Soft QoS Provisioning for Wireless Sensor Networks: a Cooperative Communications Approach

Xuedong Liang*, Min Chen†, Victor C.M. Leung* and Ilangko Balasingham‡

*Dept. of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada V6T 1Z4
Email: {xuedongl, vleung}@ece.ubc.ca

†School of Computer Science and Engineering, Seoul National University, Korea, 151-742
Email: minchen@ieee.org

‡Dept. of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway N-7491
Email: ilangkobat@iet.ntnu

Abstract—To achieve full cooperative diversity gains while still obtaining spectral and energy efficiency, cooperative communications with relay selection schemes, i.e., only the *best* relay is selected from multiple relaying candidates to cooperate with the communication, have been extensively studied in recent research. In this paper, we review the recent research on optimal relay assignment for cooperative communications, and investigate the use of cooperative communications with adaptive relay selection for soft QoS provisioning in resource-constrained wireless sensor networks. We propose *EEARS*, an energy-efficient adaptive relay selection scheme, which is based on a multi-agent reinforcement learning framework. In *EEARS*, optimal relays, in terms of outage probability, spectral and energy efficiency, are selected distributedly from multiple relaying candidates to participate in the communication, without the needs of prior knowledge of the wireless network model and centralized control. Simulation results show that *EEARS* fits well in dynamic environments, and is effective in improving the satisfying level of soft QoS provisioning for WSNs, i.e., increasing the spectral and energy efficiency, and reducing the amount of time that QoS violation occurs, by exploiting spatial and time diversities. Furthermore, compared with schemes using fixed transmitting power, *EEARS* can achieve a higher energy efficiency by varying the transmission power level according to wireless channel conditions.

I. INTRODUCTION

Due to low-cost node platforms, self-organizing manner and ease of deployment, wireless sensor networks (WSNs) have numerous potential applications, e.g., medical care, battlefield surveillance, wildlife monitoring, and disaster response. In these mission-critical applications, a set of QoS (quality of service) requirements (e.g., delay, packet delivery ratio, network lifetime, throughput and communication bandwidth) on network performances must be satisfied. However, providing guaranteed QoS is almost impossible in dynamic WSNs [1], [2], due to the dynamic network topology, time-varying wireless medium, and severe constraints on power supply, computation power, and communication bandwidth. Therefore, it is more practical to provide soft QoS than guaranteeing hard QoS, especially in multi-hop WSNs [3]. In soft QoS provisioning framework, when a data flow with certain QoS requirements is in connection, there may exist transient amount of time that the QoS requirements cannot be met. The level of soft QoS provisioning can be quantified by the fraction of total QoS violation time over the total connection time, and the ratio should not be higher than an application-determined threshold.

For a QoS-support route, QoS violations may occur because the intermediate routers cannot fulfill the QoS attributes that they have been assigned or promised in the QoS route discovery and establishment procedure, which might be caused by network topology changes, concurrent transmission interferences, thermal noise, shadowing, and multi-path fading. For instance, as shown in Fig. 1, for two adjacent routers, e.g., n_l and n_m, along the established route, the link between n_l and n_m may experience severe channel fading and thus cannot meet the assigned QoS requirements. Retransmitting the packet, e.g., using the ARQ mechanism, from n_l to n_m might not be effective in this case, as the link between n_l and n_m may remain in deep fading for a long period because of strong time correlation in a slowly varying wireless medium [4].
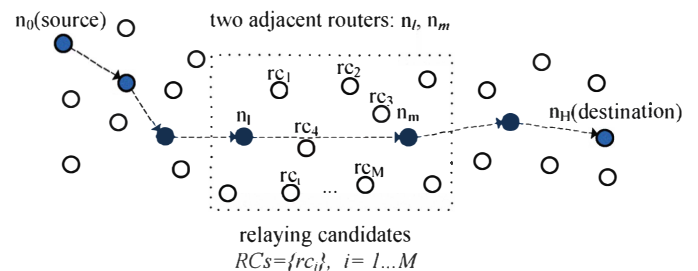

Fig. 1. Cooperative communications with relay selection in multi-hop WSNs

As nodes in WSNs are often spatially deployed, the channel fading gains for different links are assumed to be statistically independent in WSNs [5]. Therefore, there might exist some nodes, which are neighboring nodes of both n_l and n_m, overhear the packet transmission between n_l and n_m, due to the broadcast nature of the wireless medium. Among these multiple neighboring nodes, a node, e.g., rc_i, can be selected to help in the packet delivering between n_l and n_m by retransmitting the overheard packet to n_m, even it has not been assigned any routing tasks in the route discovery and

establishment procedure. By doing so, the receiver n_m may receive two copies of the original signal, which are transmitted over the two independent paths (n_l - n_m, n_l - rc_i - n_m) and experience different channel fading. Node n_m can combine the signals received from node n_l and rc_i, e.g., by applying MRC (maximum-ratio-combining) [6] for optimal packet decoding, or simply choose the signal with higher SNR (signal-to-noise-ratio) and decode it. Therefore, channel diversity gains can be achieved, i.e., a packet transmission failure occurs only when both of the two independent links experience deep channel fading simultaneously. This is known as cooperative communications with relay selection (or optimal relay assignment), that only one node is selected among multiple candidates as an optimal relay to help in the packet transmission, and other nodes remain silent. By exploiting the diversities of the wireless medium, cooperative communications with relay selection can be applied in soft QoS provisioning for WSNs to assure the QoS attributes and increase the satisfying level of soft QoS provisioning.

Compared with conventional multi-node cooperative communication systems, where all the available relays participate in the communication by retransmitting signals, cooperative communications integrated with adaptive relay selection has been demonstrated to have the potential of achieving full cooperative diversity gains while still obtaining higher spectral and energy efficiency, as well as lower transmission delay. However, it is challenging to find optimal relay selection policies in dynamic and distributed WSNs (e.g., when to cooperate? how to cooperate? and whom to cooperate with?), wherein the network state information is inherently imprecise and tend to vary.

A number of adaptive relay selection schemes have been proposed recently. In the literature, most of the cooperative protocols choose optimal relays based on the network state information (e.g., network topology, distance, channel coefficients, and SNR), and assume that full or partial CSI (channel state information) is available at the source, destination and all of the potential relays. However, the dynamic and distributed nature of WSNs implies that the network state information tend to be varying, and the available information is often inaccurate and incomplete. Therefore, algorithms relying on the network state information cannot achieve optimal network performance in dynamic environments. Moreover, significant communication overhead will be incurred in acquiring and disseminating of such information to all of the cooperative participants, especially for the cooperative protocols, as in [7], [8], that instantaneous CSI is required at all the potential relays for optimal relay selection. Thus, research on distributed, lightweight and adaptive relay selection scheme is still needed.

In this paper, we first review the recent research on optimal relay assignment for cooperative communications. Then, we investigate the use of cooperative communications with adaptive relay selection for soft QoS provisioning in WSNs, and propose *EEARS*, an energy-efficient adaptive relay selection scheme for cooperative communications. In *EEARS*, for each pair of adjacent intermediate routers along a multi-hop route,

an optimal relay in terms of packet outage probability, spectral and energy efficiency is distributedly selected from multiple relaying candidates. The selected optimal relay participates in the communication between the routers by retransmitting the overheard packet. To further improve the network performance, a power control scheme is integrated with the optimal relay selection, i.e., the selected relay retransmits the data packet using an adaptive transmission power level according to the inter-user channel conditions.

Simulation results show that *EEARS* fits well in dynamic environments, and is effective in improving the satisfying level of soft QoS provisioning for WSNs, i.e., reducing the amount of time that QoS violation occurs by exploiting spatial and time diversities. Furthermore, compared with schemes using fixed transmitting power, *EEARS* can achieve higher spectral and energy efficiency by utilizing the adaptive transmission power control scheme.

The rest of the paper is organized as follows. We review the recent research on optimal relay assignment in Section II. The background information of reinforcement learning and its applications in WSNs are provided in Section III. The system model is presented in Section IV. Section V describes the algorithm overview and design issues of *EEARS*. The design and implementation of the reinforcement learning based algorithm are illustrated in Section VI. The performance analysis is presented in Section VII. Finally, Section VIII concludes the paper and discusses the future research directions.

II. RELATED WORK

A number of relay selection protocols for cooperative communications have been proposed. Based on the criteria used for optimal relay assignment, the schemes can be classified into three main categories.

A. Pre-defined and random relay assignments

The simplest solutions for optimal relay assignment are assigning the relays in advance, or choosing the relays randomly in runtime, as the protocols proposed in [9], [10]. The pre-defined and random schemes can reduce the design complexity and network overhead. However, such schemes cannot achieve optimal performance in dynamic environments and lack of the capacity of dealing with network dynamics.

B. Distance-based relay assignment

An intuitive scheme of optimal relay assignment is using distance, either towards the source or the destination, as the criterion of optimal relay selection. In distance-based cooperative protocols, [10] chooses a node which is the closest one to the destination as the optimal relay. In contrast, a relaying candidate which is the closest one to the source, will be chosen as the optimal relay in [11].

However, it is well understood that the communication between senders and receivers with similar distances may have significant differences in terms of received signals' strengths and SNRs, due to the interferences, shadowing and multi-path fading effects on the wireless links. Therefore, the use of

distance as the criterion of optimal relay assignment cannot reflect the channel state appropriately.

C. SNR (or channel gain)-based relay assignment

The most intuitive solution of optimal relay assignment is that choosing relays, which have the highest SNRs or the maximum wireless channel gains with both the source and destination.

A cooperative relay framework which accommodates the physical, MAC and network layers for wireless ad-hoc networks was proposed in [12]. In the network layer, for a link between a source and a destination, two relaying candidates which have the two highest of the minimum SNR of the relay channels (from the source to the relay, and from the relay to the destination) are selected as the optimal relays.

The authors in [7] proposed an adaptive relay selection scheme for cooperative communication protocols, based on the CSI at the source and the relays. The optimal relay is the node which has the maximum instantaneous scaled harmonic mean function of its source-relay and relay-destination channel gains.

In [9], the source chooses N relays among all the relaying candidates, whose received signals' SNRs are the N highest, to participate in the communication between the source and destination.

In [8], each relaying candidate distributedly assesses the link qualities between the source-relay and relay-destination, by using the signaling messages, e.g., RTS (request-to-send) and CTS (clear-to-send) signals. The optimal relay is selected based on the instantaneous channel measurements.

The optimal relay assignment scheme [13] is integrated with a power control mechanism in the physical layer. The relaying candidates also use RTS and CTS signals to assess the link qualities and individually compute the required transmission power that can meet the desired link qualities. Different from [8], the source also participate in the competition procedure of relay selection.

Opportunistic single relay selection protocols under an aggregate power constraint were presented in [14]. For decode-and-forward (DaF) [15] protocol with reactive relay selection, the optimal relay is the candidate which has the maximum instantaneous channel gain between the relay and destination. For DaF protocol with proactive relay selection (the optimal relay is chosen prior to the transmissions of source-destination and relays-destination), the optimal relay is the candidate that can maximize the minimum of the weighted channel strengths between the source and destination.

However, the use of SNR as the unique relay selection criterion is not sufficient in dynamic WSNs. It has been shown in [9] that received SNR based selection schemes behave similarly to random selection schemes, or even slightly worse in some scenarios.

III. REINFORCEMENT LEARNING AND ITS APPLICATIONS IN WSNs

Reinforcement learning provides a framework in which agents learn control policies in dynamic environments based on experiences and rewards. In a standard reinforcement learning model, an agent is connected to the environment via perception and action, as shown in Fig. 2. On each step of interaction, the agent receives an input, i, some indication of the current state, s, of the environment; the agent then choose an action, a, to generate as output. The action changes the state of the environment, and the value of the state transition is communicated to the agent through a scalar reinforcement learning signal, r. The agent's behavior, B, should choose actions that tend to increase the long-term sum of values of the reinforcement signal [16].
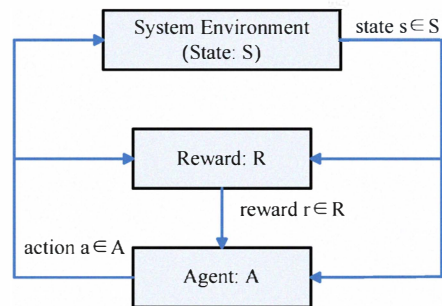

Fig. 2. A standard reinforcement learning model

The underlying concept of reinforcement learning is Markov Decision Process (MDP). A MDP models an agent acting in an environment with a tuple (S,A,P,R), where S is a set of states, A denotes a set of actions. $P(s'|s,a)$ is the transition model that describes the probability of entering state $s' \in S$ after executing action $a \in A$ at state $s \in S$. $R(s,a,s')$ is the reward obtained when the agent executes a at s and enter s'. The goal of solving a MDP is to find an optimal policy, $\pi : S \mapsto A$, that maps states to actions such that the cumulative reward is maximized. Detailed information on reinforcement learning can be found in [16].

Multi-agent systems (MASs) are systems that multiple agents are connected to an environment and all the agents may take actions to change the state of the environment. In MASs, from an agent's perspective, the agent itself can take actions and change the state of the environment. Furthermore, other agents in the system can also change the environmental state by taking their actions. Therefore, when an agent chooses its actions, the agent should consider other agents' possible actions. WSNs can be regarded as multi-agent systems, where sensor nodes can be considered as agents, the wireless medium and data flows can be regarded as environment. The agents may take actions (e.g., packet sending, receiving and forwarding) to change the state of environment. Moreover, the agents interact (contend and/or collaborate) with others due to the shared and contention nature of the wireless medium.

IV. SYSTEM MODEL

We consider a WSN in which sensor nodes are uniformly and randomly distributed in a certain area. Source and destination nodes are randomly chosen to establish a multi-hop route, which might be used for multiple data session communication.

AODV (Ad hoc On-Demand Distance Vector) and CSMA/CA are employed as the underlying network and MAC layer protocols, respectively. DaF is used as the cooperative transmission scheme. To achieve optimal decoding performance, MRC is utilized at the receiver for packet decoding by combining the multiple signals received from the sender and relay. We assume that the nodes can adjust their transmission power level dynamically, i.e., nodes can choose transmission power levels within the range of $[P_0, P_1, ..., P_t, ..., P_T]$.

We consider two QoS metrics, i.e., end-to-end delay T_{end} and packet delivery ratio P_{end}, that should be satisfied by the underlying communication network. QoS violation occurs if any of the two metrics measured cannot meet the desired QoS constraints, when a data session is in connection.

The wireless channel model is assumed to be a narrow band Raleigh block fading channel with additive white Gaussian noise (AWGN) [17]. For any two nodes, e.g., i and j, the channel gain h_{ij} is modeled as zero-mean circular symmetric complex Gaussian random variables with zero mean. For the links between a pair of nodes, the channel fading gain coefficients are assumed to be reciprocal, i.e., $h_{ij} = h_{ji}$, and the expectation $E(|h_{ij}|^2) = 1$. The channel fading gain coefficients are constant for a given transmitted block, or a codeword, but are independent and identical distributed (i.i.d.) for different blocks [9]. For different links, the channel fading gain coefficients are statistically i.i.d., which is a reasonable assumption as the nodes are usually spatially deployed [4].

V. COOPERATIVE COMMUNICATIONS WITH ADAPTIVE RELAY SELECTION

A. Algorithm Overview

In *EEARS*, first, for a data session with certain QoS requirements, a multi-hop route is discovered and established by using the AODV routing protocol. Then, for each pair of adjacent intermediate routers (a pair of 1-hop sender and receiver) along the established route, a number of nodes will be selected as a set of relaying candidates for the communication between the sender and receiver. If the link between the sender and receiver cannot meet the desired qualities, an optimal relay will be chosen from the set of relaying candidates to help in the packet delivering by retransmitting the packet. The selection of relaying candidates is valid for the lifetime of the established route. While the optimal relays are selected in a session-by-session manner for the communication between the source and destination, as the wireless medium is assumed to be slowly varying, as well as for the purpose of reducing the computation overhead and design complexity.

The cooperative transmission scheme operates in two phases, namely, direct transmission and relaying transmission phases. If a packet transmission fails in the direct transmission phase, the relaying transmission phase will be invoked, and an optimal relay among the relaying candidates is selected to retransmit the packet to the receiver using an appropriate transmitting power level. Then, the receiver combines the signals received from both the sender and relay for optimal packet decoding.

When a data session starts, each relaying candidate assesses it capabilities of being an optimal relay, and may choose to participate in the communication, or decide to remain silent. When the data session completes, each candidate will get a feedback from the corresponding receiver. The feedback contains the QoS information of the link between the sender and receiver, e.g., 1-hop delay and packet delivery ratio, which represents the quality of the cooperative communication, and can be regarded as an immediate reward (could be positive or negative) given by the environment. The candidates then use the immediate reward and the expected long-term reward in the remaining path to update the decision policy, i.e., the optimal decision of relay assignment will be strengthened; and the sub-optimal decisions will be weakened by a series of trial-and-error interactions with the dynamic environment. From the updated policy, the candidates can decide whether it should participate in the communication with an appropriate transmitting power level, or remain silent, according to the environmental state.

Once the algorithm reaches convergence, the relaying candidates are able to use the learned policy to take appropriate actions, i.e., the candidate which can make the most contributions in terms of outage probability, spectral and energy efficiency will be more likely to be selected as the optimal relay in the following data sessions.

B. Relaying Candidate Selection

In *EEARS*, the relaying candidate selection is integrated with the route finding mechanism, and the selection is valid during the route's lifetime. The route finding mechanism in *EEARS* is based on the AODV routing protocol with QoS extension, i.e., using the mechanism of route request (RREQ), route reply (RREP) and route error (RERR) messages to discover and maintain an initial QoS-support route.

As shown in Fig. 1, a node determines that it is a relaying candidate rc for the adjacent routers n_l and n_m, if it has heard both of the $RREQ$ transmitted by n_l and the $RREP$ replied by n_m, and has not been selected by n_l as the next hop router in the route discovery procedure.

Formally, the set of relaying candidates $RCs = N_l \cap N_m$, where N_l and N_m are the sets of the immediate neighboring nodes of n_l and n_m, respectively. Thus, for each relaying candidate $rc_i \in RCs$, rc_i is a common immediate neighboring node of both n_l and n_m. Ideally, all of the relaying candidates are connected to both n_l and n_m. However, the link qualities tend to vary over time, and the relaying candidates may have different duty cycles, processing and queuing delays, and mobility patterns, which have significant impacts on the performance of the cooperative communication.

If the network topology changes due to node mobilities, channel interferences, shadowing, and multi-path fading, the re-selection of the relaying candidates will be invoked by the error messages passed from the network layer protocol. Then, the sets of relaying candidates are either re-selected for all the intermediate routers, or only re-selected at the area where link failure occurs.

C. Optimal Relay Assignment

Based on the mathematical properties, the QoS attributes of the end-to-end delay T_{end} and packet delivery ratio P_{end} are additive and multiplicative metrics, respectively. For instance, for a H-hop route, the end-to-end delay is the accumulated delay of each link along the route, i.e., $T_{end} = \sum_{i=0}^{H} T_i$, where T_i is the delay experienced at link i. The end-to-end packet delivery ratio is a product of the packet delivery ratio of each link along the route, i.e., $P_{end} = \prod_{i=0}^{H} P_i$, where P_i is the packet delivery ratio at link i.

For simplicity, we assign identical QoS constraints on each link along the multi-hop route. That is, to meet the end-to-end QoS requirements on T_{end} and P_{end} for a H-hop route, each link should satisfy the metrics of link qualities that $T_i \leq \frac{1}{H} T_{end}$ and $P_i \geq P_{end}^{\frac{1}{H}}$.

The optimal relay assignment scheme is based on a multi-agent reinforcement learning algorithm, i.e., each node is implemented with a Q-learning algorithm [18], a model-free method which learns the value of a function $Q(s, a)$ to find the optimal decision policy. In *EEARS*, the Q-value represents the quality of cooperative communication, i.e., the contribution that the selected relay may make in terms of packet outage probability, spectral and energy efficiency.

The feedback can be regarded as an immediate reward from the environment in the context of reinforcement learning, which represents the quality of the cooperative communication, i.e., the contribution that the selected relay has made in terms of outage probability, spectral and energy efficiency. Each relay then uses the immediate reward and the long-term expected reward to update the corresponding Q-value, which has influence on the future decisions on optimal relay assignment.

D. Cooperative Transmission Scheme

In the two phase operations of *EEARS*, the relaying transmission phase will be invoked only when the packet transmission fails in the direct transmission phase.

In the direct transmission phase, the sender transmits a data packet to the receiver and all of the relays, then the receiver and all the relays attempt to decode the packet. If the receiver can decode the packet successfully, it will send an ACK packet to the sender and all the relaying candidates, and the packet transmission is finished; otherwise, the receiver stores the received signal and defers the decoding to the relaying transmission phase, and sends a $NACK$ packet to the sender and all the relaying candidates, notifying the packet transmission failure in this phase.

The relaying transmission phase will be invoked in case a $NACK$ is received by the relaying candidates in the direct transmission phase, or neither an ACK nor a $NACK$ is received within a certain amount of time. The relaying candidates are then aware of the failure of packet transmission in the direct transmission phase, and one of the relays, among those which successfully received and decoded the data packet

in the direct transmission phase, will re-encode and retransmit the packet to the receiver. The receiver combines the signals received in both of the direct transmission and the relaying transmission phases and applies MRC for packet decoding. If the receiver can decode the combined signal successfully, it sends an ACK to the sender and all the relaying candidates; otherwise, it sends a $NACK$ packet.

VI. ALGORITHM DESIGN OF OPTIMAL RELAY SELECTION

A. Design Objective: Analysis on Benefits and Costs of Cooperative Communication

One of the main benefits of utilizing cooperative communications is combating the multiple fading effects in wireless networks, and thus improve the network performance, in terms of transmission reliability, robustness, adaptivity, network throughput and lifetime.

However, the use of cooperative communications also associates with certain costs because of conducting extra tasks of signal processing, packet receiving and transmitting. For the overhead introduced, the relays' transmission is a major concern in designing cooperative communication protocols, as the nodes are often battery-powered and WSNs usually encounter severe resource constraints on power supply. Furthermore, the relays' transmission also increases the network's background noise, and thus has negative effect on other nodes' decoding performance, even those nodes are out of the communication range of the relay. Thus, using a lower transmitting power is not only for energy saving, but also for the purpose of reducing the concurrent transmission interferences may caused.

Therefore, the relaying candidate which can make the most contribution in improving the network performance, in terms of transmission reliability and spectral efficiency, and requires the minimal transmitting power, should be chosen as the optimal relay.

B. Algorithm Design

In the context of the multi-agent reinforcement learning framework, for an agent, the state, action, and reward are defined as follows.

a) State: From an agent's perspective, the state is the locally observed network configuration and events (data flows in connection). The state evolves with the actions taken by the candidates and the state spaces are defined as

$$S = \{V_{rc}, V_f\}, \tag{1}$$

where $V_{rc} = \{rc_0, rc_1, ..., rc_i, ..., rc_M\}$ is the set of relaying candidates, and $V_f = \{f_0, f_1, ..., f_i, ..., f_S\}$ is the set of data flows in connection.

b) Action: When a relaying candidate observes data flows in connection, the candidate can either remain silent, or choose an appropriate transmitting power level to participate in the communication. The action spaces are defined as

$$A = \{(f_i, a_i)\}, i = 0, 1, .., i..., S, \tag{2}$$

where $a_i = (CC, P_t)$. For an action a_i, $CC = 0$ stands for that the candidate rc_i does not cooperate with the communication of the data flow f_i, and $CC = 1$ represents that rc_i cooperates with the data flow f_i with the transmitting power level P_t, which is chosen from the range of $[P_0, P_1, ..., P_t, ..., P_T]$.

c) Reward function: The reward function is designed to reflect the contribution (improvement on link qualities) made by the relay, as well as the costs associated (energy consumed by the relay). The reward function is defined as

$$R = \omega_1 \frac{(P_l^E - P_l^R)}{P_l^R} + \omega_2 \frac{1}{N_d} \sum_{i=0}^{N_d} \frac{(T_{l_i}^R - T_{l_i}^E)}{T_{l_i}^R}$$
$$+ \omega_3 \frac{1}{N_d} \sum_{i=0}^{N_d} \frac{(P_m - P_t)}{P_m} \cdot \left(\frac{L_d^i}{R_d} \right), \qquad (3)$$

where P_l^E and P_l^R are the experienced and required 1-hop packet delivery ratio, respectively. N_d is the total number of packets received by the receiver in the data session. $T_{l_i}^R$ and $T_{l_i}^E$ are the required and experienced 1-hop delay for packet i, respectively. P_m is the medium power level and $P_m = \frac{P_0 + P_T}{2}$. P_t is the selected transmission power level for relaying transmission. L_d^i is the length of the packet i in bits (including overhead), and R_d is the bit transmission rate of the transceiver in bps (bits per second).

ω_1, ω_2, and ω_3 are the weighting factors for the metrics of packet delivery ratio, delay, and energy efficiency, respectively. The values of the weighting factors can be adjusted to adapt to the data session's QoS requirements.

In Eq. 3, the first term represents the improvement on the metric of packet delivery ratio made by using the relaying transmission. The second term stands for the improvement on the metric of 1-hop transmission delay. The third term represents the metric of energy efficiency, compared with using a fixed transmission power level.

The reward represents a weighted quality of the cooperative communications, when rc_i is chosen as the optimal relay to participate in the communication with the transmission power level of P_t in the data session.

The updating of Q-value iterates in each relay assignment procedure. Distributed value function - distributed reinforcement learning algorithm (DVF-DRL) [19] is used in the updating iteration.

For the 1-hop communication between n_l and n_m, at iteration t, a relaying candidate rc_i is selected as the optimal relay to retransmit the data packet to the receiver n_m. The Q-value is updated as in (4).

$$Q_{rc_i}^{t+1}(s_{rc_i}^t, a_{rc_i}^t) = (1 - \alpha)Q_{rc_i}^t(s_{rc_i}^t, a_{rc_i}^t) + \alpha(r_{rc_i}^{t+1}(s_{rc_i}^{t+1}) +$$
$$\gamma w(rc_i, n_m) \max_{a_{n_m} \in A_{n_m}} Q_{n_m}^t(s_{n_m}^t, a_{n_m}^t) +$$
$$\gamma \sum_{rc_i' \in V_{rc_i'}} w(rc_i, rc_i') \max_{a_{rc_i'} \in A_{rc_i'}} Q_{rc_i'}^t(s_{rc_i'}^t, a_{rc_i'}^t)), \quad (4)$$

where α is the learning rate, which models the updating rate of the Q-value. r denotes the immediate reward of execution

Parameters	Value
Number of sensor nodes	100
Simulation area	200 m × 200 m
Wireless channel model	Raleigh shadowing wireless model
Path loss exponent	2.4
Collision model	Additive interference model
Physical and MAC layer	IEEE 802.15.4 standard
Packet length	40 bytes
Transmitting power level	[-25, -15, -10, -7, -5, -3, -1, 0] dbm
Node's initial energy	12 J
Data transmission rate	250 kbps
Simulation time	400 s
Number of simulation runs	10
ω_1	0.4
ω_2	0.4
ω_3	0.2
α	0.1
γ	0.5
$w(rc, n_m)$	0.5
$w(rc_i, rc_i')$	$\frac{1}{2I}$, I is the number of routers.

of the action, i.e., the contribution that the selected relay has made. The weight of future reward is defined by γ. V_{rc_i}' is the set of nodes within rc_i's neighborhood which are selected as routers by other data flows in the network. $w(rc_i, n_m)$ and $w(rc_i, rc_i')$ are the weighting factors for modeling the expected reward at the receiver, and the values of the routers in V_{rc_i}', respectively.

Eq. 4 shows that the candidate rc_i's Q-value is a weighted sum of rc_i's Q-value at the current state, the action's immediate reward, the maximum Q-value of the receiver n_m, and the values of all the nodes selected by other data flows as routers within rc_i's neighborhood.

VII. PERFORMANCE EVALUATION

To study the network performance of *EEARS*, We compare it with *CRP* [12], a cooperative routing protocol which selects two relays from the neighboring nodes based on the link SNR and two-hop neighborhood information.

A. Simulation Environment

We simulate a WSN where 100 sensor nodes are randomly distributed in a 200m × 200m area. A CBR (constant bit rate) traffic with $5p/s$ is used as the communication pattern, and the source and destination nodes are chosen randomly in each simulation run. The number of background traffic data flows varies from 1 to 10 in the simulations. We define the lifetime is the time when the first node exhausts its battery's energy.

Castalia [20] wireless sensor network simulator (the data link layer is modified to facilitate *MRC* combining and decoding), which is built based on OMNeT++ [21] discrete event simulation platform, is used as the simulation environment.
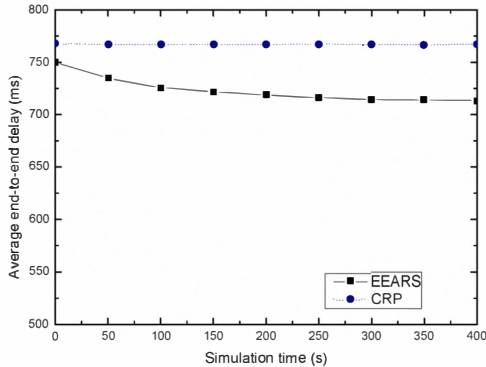
Table I lists the detailed simulation parameters.

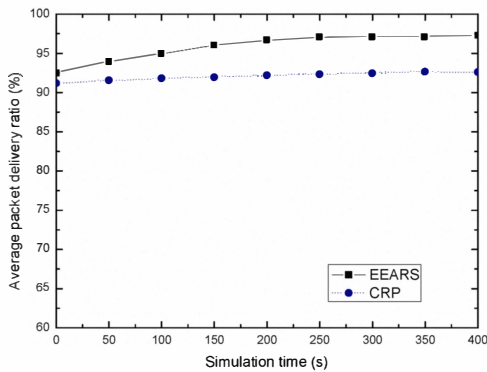Fig. 3.   Average end-to-end delay

Fig. 5.   Average network throughput with a varying number of data flows

Fig. 4.   Average packet delivery ratio

Fig. 6.   Average network lifetime with a varying number of data flows

B. Comparison with CRP

The average end-to-end delay and packet delivery ratio with a background traffic of 4 CBR data flows are shown in Fig. 3 and Fig. 4, respectively.

The simulation results show that *EEARS* outperforms *CRP* in both of the two metrics. The reason is that *EEARS* utilizes cooperative communications only in case the direct transmission fails, i.e., the relays will be involved in the cooperative transmission only when the link between the sender and the receiver is of poor quality. In contrast, *CRP* selects two relays for each packet transmission, which increases the probabilities of channel access contention and packet collision, and thus leads to lower spectral efficiency. Moreover, *CRP* considers SNR as the unique relay selection criterion, which is not sufficient in dynamic WSNs. For instance, a relay with high SNR to both the sender and the receiver may suffer severe channel access contention and/or processing and queuing delay, or it may operate in a low duty cycle for energy conservation. For *EEARS*, the relay which can improve the performance on both of the outage probability and spectral efficiency will be selected as the optimal relay, by strengthening the optimal decision and weakening the sub-optimal decisions of relay assignment. Furthermore, the use of adaptive transmission power level in *EEARS* effectively reduces the interferences caused by concurrent transmissions, and thus nodes in *EEARS* can achieves a better decoding performance.
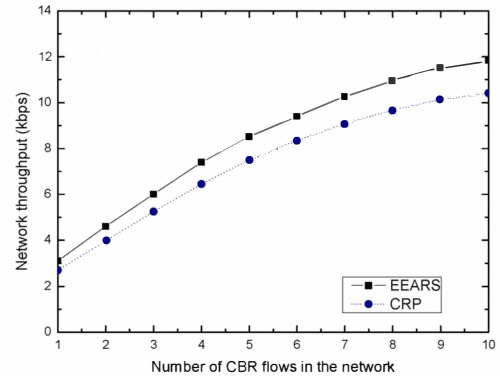
The aggregated network throughput and network lifetime with a varying number of *CBR* data flows are shown in Fig. 5 and Fig. 6, respectively.

We can observe that *EEARS* and *CRP* behave similarly on both of the two QoS metrics when the number of data flows is low (from the number of 0 to 5 CBR data flows). However, when the number of data flows increases, *EEARS* performs better than *CRP*, i.e., *EEARS* has a higher network throughput and a longer network lifetime. The reason is that when the number of data flows increases, the channel access contention becomes higher (nodes are more likely to contend with other nodes to access the channel), as the wireless medium is shared by all the nodes in the network. Moreover, the network background noise also becomes higher when more nodes transmit signals. In this situation, efficient allocation of network resources and optimal relay assignments are more important than WSNs with lighter traffic load. By adaptive selecting optimal relays and varying the transmission power level, *EEARS* can reduce concurrent transmission interferences, probability of packet collision, and energy consumption, and thus achieves better overall network performance than *CRP*.

Fig. 7 illustrates the satisfying level of soft QoS provisioning for the measured route, when the background traffic varies.

The result shows that *EEARS* can achieve a higher satisfying level than *CRP*, especially when the number of background data flows increases. This is because that for *EEARS*, the use of reinforcement learning based algorithm allows a more efficient handling of network dynamics. When the background
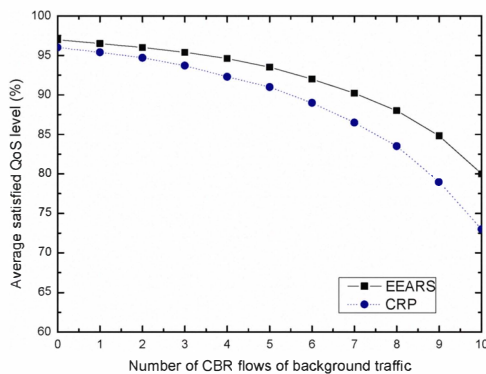
Fig. 7.   Impact of background traffic on satisfying level of soft QoS

traffic increases, nodes are more likely to contend with others to access the channel, or being selected as routers/relays by other data flows, due to the shared and contention nature of the wireless medium. In *CRP*, the source explicitly assigns optimal relays based on the measured signal's SNR without consideration of the states of the candidates, thus the scheme lacks of the flexibility of handling network dynamics. In comparison, *EEARS* is more adaptive in relay selection since the optimal relay is distributedly determined by each pair of adjacent routers along the route through experiences and rewards, and thus achieves the capability of handling network dynamics. The simulation results also verify that *EEARS* is more adaptive and flexible than *CRP* in dynamic environments.

VIII. Conclusions and Future Research

In this paper, we have reviewed the recent research on optimal relay assignment for cooperative communications, and investigated the use of cooperative communications with adaptive relay selection for soft QoS provisioning in resource-constrained WSNs. To further improve the network performance, the proposed optimal relay assignment scheme, *EEARS*, is integrated with a power control scheme, which is based on a multi-agent reinforcement learning framework. Simulation results have shown that *EEARS* can effectively improve the network performance on transmission reliability, spectral and energy efficiency, and thus increase the satisfying level of soft QoS provisioning for WSNs. In dynamic environments, *EEARS* also performs well in terms of a number of QoS metrics.

In future research, adaptive assignment of QoS constraints will be considered in the route discovery and establishment procedure. Moreover, we will examine the joint use of reinforcement learning and cooperative game theory in efficient allocation of network resources for differential QoS provisioning and network optimization in WSNs, and achieving system fairness in assigning optimal relays as well.

Acknowledgment

References

[1] Q. Zhang and Y.-Q. Zhang, "Cross-layer design for QoS support in multihop wireless networks," *Proceedings of the IEEE*, vol. 96, no. 1, pp. 64–76, Jan. 2008.

[2] P. Mohapatra, J. Li, and C. Gui, "QoS in mobile ad hoc networks," *IEEE Wireless Communications*, vol. 10, no. 3, pp. 44–52, Jun. 2003.

[3] A. Veres, A. Campbell, M. Barry, and S. Li-Hsiang, "Supporting service differentiation in wireless packet networks using distributed control," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 2081–2093, Oct. 2001.

[4] A. K. Sadek, Y. Wei, and K. J. Liu, "When does cooperation have better performance in sensor networks?" in *Proc. IEEE The 3rd Sensor and Ad Hoc Communications and Networks (SECON'06)*, Reston, Virginia, USA, Sep. 2006, pp. 188–197.

[5] A. K. Sadek, W. Su, and K. J. R. Liu, "Multinode cooperative communications in wireless networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 341–355, Jan. 2007.

[6] D. G. Brennan, "Linear diversity combining techniques," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 331–356, Feb. 2003.

[7] A. Ibrahim, A. Sadek, W. Su, and K. Liu, "Cooperative communications with relay-selection: when to cooperate and whom to cooperate with?" *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2814–2827, Jul. 2008.

[8] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 659–672, Mar. 2006.

[9] T. E. Hunter and A. Nosratinia, "Distributed protocols for user cooperation in multi-user wireless networks," in *Proc. IEEE the 47th annual Global Telecommunications Conference (GLOBECOM'04)*, Dallas, Texas, USA, Nov. 2004, pp. 3788–3792.

[10] M. Chen, X. Liang, V. C. Leung, and I. Balasingham, "Multi-hop mesh cooperative structure based data dissemination for wireless sensor networks," in *Proc. The 11th International Conference on Advanced Communication Technology (ICACT'09)*, Phoenix Park, Korea, Feb. 2009, pp. 102–106.

[11] G. Jakllari, S. V. Krishnamurthy, M. Faloutsos, P. V. Krishnamurthy, and O. Ercetin, "A cross-layer framework for exploiting virtual MISO links in mobile ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 579–594, Jun. 2007.

[12] Y. Lin, J.-H. Song, and V. W. Wong, "Cooperative protocols design for wireless ad-hoc networks with multi-hop routing," *Mobile Networks and Applications*, vol. 14, no. 2, pp. 143–153, Apr. 2009.

[13] Z. Zhou, S. Zhou, J.-H. Cui, and S. Cui, "Energy-efficient cooperative communication based on power control and selective single-relay in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 8, pp. 3066–3078, Aug. 2008.

[14] A. Bletsas, H. Shin, and M. Z. Win, "Cooperative communications with outage-optimal opportunistic relaying," *IEEE Transactions on Wireless Communications*, vol. 6, no. 9, pp. 3450–3460, Sep. 2007.

[15] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[16] L. P. Kaelbling, M. L. Littman, and A. P. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, May 1996.

[17] M. K. Simon and M.-S. Alouini, Eds., *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis.* USA: John Wiley & Sons, Inc, 2000.

[18] R. S. Sutton and A. G. Barto, Eds., *Reinforcement Learning: An Introduction.* Cambridge, MA, USA: MIT Press, 1998.

[19] J. Schneider, W.-K. Wong, A. Moore, and M. Riedmiller, "Distributed value functions," in *Proc. The 16th International Conference on Machine Learning*, Bled, Slovenia, Jun. 1999, pp. 371–378.

[20] (2010) The Castalia website (http://castalia.npc.nicta.com.au/). [Online]. Available: http://castalia.npc.nicta.com.au/

[21] (2010) The OMNeT++ website (http://www.omnet++.org/). [Online]. Available: http://www.omnet++.org/