportional to N_0 and proportional to BER_d.

Then, the packet with d_{tnew}, d_{pnew}, T_{dnew} and P_{dnew} will be dispatched by QDAD module to the children cluster head in SPT.

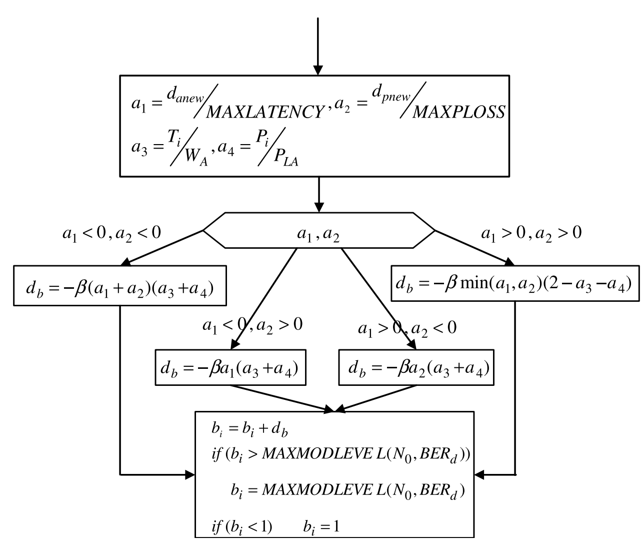

Fig. 4. Algorithm of MLC.

As for the leaf cluster head, only NQC module is used to calculate P_i and T_i. And the QDA(QoS Difference Accumulator) is used to calculate d_{tnew} and d_{pnew} based on the packet received from its parent cluster head. Then, the modulation level will be updated by MLC.

The major idea in the protocol design is to consider the queuing process of the whole SPT as a feedback control system. Our objective is to achieve the desired end-to-end QoS performance. The difference between the actual end-to-end QoS performance and the target performance are fed back to each node in SPT as the control packet. If a path in the SPT is found to have violated one of the application's QoS requirements, d_t and d_p will become negative in the received control packet of the nodes along the path. In such case, the modulation levels of those nodes will be increased so as to improve the performance. The amount of modulation level to increase on each node should be proportional to the amount of packet loss and latency introduced by the node, and also should be proportional to the severity of the violation. If the actual QoS performances of the path are better than the application's requirements, the modulation levels of the nodes along the path will be relaxed to save energy. The protocol is heuristic in nature. It can only obtain the sub-optimal performance. However, since the protocol can adjust the communication parameters online, it can be used for dynamic QoS provisioning.

Since the protocol only depends on the information exchanged between parent node and children nodes, the overhead of protocol is small. Assume that the distributed online protocol will adjust the modulation levels every T_c seconds. The control packet contains four float numbers. Each float number needs 32 bits for binary representation. Then, the size of the control packet is only 128 bits. Since each cluster head will receive only one packet in each adjustment, the overhead of the protocol is just $128k_c$ bits. However, the average total bits generated by all sensor nodes in T_c seconds are $N_{\text{total}}\lambda T_c s$. Therefore, the overhead of the protocol is negligible.

It should be noted that, although the distributed online protocol is designed to provide the dynamic QoS guarantee, It can also be used to provide the fixed QoS guarantee.

6. Experimental results

In this section, we evaluate the performance of the proposed integrated wireless transmission system. Our simulations are organized as follows: First, to test the energy saving performance of the integrated system in providing the fixed QoS guarantee by the centralized off-line protocol. Second, to show the energy saving and the adaptation capability of the integrated system in providing the dynamic QoS guarantee by the distributed online protocol. In order to test the integrated system, a simulator is developed using the PASREC software [16], which is a discrete-event simulation language. To understand the performance of energy saving and QoS guarantee, we need to compare the integrated system with a baseline protocol. However, to our best knowledge, there is not other known baseline approach in WSN for the similar application scenario, in which the latency and packet loss ratio introduced by queuing are both considered. Therefore, we compare the integrated system with a baseline protocol, in which all nodes send the packet at the highest speed, as studied in the PAMAS protocol [16]. In the experiments, the related system parameters are summarized in Table 2.

Table 2
The system parameters

$\frac{N_0}{2} = -134$ dBm/Hz	$B = 20$ kHz	$\lambda = 4$
$N = 5$	$\alpha = 0.5$	$P_b = 10^{-3}$
$P_{ct} = 98.2$ mw	$P_{cr} = 112.6$ mw	

6.1. The energy saving performance in fixed QoS provisioning

To evaluate the energy saving performance of the integrated system by the centralized off-line protocol, we consider a scenario with 195 nodes distributed uniformly in $200m \times 200m$. Therefore, $N_{total} = 195$ and $M = 200m$.

In our experiments, we evaluate the proposed integrated system and the baseline protocol for multiple rounds. In each round, the network will be grouped into k_c clusters. The SPT will be constructed among the cluster heads. For the baseline protocol, the modulation level will always be set to 4. And the optimal modulation level for each cluster head in the proposed integrated system will be found by sink according to the centralized off-line protocol. Then, the intra-cluster communication and inter-cluster communication will be started. At the end of each round, the percent of alive nodes by the integrated system and the baseline protocol will be saved and compared.

Fig. 5 shows the network topology in one round by the integrated system. The mean end-to-end latency for each cluster are $W_{A0} = 1.668$, $W_{A1} = 0.667$, $W_{A2} = 1.668$, $W_{A3} = 0.667$, and $W_{A4} = 0.667$ ms. The mean packet loss ratio for

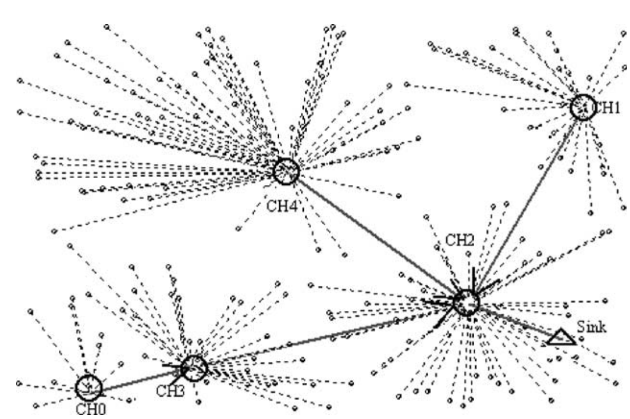

Fig. 5. The network topology in one round.

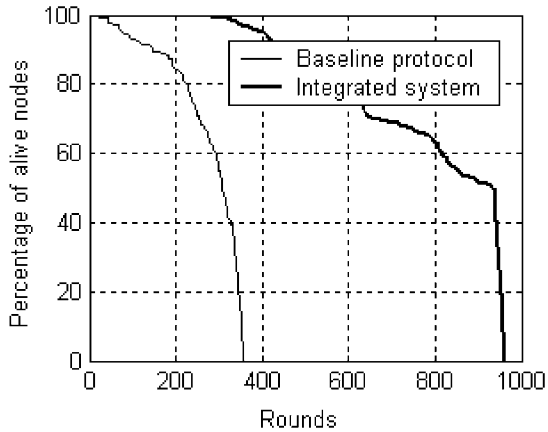

Fig. 6. Percentage of alive nodes, $T=3$ ms. $P_l=1e$-3, $k_c=5$.

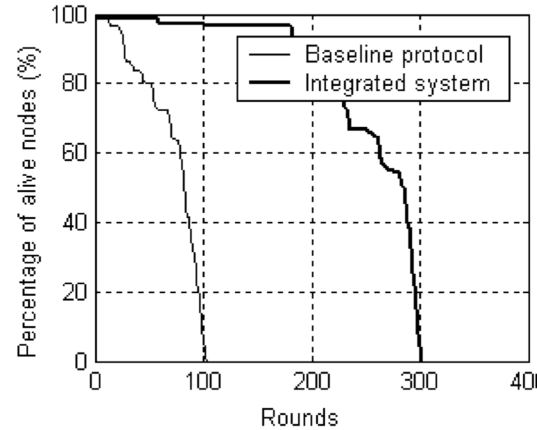

Fig. 8. Percentage of alive nodes, $T=1.5$ ms. $P_l=1e$-3, $k_c=7$.

each cluster are $P_{LA0}=3.552e-4$, $P_{LA1}=1.57e-4$, $P_{LA2}=1.24e-4$, $P_{LA3}=2.24e-4$, and $P_{LA4}=3.12e-4$. And the modulation level for each cluster are $b_0=1$, $b_1=1$, $b_2=3$, $b_3=1$ and $b_4=1$.

The overall network energy efficiency in terms of percent of alive nodes after each round is evaluated for the proposed integrated system and the baseline protocol in three scenarios with different T, P_l and k_c. The percentage of alive nodes of the proposed integrated system and the baseline protocol are shown in Figs. 6–8.

If we define the network lifetime of WSN as the duration of more than 60% of network nodes being alive, we then observe that the network lifetime of WSN with the integrated system, denoted as $t1$ and baseline protocol, denoted as $t2$ in the three test scenarios are (812,274), (567, 126) and (260,76) respectively. And the improvement of the proposed integrated system, denoted as $\frac{t2-t1}{t1}$ in the three test scenarios are 1.96, 3.5 and 2.42, respectively.

From the experimental results including those shown in Figs. 6–8, it can be seen that the improvement of energy saving by the proposed integrated system is significant. If the QoS requirements are looser, the energy saving performance will be more significant.

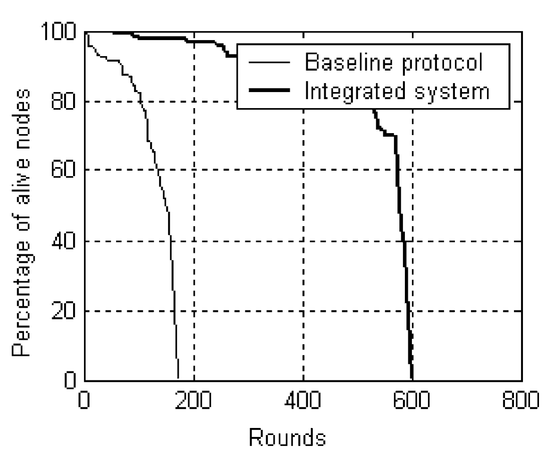

Fig. 7. Percentage of alive nodes, $T=3$ ms. $P_l=1e-3$, $k_c=7$.

The actual end-to-end latency by the proposed integrated system and the baseline protocol are shown in Fig. 9. As shown in Fig. 9, the actual end-to-end latency by the baseline protocol is always much better than the application's requirements, which is not necessary. On the other hand, the actual end-to-end latency by the proposed integrated system is almost equal to the application's requirements.

6.2. The performance in dynamic QoS provisioning

In order to evaluate the performance of energy saving and the adaptation capability of the integrated system in guaranteeing the application's dynamic QoS requirements, the following experiments are carried out. In our experiments, the SPT is assumed to be a complete binary tree with 63 nodes. In order to test the adaptation capability of the integrated system, the end-to-end QoS requirements are varied with time and the actual end-to-end QoS performance and energy dissipation are measured.

In the first experiment, the mean end-to-end latency requirements are varied as the dotted line in Fig. 10.

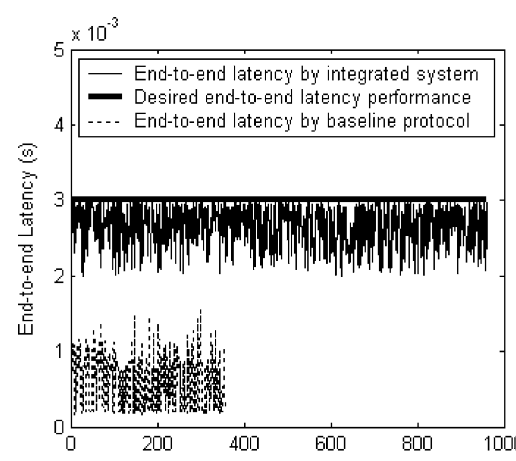

Fig. 9. End-to-end QoS performance by two protocols.

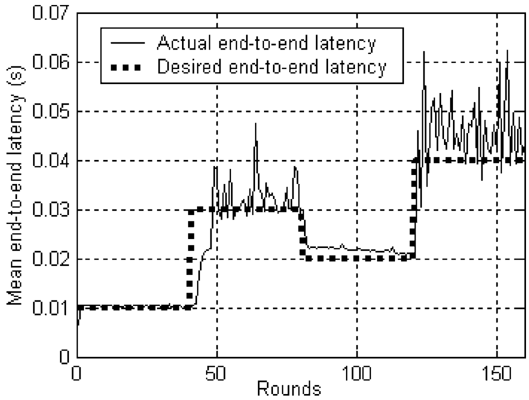

Fig. 10. Mean end-to-end latency during each round.

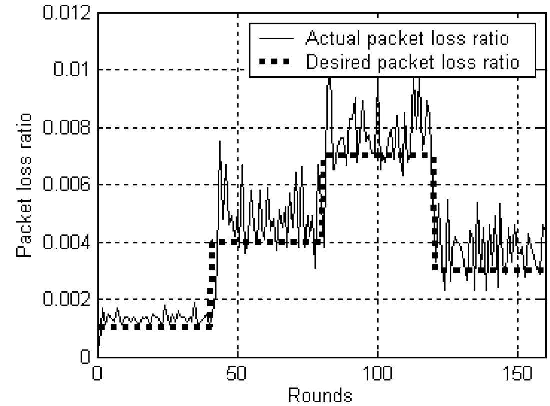

Fig. 12. Mean end-to-end packet loss ratio during each cycle.

The mean end-to-end packet loss ratio requirement is fixed at 0.5. The actual mean latency is shown in Fig. 10. If the baseline protocol is used, the latency will always be 0.0066. The percentage of energy saving during each cycle achieved by the proposed algorithm is shown in Fig. 11.

In the second experiment, the mean end-to-end packet loss ratio requirements varies, as shown by the dotted line in Fig. 12. The mean end-to-end latency requirement is fixed at 0.1 s. The actual mean packet loss ratio is shown in Fig. 12. If the baseline protocol is used, the packet loss ratio will always be about 10^{-5}. The percentage of energy saving during each cycle achieved by the proposed algorithm is shown in Fig. 13.

From the experimental results including those shown in Figs. 10–13, it can be seen that the amount of energy saving by the proposed integrated system is varied with the latency and packet loss requirements. If the latency and packet loss requirements are looser, more energy will be saved. In most scenarios, the proposed integrated system achieves about 30–40% energy saving compared to the baseline protocol. In some scenarios with much relaxed QoS requirements, the protocol can even achieve more than 50% energy saving. In addition, the protocol is able to adapt to the time-varying QoS requirements very well. Another point should be noticed is that there are some peaks in the experimental result figures. The reason is that the end-to-end latency and packet loss ratio are random. The objective of the integrated system is to ensure the mean end-to-end QoS performance stabilized at the desired level. From the experimental results, we can see the actual end-to-end latency and packet loss ratio are surrounding with the required dynamic QoS performance. And in the investigated scenarios, we found that when the required QoS performance varied, the mean of the actual QoS performance could be adjusted from the old level to the new level in about 10 rounds. Therefore, the QoS performance can be stabilized quickly.

7. Conclusion

In this paper, an integrated energy and QoS aware wireless transmission system is developed for the WSN. The proposed system is able to minimize the energy consumption under the fixed or dynamic QoS constraints. In the integrated wireless transmission system, the QoS requirements from the application layer, the modulation and transmission schemes in the data link layer and physical layer are jointly analyzed and integrated into a single framework. In the framework, the cluster based protocol is used. The cluster heads form a multi-hop backbone, where

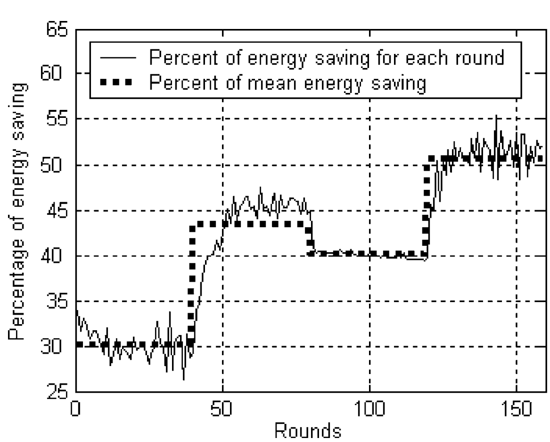

Fig. 11. Energy conservation during each cycle.

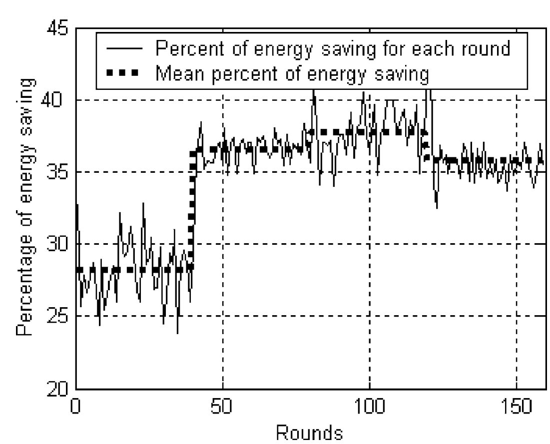

Fig. 13. Energy conservation during each round.

the modulation level and transmit power of the cluster heads can be adjusted adaptively. The intra-cluster QoS performance and inter-cluster QoS performance are analyzed. Based on the results, the mean end-to-end latency and packet loss ratio and the overall energy consumption are modelled. According to the model, a centralized off-line protocol and a distributed online protocol are designed in the integrated framework to adjust the modulation level and transmit power of each cluster head to guarantee the fixed and dynamic QoS requirements.

Extensive simulation results under various experimental settings have demonstrated the energy-saving performance of the proposed integrated system. The ability to adapt to time-varying delivery quality requirements is also demonstrated.

In our future work, we will analyze the robustness and convergence of the proposed distributed online protocol. A guideline to choose the parameters will be developed. In addition, the optimal packet transmission for the WSN with multiple sinks will be investigated within the proposed integrated wireless transmission system.

Acknowledgements

The authors would like to thank the editors and anonymous reviewers for their valuable comments. This project is supported by the National Science Foundation of China (Grant No. 60202005).

References

[1] C. Schurgers, O. Aberborne, and M. Srivastava, Modulation scaling for energy-aware communication systems, in ISLPED, pp. 96–99, 2001.

[2] B. Prabhakar, E. Uysal-Biyikoglu and A.E. Gamal, Energy-efficient Transmission over a Wireless Link via Lazy Packet Scheduling, in Proceedings of IEEE INFOCOM, Anchorage, Alaska, vol. 1, pp. 386–394, March 2001.

[3] A.E. Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, Energy-efficient scheduling of packet transmissions over wireless networks, in Proceedings of IEEE INFOCOM, New York, USA, vol. 3, pp. 1773–1782, June 2002.

[4] N. AbouGhazaleh, P. Lanigan, S. Gobriel, D. Mosse and R. Melhem, Dynamic Rate-Selection for Extending the Lifetime of Energy-Constrained Networks, in Proceedings of IEEE International Conference on Performance, Computing, and Communications, pp. 553–558, April 2004.

[5] Y. Yu, B. Krishnamachari and V.K. Prasanna, Energy-Latency Tradeoffs for Data Gathering in Wireless Sensor Networks, in Proceedings of IEEE INFOCOM, Hong Kong, P.R.China, March 2004.

[6] V. Raghunathan, C. Schurgers, S. Park, and M. Srivastava, Energy-Aware Wireless Microsensor Networks, IEEE Signal Processing Magazine, vol. 19, pp. 40–50, March 2002.

[7] J.A. Stankovic, T. Abdelzaher, C. Lu, L. Sha and J. Hou, Real-Time Communication and Coordination in Embedded Sensor Networks, in Proceedings of the IEEE, vol. 7, pp. 1002–1022, July 2003. (invited paper).

[8] J.A. Stankovic, T. He, T.F. Abdelzaher, M. Marley, G. Tao, S. Son, and C. Lu, Feedback Control Scheduling in Distributed Systems, in IEEE RTSS, pp. 59–70, London, UK, December 2001.

[9] T. He, J.A. Stankovic L. Chenyang, T.F. Abdelzaher, SPEED: a stateless protocol for real-time communication in sensor networks, in IEEE ICDCS pp. 46–55, May 2003.

[10] S. Mukhopadhyay, D. Panigrahi and S. Dey, Data aware, Low cost error correction for wireless sensor networks, in IEEE WCNC, pp. 2492–2497, March 2004.

[11] Y. Yuan, Z.K. Yang, J.H. He, W.Q. Chen, An Adaptive Code Position Modulation Scheme for Wireless Sensor Networks to appear at IEEE Communications Letters 2005.

[12] W.R. Heinzelman, A. Chandrakasan, and H. Balarislman, An Application-Specific Protocol Architecture for Wireless Microsensor Networks, IEEE Transactions on Wireless Communications, vol. 1, no. 4, pp. 660–670, Oct 2002.

[13] D.J. Lu, Random process and application (in Chinese), Tsinghua Press, China, 1986.

[14] J. Kennedy, R.C. Eberhart, Particle swarm optimization, Proceedings of IEEE Conference on Neural Networks, vol. IV, 1995 pp. 1942–1948.

[15] PARSEC Project. [Online]. Available: http://pcl.cs.ucla.edu/projects/parsec.

[16] C.S. Raghavendra, S. Singh. PAMAS-power aware multi access protocol with signaling for ad hoc networks, Computer Communication Review, July 1998.