Cooperative Communications with Relay Selection for QoS Provisioning in Wireless Sensor Networks

Xuedong Liang*, Ilangko Balasingham† and Victor C.M. Leung‡
*Rikshospitalet University Hospital, University of Oslo, Oslo, Norway N-0027
Email: xuedongl@medisin.uio.no
†Dept. of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway N-7491
Email: ilangkobat@iet.ntnu
‡Dept. of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada V6T 1Z4
Email: vleung@ece.ubc.ca

***Abstract*—Cooperative communications have been demonstrated to be effective in combating the multiple fading effects in wireless networks, and improving the network performance in terms of adaptivity, reliability and network throughput. In this paper, we investigate the use of cooperative communications with adaptive relay selection for resource-constrained wireless sensor networks, and propose *QoS-RSCC*, a QoS-support multi-agent reinforcement learning based relay selection scheme for cooperative communications. In *QoS-RSCC*, optimal relays, in terms of outage probability and channel efficiency, are selected distributedly from multiple relaying candidates for the intermediate routers along the multi-hop route, without the needs of prior knowledge of the wireless network model and centralized control. We compare the network performance of *QoS-RSCC* with *CRP* [1], and investigate the impacts of network traffic load, channel bit error rate, and node's mobility on the network performance. Simulation results show that *QoS-RSCC* can achieve a near-optimal performance on both diversity gains and channel efficiency, and fits well in dynamic environments.**

I. Introduction

In recent years, cooperative communications have been proposed to exploit the spatial and time diversity gains in wireless networks by utilizing the broadcast nature of the wireless medium [2], [3]. Users in cooperative communication systems work cooperatively by relaying data packets for each other, and thus forming multiple transmission paths or virtual MIMO (multiple-input-multiple-output) system to the destination without the need of multiple antennas at each user.

In cooperative communication systems, the cooperative communication protocols which define cooperative partner assignment, power allocation, system fairness, coding and transmission schemes, are the keys to the performance of the cooperative communication systems. A large number of cooperative communication protocols have been proposed in the last years. In [4], a variety of cooperative diversity protocols were proposed, namely, amplify-and-forward, decode-and-forward, selection relaying, and incremental relaying cooperative protocols. The performance of the proposed protocols in terms of outage events and associated outage probabilities were evaluated respectively. Coded cooperation was proposed in [5], which integrates cooperation with channel coding and works by sending different parts of each user's code word via two independent fading paths. The authors in [6] implemented a cooperative strategy for mobile users in CDMA (code division multiple access) systems, where the mobile users are active and use different spreading codes to avoid multiple access interferences. CoopMAC, a cooperative MAC (medium access control) protocol for IEEE 802.11 wireless networks, was presented in [7]. CoopMAC can achieve performance improvements by exploiting both the broadcast nature of the wireless medium and cooperative diversity.

The conventional multi-node cooperative communication systems, where all the available relays actively participate in the communication by re-transmitting signals, have the potential of achieving full cooperative diversity gains. For instance, for a pair of sender and receiver with N relays participating in the cooperative communication, a packet transmission failure occurs only when all the $N + 1$ links (sender-receiver plus sender-relays-receiver) experience deep channel fading simultaneously. However, the channel efficiency of the multi-node cooperative communication system is much lower than the non-cooperative communication system, because the total number of $N + 1$ time slots (assuming CSMA/CA or TDMA is used as the underlying MAC protocol) is needed for the packet transmission. Besides, the packet transmission suffers extra delay due to the receiver deferring the packet decoding until all relays have completed their transmissions.

To achieve full cooperative diversity gains while still obtaining high channel efficiency and low transmission delay, selective single relay cooperative schemes, i.e., only one optimal relay is selected from multiple relaying candidates to cooperate with the communication, have been extensively studied in recent research. A cooperative relay framework which accommodates the physical, MAC and network layers for wireless ad-hoc networks was proposed in [1]. In the network layer, cooperative diversity gains can be achieved by selecting two cooperative relays based on the average link SNR (signal-to-noise ratio) and the two-hop neighborhood information. The authors in [8] proposed a cooperative protocol based on relay selection technique using the channel state information (CSI) at the source and the relays. The optimal relay is the node which has the maximum instantaneous scaled harmonic

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

mean function of its source-relay and relay-destination channel gains. In [9], distributed cooperative protocols, including random selection, received SNR selection and fixed priority selection, were proposed for cooperative partner selection. A cooperative communication scheme combining relay selection with power control was proposed in [10], where the potential relays compute individually the required transmission power to participate in the cooperative communication. The authors in [11] proposed a cooperative-based routing algorithm, namely, MPCR (minimum power cooperative routing). The analysis and simulation results showed that MPCR can choose the minimum-power route while guaranteeing the desired QoS, by making full use of the cooperative communications. Opportunistic single-relay-selection with decode-and-forward and amplify-and-forward protocols under an aggregate power constraint were presented in [12]. The results showed that the cooperative diversity benefits can be achieved even when cooperative relays act as passive relays and give priority to the transmission of a single opportunistic relay. An opportunistic relay selection scheme based on local measurements of the instantaneous channel conditions was proposed in [13].

Most of the existing cooperative protocols assume that full or partial CSI is available at the source, destination and all of the potential relays. However, significant communication overheads will be incurred in acquiring and disseminating of CSI to all of the cooperative participants, especially for the cooperative protocols, as in [8], [13], that instantaneous CSI is required at all the potential relays in calculation of relay selections. Furthermore, the use of CSI as the unique relay selection criterion is not sufficient in dynamic WSNs. The relaying candidates' working status, e.g., duty cycles, processing and queuing delays, and mobility patterns have significant impacts on the quality of cooperative communication, and should be taken into consideration in the relay selection. However, it is challenging to find optimal relay selection policies in dynamic WSNs, (e.g. when to cooperate? how to cooperate? and whom to cooperate with?) where the network state information is inherently imprecise and tend to vary. Thus, research on distributed, lightweight and adaptive relay selection scheme is still needed.

In this paper, we investigate the use of cooperative communications for QoS provisioning in wireless sensor networks (WSNs), and propose *QoS-RSCC*, a QoS support adaptive relay selection scheme for cooperative communications. In *QoS-RSCC*, for each pair of routers along the multi-hop route, an optimal relay in terms of packet outage probability and channel efficiency is distributedly selected from multiple relaying candidates according to the QoS requirements of the data session. The optimal relay will participate in the cooperative communication between the routers by re-transmitting the packet. The relay selection scheme is based on a multi-agent reinforcement learning algorithm, i.e., the optimal relay selection policy is learned collaboratively by the routers from a series of trial-and-error interactions with the dynamic network, without the needs of prior knowledge of the network model and centralized control. Simulation results show that *QoS-RSCC* is effective in QoS provisioning for WSNs and can achieve a near-optimal performance on cooperative diversity gain and channel utilization efficiency.

The rest of the paper is organized as follows. The background information of reinforcement learning and its applications in WSNs are provided in Section II. Section III describes the architecture overview and design issues of *QOS-RSCC*. The design and implementation of the reinforcement learning algorithm are illustrated in Section IV. The performance analysis is presented in Section V. Finally, Section VI concludes the paper and discusses the future research directions.

II. Reinforcement Learning and Its Applications in WSNs

Reinforcement learning provides a framework in which an agent can learn control policies in dynamic environment based on experiences and rewards. In the standard reinforcement learning model, an agent is connected to the environment via perception and action, as shown in Fig. 1. 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 [14].

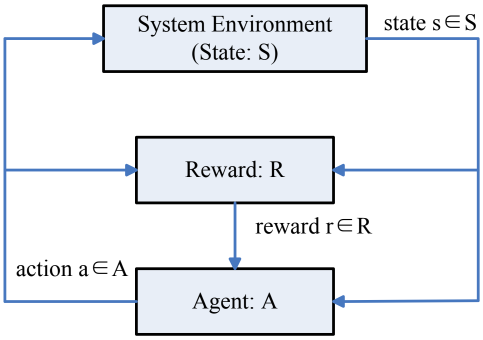

Fig. 1. A standard reinforcement learning model

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 [14].

Multi-agent systems (MASs) are systems that multiple agents are connected to the environment and may take actions to change the states of the environment. WSNs can be regarded as multi-agent systems, where sensor nodes can be considered

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

as agents, the wireless medium and data flows can be regarded as environment. The agents may take actions (e.g., 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.

III. Cooperative Communications with Adaptive Relay Selection

A. Architecture Overview

We consider a WSN with multi-hop communications, where each of the sensor nodes may establish route to the sink for data packet sending. Decode-and-forward [4] is used as the cooperative transmission scheme. To achieve optimal decoding performance, MRC (maximal ratio combining) [15] is utilized at the receiver for packet decoding by combining the multiple signals received from the sender and the relay(s). AODV (Ad hoc On-Demand Distance Vector) and CSMA/CA are employed as the underlying network and MAC layer protocols, respectively.

In *QoS-RSCC*, when a route is established, for each pair of adjacent routers along the route, a number of nodes will be selected as the set of relaying candidates for the cooperative communication between the two routers. The cooperative transmission scheme operates in two phase, namely, direct transmission and relaying transmission phases. If the packet transmission fails in the direct transmission phase, the relaying transmission phase will be invoked, and an optimal relay among all the relaying candidates is selected to re-encode and re-transmit the packet to the receiver. Then, the receiver combines the received signals from both the sender and relay for optimal packet decoding.

Each time an optimal relay is selected to participate in the cooperative communication, the decision maker of the relay assignment will receive an immediate reward given by the environment, which represents the quality of the cooperative communication. The decision maker 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. When the algorithm reaches convergence, the routers 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 and channel efficiency will be more likely to be selected as the optimal relay.

B. Relaying Candidate Selection

In *QoS-RSCC*, the relaying candidate selection is integrated with the route finding mechanism, and the selection is valid during the lifetime of the established route. The route finding mechanism in *QoS-RSCC* 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-satisfied route.

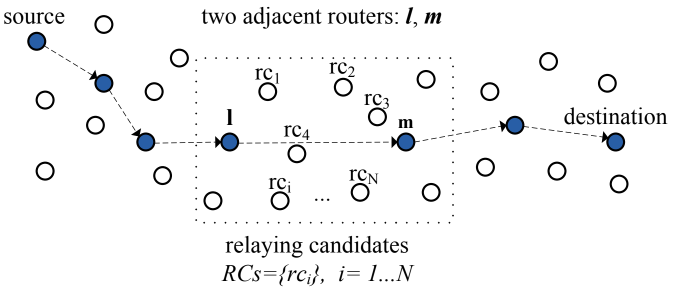

Fig. 2. Cooperative communications with relay selection in multi-hop WSNs

When a source has packets to send, it will initiate a route discovery procedure. Once the route from the source to the destination is discovered, for two adjacent routers along the established route, e.g., l and m (a pair of one-hop sender and receiver), as shown in Fig. 2, a set of relaying candidates (RCs) is selected as the cooperative partners for l and m.

A node determines that it is a relaying candidate (rc) for l and m, if it

- has heard the $RREQ$ transmitted by l, and
- has heard the $RREP$ replied by m, and
- has not been selected by l as the next hop router

in the route discovery procedure.

Formally, RCs =$N_l \cap N_m$, where N_l and N_m are the sets of the immediate neighboring nodes of l and m, respectively. Thus, for each relaying candidate $rc_i \in RCs$, rc_i is the common immediate neighbor of both l and m. Ideally, all of the relaying candidates are connected to both l and 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 established route is broken due to network topology changes, wireless channel interferences and node mobilities, RERR messages will be broadcasted or unicasted by the routers which have detected the route failures, as defined in the AODV routing protocol. Upon receiving the RERR messages, the re-selection of the relaying candidates will be invoked by the involved routers. Then, the sets of relaying candidates are either re-selected for all the intermediate routers from the source to the destination, or re-selected locally at the area where the link failure occurs, depending on the route repairing mechanism used in the AODV protocol.

C. Adaptive Optimal Relay Assignment

Once the route is established and the set of relaying candidates is selected for each pair of adjacent routers, one of the relays should be chosen to participate in the cooperative communication between the routers, if the direct link between the two routers cannot satisfy the QoS requirements of the data session.

In *QoS-RSCC*, for two adjacent routers, the decision of optimal relay assignment is made by the sender, i.e., the sender chooses an optimal relay from the set of relaying candidates as a cooperative partner. For simplicity, we assume that the number of relaying candidates e.g. N, is identical for each pair of routers. Thus, for l and m, as shown in Fig. 2, there are N options available for the optimal relay assignment.

The optimal relay assignment scheme is based on a multi-agent reinforcement learning algorithm, i.e., the router is implemented with a Q-learning algorithm [16], a model-free method which learns the value of a function $Q(s, a)$ to find the optimal decision policy. In *QoS-RSCC*, the Q-value represents the quality of cooperative communication, i.e., the expected contribution that the selected relay can make in terms of packet outage probability and channel utilization efficiency.

Each time a relay is selected to participate in the communication by re-transmitting the packet, the receiver combines the signal received from both the sender and the selected relay for packet decoding. Then the receiver sends a feedback to the sender, which contains of the information of the improvements on packet outage probability and channel utilization efficiency. The feedback can be regarded as an immediate reward (could be positive or negative) from the environment in the context of reinforcement learning, which represents the quality of the cooperative communication, i.e., the contribution of the selected relay has made in terms of outage probability and channel efficiency. The sender then uses the immediate reward and the long-term expected reward in the remaining path to update the corresponding Q-value, which will influence the future decisions of optimal relay assignment.

D. Cooperative Transmission Scheme

The cooperative transmission scheme in *QoS-RSCC* operates in two phases, namely, direct transmission and relaying transmission phases. 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 of the relays attempt to decode the data packet. If the receiver can decode the packet successfully, it will send an ACK packet to the sender, 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, notifying the packet transmission failure in this phase.

The relaying transmission phase will be invoked in case a $NACK$ is received by the sender in the direct transmission phase, or neither an ACK nor a $NACK$ is received within a certain amount of time. The sender is then aware of the failure of packet transmission in the direct transmission phase, and one of the relays, among those which successfully received the data packet in the direct transmission phase, will be selected by the sender to re-encode and re-transmit the packet to the receiver. The receiver combines the signals received in both of the direct transmission and the relaying transmission phases and decodes the signal. If the receiver can decode the combined signal successfully, it sends an ACK to the sender; otherwise, it sends a $NACK$ packet.

The algorithm will reach convergence after a certain amount of time, depending on the network size, node mobility and density. Then, the sender is able to use the learned policy to take appropriate actions, i.e., the relaying candidate which can contribute most in improving the performance of cooperative communication will be more likely to be selected as the optimal relay.

To adapt to the dynamic nature of WSNs, *QoS-RSCC* explores the environment with a certain probability ε, namely ε-greedy method [16]. That is, with the probability of $1 - \varepsilon$, the candidate which is expected to be able to make the most contributions will be selected as the optimal relay; and with the probability of ε, a randomly chosen candidate will be selected.

Thus, without maintaining and disseminating CSI at the source, destination, intermediate routers and relays, or using complicated channel state prediction algorithms, optimal relays can be selected through experiences and rewards in dynamic WSNs.

IV. Q-learning Algorithm Design and Implementation

In the context of reinforcement learning, for a pair of sender and receiver, the states, actions, and rewards are defined as follows.

a) State: The states are the data session's quantized satisfaction/violation levels of the QoS metrics, which evolve with the actions taken by the agents.

$$S = \left\{ L_i^{QoS} \right\}, \quad i = 1...K \tag{1}$$

b) Action:

$$A = \{a_i\}, \quad i = 1...N \tag{2}$$

The execution of a_i represents that relay rc_i is selected to participate in the cooperative communication.

c) Reward function:

$$Rwd = \begin{cases} \omega_1 \frac{1-P^o_{s,rc_i,rec}}{1-P^o_{s,rec}} - \omega_2 \frac{T_{ACK}-T_{TX}}{T_{AVR}} & \text{ACK received (a)} \\ -\omega_2 \frac{T_{NACK}-T_{TX}}{T_{AVR}} & \text{NACK received (b)} \\ -R_c & \text{neither ACK nor NACK received (c)} \end{cases} \tag{3}$$

Eq. (3a) is used to calculate the reward when the packet transmission is successful. The first term represents the performance improvement on outage probability, and the second term represents the channel utilization efficiency, respectively. $P^o_{s,rc_i,rec}$ is the outage probability of the cooperative communication wherein rc_i is selected as the optimal relay to re-transmit the data packet to the receiver. $P^o_{s,rec}$ is the outage probability of the direct transmission between the sender and the receiver. ω_1 and ω_2 are the weighting factors for the metrics of outage probability and channel efficiency, respectively. The values of the weighting factors can be adjusted to adapt to the data session's QoS requirements. T_{ACK} and T_{TX} are the ACK receiving time and the data packet re-transmitting time at the sender and the selected relay rc_i, respectively.

$P^o_{s,rec}$ and $P^o_{s,rc_i,rec}$ can be calculated as in (4) and (5), respectively [11], [15].

$$P^o_{s,rec} = 1 - \exp(-\frac{2^R - 1}{SNR_{s,rec}}) \tag{4}$$

$$P^o_{s,rc_i,rec} = 1 - \exp(-\frac{2^R - 1}{SNR_{s,rec} + SNR_{rc_i,rec}}) \tag{5}$$

R is the data transmission rate of the transceiver in bits per second. $SNR_{s,rec}$ and $SNR_{rc_i,rec}$ are the SNR of the data packets at the receiver which are received from the sender and the selected optimal relay rc_i, respectively. The received packets' SNR can be acquired from the underlying data link layer protocol employed at the receiver. Thus, the outage probability $P^o_{s,rec}$ and $P^o_{s,rc_i,rec}$ can be calculated at the receiver, and encapsulated in the ACK or $NACK$ packet, then sent back to the sender for the evaluation of the quality of the relay assignment.

T_{AVR} is the average amount of time needed for the data packet transmission between the relaying candidate rc_i and the receiver, without any channel contention, processing and queuing delays. T_{AVR} can be calculated as

$$T_{AVR} = T_{BO} + T_{P_D} + T_{TA} + T_{P_{ACK}} + T_{IFS} \tag{6}$$

where T_{BO} is the average backoff time at rc_i without any channel contention, and the value is determined by the underlying MAC layer protocol. T_{P_D} is the data packet transmission time and $T_{P_D} = \frac{l_d}{R}$, where l_d is the packet size (including overheads) in bits. T_{TA} and T_{IFS} are the transceiver's transmitting to receiving turnover time and the inter frame space (IFS), respectively, which are defined by the underlying communication protocols. $T_{P_{ACK}}$ is the amount of time for the ACK packet transmission.

The positive reward represents the quality of the cooperative communication in terms of outage probability and channel efficiency, when rc_i is involved in the cooperative communication.

Eq. (3b) and (3c) are used to calculate the rewards when the packet transmission fails, where T_{NACK} is the $NACK$ receiving time at the sender, R_c is a constant value which can be practically tuned in the simulation. The negative values represent the relative channel occupation time caused by the unsuccessful packet re-transmission, when rc_i is selected as the optimal relay.

The updating of Q-value iterates in each relay assignment procedure. Distributed value function - distributed reinforcement learning algorithm (DVF-DRL) [17] is used in the updating iteration.

For the 1-hop communication between l and m, at iteration t, a relaying candidate is selected as the optimal relay to re-transmit the data packet to the receiver. The Q-value corresponding to the selected relay is updated as in (7).

$$\begin{aligned} Q^{t+1}_l(s^t_l, a^t_l) = (1-\alpha)Q^t_l(s^t_l, a^t_l) + \alpha(r^{t+1}_l(s^{t+1}_l) + \\ \gamma w(l,m) \max_{a_m \in A_m} Q^t_m(s^t_m, a^t_m) + \\ \gamma \sum_{l' \in V_l} w(l,l') \max_{a_{l'} \in A_{l'}} Q^t_{l'}(s^t_{l'}, a^t_{l'})) \end{aligned} \tag{7}$$

where α is the learning rate, which models the updating rate of the Q-value. r denotes the immediate reward of execution of the action, i.e., the contribution that the selected relay has made in terms of outage probability and channel efficiency. The weight of future reward is defined by γ. V_l is the set of nodes within l's neighborhood which are selected as routers by other data flows in the network. $w(l,m)$ and $w(l,l')$ are the weighting factors for modeling the expected reward in the remaining path to the destination and the rewards of the routers in V_l, respectively.

Eq. 7 shows that the sender l's Q-value is a weighted sum of l's Q-value at the previous state, the action's immediate reward, the maximum Q-value of the receiver m, and the Q-values of all the nodes selected by other data flows as routers within l's neighborhood.

Considering the multi-hop route as shown in Fig. 2, node i_0 (the source) originates a packet towards the destination $dest$. Along the route, a number of nodes (denoted as $i_1, i_2, \ldots i_M$) are selected as the intermediate routers sequentially from the source to the destination. For node i_0, the Q-value is updated as in (8), recursively rewritten from (7).

$$\begin{aligned} & Q_{i_0}(s_{i_0}, a_{i_0}) = \\ & (1-\alpha) \sum_{n=0}^{M} (\alpha\gamma)^n \max_{a_{i_n} \in A_{i_n}} Q_{i_n}(s_{i_n}, a_{i_n}) \prod_{j=0}^{n} w(i_j, i_{j+1}) \\ & \qquad + \sum_{n=0}^{M} \alpha^{n+1}\gamma^n r(s_{i_n}, a_{i_n}) \prod_{j=0}^{n} w(i_j, i_{j+1}) \\ & + \sum_{n=0}^{M} (\alpha\gamma)^{n+1} \sum_{i' \in V_n} \max_{a_{i'} \in A_{i'}} Q_{i'}(s_{i'}, a_{i'}) \prod_{j=0}^{n} w(i_j, i^{'}_{j+1}) \\ & \qquad + \alpha^{M-1}\gamma^M Q_{dest}(s_{dest}, a_{dest}) \prod_{j=0}^{M} w(i_j, i_{j+1}) \end{aligned} \tag{8}$$

The first term is the weighted sum of the maximum Q-values of the routers sequentially from the source to the destination. The second term is the weighted sum of the immediate rewards achieved by the selected optimal relays. we can observe that both of the first and second terms are contributed by the intermediate routers and the selected optimal relays. The third term defines the weighted sum of the maximum Q-values of the routers for other data flows, i.e., the nodes which may have impacts on the performance of the measured data session, although they are not directly involved in the routing procedure. The weighted Q-value of the destination, which is set as a constant value is modeled in the last term.

Eq. 8 illustrates that although the source node only has locally observed network state information, and only communicates with its adjacent routers and the routers for other data session within its neighborhood, it can estimate the end-to-end QoS performance of the route to the destination, by calculating the weighted sum of its own immediate reward, the rewards that are expected to be achieved by the intermediate routers and the selected relays in the remaining path to the destination, and the rewards of all of the routers for other data flows in the network. Therefore, nodes in *QoS-RSSCC* can work in a cooperative manner by choosing actions to maximize the global rewards.

The pseudo code of the Q-learning algorithm is listed at Algorithm 1.

Algorithm 1 The Q-learning algorithm at sensor node l

```
  begin
  initialization
   select a number of nodes as the set of relaying candidates
   setup an action list corresponding to the set of relaying candidates
loop
    if it has packets to send/forward then
        transmit the packet to the receiver (direct transmission)
        if the packet transmission fails in the direct transmission phase then
            with the probability of $1 - \varepsilon$, assign the node with the highest
            Q-value as the optimal relay for packet re-transmission
        else
            with the probability of $\varepsilon$, randomly assign a node as the optimal
            relay for packet re-transmission
        end if
    end if
    if ACK/NACK received in the cooperative transmission phase then
        calculate the reward using Eq. (3a)/(3b) and update the Q-value
    else
        calculate the reward using Eq. (3c) and update the Q-value
    end if
    if RERR received then
        re-select the set of relaying candidates
    end if
end loop
```

V. Performance Evaluation

To study the network performance of *QoS-RSCC*, We compare it with *CRP* [1], a cooperative routing protocol which selects two relays from the neighboring nodes based on the link SNR.

A. Simulation Environment

We simulate a WSN where 100 sensor nodes are randomly distributed in a 200m × 200m area. We assume that the nodes are stationary in the simulations, except in the mobile scenario where 10 nodes are randomly chosen as mobile nodes and others keep stationary. 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.

TABLE I
Simulation Parameters

Parameters	Value
Number of sensor nodes	100
Simulation area	200 m × 200 m
Wireless channel model	Log shadowing wireless model
Path loss exponent	2.4
Collision model	Additive interference model
Mobility model	Random waypoint model
Physical and MAC layer	IEEE 802.15.4 standard
Packet length	40 bytes
Communication range	50 m
Data transmission rate	250 kbps
Simulation time	400 s
Number of simulation runs	10
ω_1	0.8
ω_2	0.2
R_c	0.2
ε	0.1
α	0.1
γ	0.5
$w(l, m)$	0.5
$w(l, l')$	$\frac{1}{2I}$, I is the number of routers

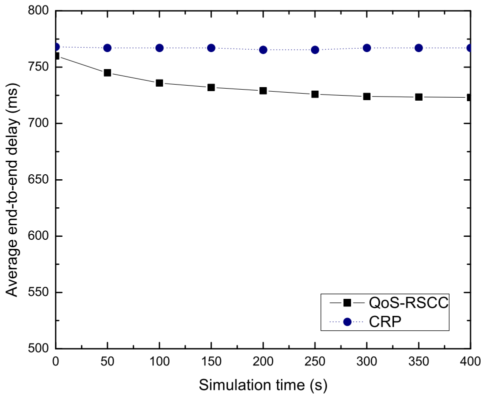

Fig. 3. Average end-to-end delay

Castalia [18] wireless sensor network simulator (the data link layer is modified to facilitate *MRC* combining and decoding), which is built based on OMNeT++ [19] discrete event simulation platform, is used as the simulation environment.

Table I lists the detailed simulation parameters.

B. *Comparison with* CRP

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

The simulation results show that *QoS-RSCC* outperforms *CRP* in both of the two metrics. The reason is that *QoS-RSCC* utilizes cooperative communication 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 low channel utilization

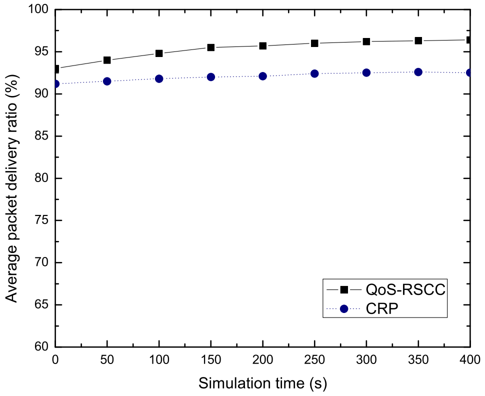

Fig. 4. Average packet delivery ratio

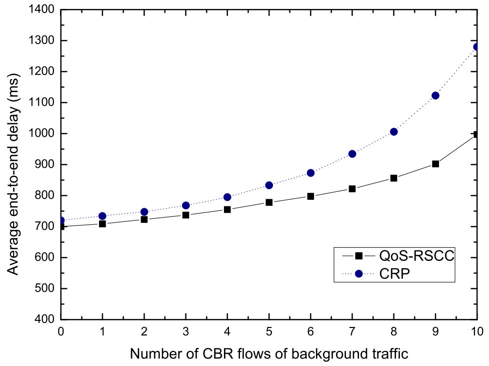

Fig. 5. Impact of background traffic on end-to-end delay

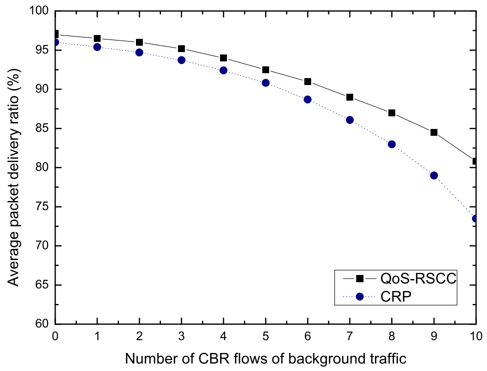

Fig. 6. Impact of background traffic on packet delivery ratio

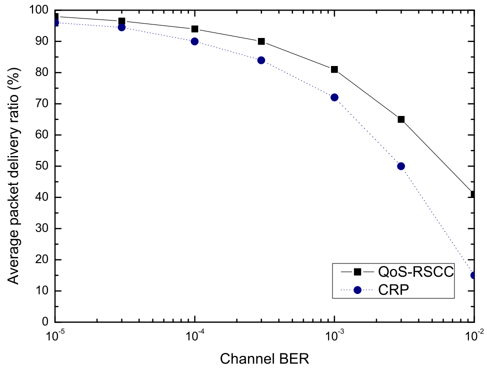

Fig. 7. The impact of channel BER on average packet delivery ratio

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 run in a low duty cycle for energy conservation. For *QoS-RSCC*, the relay which can improve the performance on both the outage probability and channel efficiency will be selected as the optimal relay, by strengthening the optimal decision and weakening the sub-optimal decisions of relay assignment. Thus, the relay selection in *QoS-RSCC* is more efficient than that in *CRP*.

Fig. 5 and Fig. 6 illustrate the average end-to-end delay and packet delivery ratio with the background traffic of varying number of *CBR* data flows, respectively.

We can observe that *QoS-RSCC* and *CRP* have similar performances on both of the two metrics when the background traffic is low (from the number of 0 to 5 CBR data flows). However, when the number of data flows of background traffic increases, *QoS-RSCC* performs better than *CRP*. The reason is that the background traffic has significant impacts on the measured source-destination data flow due to the shared and contention nature of the wireless medium. That is, nodes are more likely to contend with other nodes to access the channel, or to be selected as routers/relays by other data flows, when the background traffic increases. The simulation results also verify that *QoS-CC* is more adaptive and flexible than *CRP* in dynamic network conditions.

The impact of channel BER (bit error rate) and node mobility on average packet delivery ratio are shown in Fig. 7 and Fig. 8, respectively.

The simulation results show that *QoS-RSCC* performs better than *CRP*, especially when the channel BER becomes high and/or the network mobility level increases. It is because that in *CRP*, the source needs to explicitly assign the relays for each packet transmission, thus the scheme lacks of the flexibility of handling network dynamics. In comparison, *QoS-RSCC* is more adaptive in relay selection since the optimal

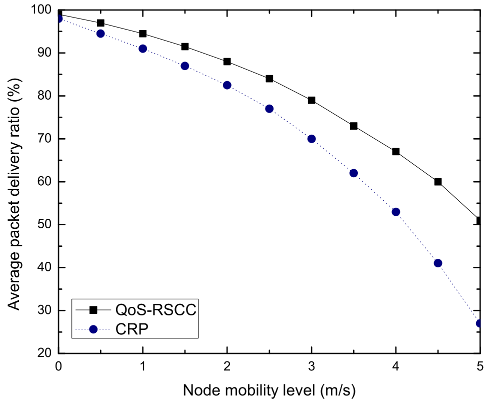

Fig. 8. The impact of node mobility on average packet delivery ratio

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

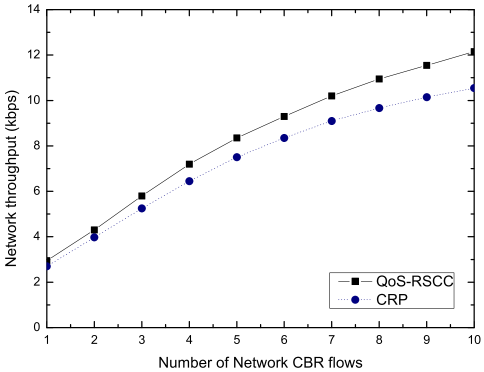

Fig. 9. The impact of CBR flows on average network throughput

relay is distributedly determined by each pair of adjacent routers along the route through experiences and rewards. The flexible nature of reinforcement learning allows it to adapt to the dynamic environment well, especially in networks with varying link qualities in mobile scenarios.

The aggregated network throughput with a varying number of data flows is shown in Fig. 9.

The result shows that *QoS-RSCC* can achieve higher network throughput than *CRP*, especially when the number of data flows increases. The reason is that the cooperative transmission scheme in *QoS-RSCC* is more efficient than that in *CRP*, i.e., the relaying transmission will be triggered only when the direct link quality is not good enough for packet transmission. For *CRP*, two relays are involved in the communication for each packet transmission. Thus, the network resources are consumed by unnecessary relaying transmissions, which increase the probabilities of channel access contention and packet collision, and thus has negative effects on the overall network performance.

We have also observed that for all the measured metrics in the simulations, *QoS-RSCC* performs better after the simulation runs for a certain amount of time (i.e., around 50s). This is because that there is a learning period in any learning based protocols, in which agents explore all the available decisions and estimate the decision qualities. When the learning procedure is finished, agents can take optimal actions according to the state information. so that the network performance are improved over time.

VI. Conclusions and Future Research

In this paper, we investigate the use of cooperative communications with adaptive relay selection for resource-constrained WSNs, and propose *QoS-RSCC*, a multi-agent reinforcement learning based optimal relay assignment scheme. Simulation results show that *QoS-RSCC* can achieve a near-optimal performance on cooperative diversity gains and performs well in terms of a number of QoS metrics in dynamic environments.

In future research, service differentiation and system fairness will be considered in the cooperative scheme design. Moreover, we will examine the use of adaptive cooperative coding scheme (e.g., channel coding) and employ power allocation scheme to improve the network performance and prolong the network lifetime.

Acknowledgment

This research is in the context of the EU project IST-33826 CREDO: Modeling and analysis of evolutionary structures for distributed services (http://www.cwi.nl/projects/credo/). This work was also supported in part by the Canadian Natural Sciences and Engineering Research Council under grant STPGP 322208-05.

References

[1] 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.

[2] A. Nosratinia, T. Hunter, and A. Hedayat, “Cooperative communication in wireless networks,” *IEEE Communications Magazine*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[3] Y.-W. Hong, W.-J. Huang, F.-H. Chiu, and C.-C. J. Kuo, “Cooperative communications in resource-constrained wireless networks,” *IEEE Signal Processing Magazine*, vol. 42, pp. 47–57, May 2007.

[4] 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.

[5] T. E. Hunter and A. Nosratinia, “Cooperation diversity through coding,” in *Proc. IEEE 2002 International Symposium on Information Theory (ISIT’02)*, Lausanne, Switzerland, Jun. 2002, p. 220.

[6] A. Sendonaris, E. Erkip, and B. Aazhang, “User cooperation diversity: System description, implementation aspects and performance analysis (Part I and Part II),” *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1948, Nov. 2003.

[7] P. Liu, Z. Tao, Z. Lin, E. Erkip, and S. hivendra Panwar, “Cooperative wireless communications: a cross-layer approach,” *IEEE Wireless Communications*, vol. 13, no. 10, pp. 84–92, Aug. 2006.

[8] 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.

[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] 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.

[11] A. Ibrahim, Z. Han, and K. Liu, “Distributed energy-efficient cooperative routing in wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3930–3941, Oct. 2008.

[12] 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.

[13] 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.

[14] 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.

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

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

[17] 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.

[18] (2009) The Castalia website. [Online]. Available: http://castalia.npc.nicta.com.au/

[19] (2009) The OMNeT++ website. [Online]. Available: http://www.omnet++.org/