ARTICLE IN PRESS

Computer Networks xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Computer Networks

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

SDRCS: A service-differentiated real-time communication scheme for event sensing in wireless sensor networks ☆

Yuyan Xue, Byrav Ramamurthy, Mehmet C. Vuran *

Department of Computer Science and Engineering, University of Nebraska, Lincoln, NE 68588, United States

ARTICLE INFO

Article history:
Received 5 October 2010
Received in revised form 20 June 2011
Accepted 21 June 2011
Available online xxxx

Keywords:
Service-differentiated real-time communication
Prioritized MAC
Receiver-contention-based forwarding
Distributed event sensing
Wireless sensor networks

ABSTRACT

Real-time communication is crucial for wireless sensor networks (WSNs) to accomplish collaborative event sensing tasks with specific timing constraints. In this work, a service-differentiated real-time communication scheme (SDRCS) is developed to provide soft real-time guarantees for event-based traffic in WSNs. SDRCS features a cross-layer packet forwarding design to integrate the real-time routing functionality with a novel prioritized medium access control scheme. Based on this design, SDRCS performs distributed packet traversal speed estimation for traffic classification and admission control. SDRCS also performs prioritized packet forwarding so that the routing decisions are locally performed for maximized packet traversal speed. SDRCS requires no extra hardware for localization, transmission power adaptation or multi-channel transmission. It also adapts well to network dynamics, such as channel quality and communication voids. Performance evaluations show that SDRCS significantly improves the on-time delivery ratio and service-differentiation granularity for mixed priority traffic flows in unsynchronized WSNs, compared with currently used communication schemes. SDRCS also provides higher end-to-end throughput in terms of supporting higher source data rates with tight end-to-end latency requirements.

1. Introduction

Wireless Sensor Networks (WSNs) have emerged as a new generation of distributed embedded systems that provide observations on the physical world at low cost and with high accuracy. Most current WSN applications, such as battlefield surveillance [2], industrial production control [3], and structural monitoring [4], pose various kinds of real-time constraints in response to the physical world [5–8,1]. In a typical real-time WSN application, as shown in Fig. 1, a number of sensor nodes are deployed to cover the sensing field. Predefined events can be detected by the nearby sensor nodes. The collected event information must be sent to a sink by a certain deadline so that the proper event response can be performed in a timely manner. Depending on the urgency of the event, the data packets associated with different events can be assigned different end-to-end deadline requirements. Only the packets that are delivered to the sink before the deadline are deemed useful. Mission-critical applications call for new service-differentiated real-time communication protocols designed for WSNs.

Providing end-to-end real-time guarantees in WSNs is extremely challenging, compared with the situation in traditional networks such as wireless local area networks (WLANs). First, WSNs utilize multi-hop communication over lossy channels. The dynamic network and channel conditions make firm real-time guarantees (e.g., a guaranteed packet reception rate and end-to-end transmission delay for a specific data rate) almost impossible. Second, the event-based traffic in WSNs may exhibit highly diverse real-time constraints [5] depending on the event locations

☆ An earlier version of this work was presented at the IEEE SenseApp '08 workshop [1].

* Corresponding author.
E-mail addresses: yxue@cse.unl.edu (Y. Xue), byrav@cse.unl.edu (B. Ramamurthy), mcvuran@cse.unl.edu (M.C. Vuran).

1389-1286/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2011.06.018

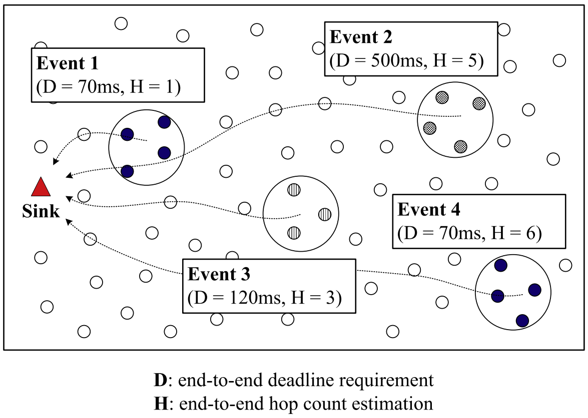


Fig. 1. A service-differentiated real-time application in event-based WSNs.

and urgencies. As a result, traditional flow-based traffic classification methods (e.g., dividing QoS traffic into data, voice, video, and control categories) may not be able to provide sufficient service-differentiation capability for the event traffic and ensure prioritized transmission. Third, the limited resources available to a sensor node restrict the design space of feasible WSN communication scheme. For example, location awareness and network synchronization may not be realistic assumptions for low-cost wireless sensor nodes. More importantly, the use of duty-cycle operation is essential for event-based applications to prolong the network lifetime.

Supporting service-differentiated real-time communication in WSNs is a cross-layer task. First, an efficient prioritized medium access control (MAC) mechanism is required to provide service differentiation so that the packets with tighter deadline requirements are given higher priority to access the channel and are delivered to the destination earlier than others. Some existing real-time communication protocols [9] use non-prioritized MAC designs, such as B-MAC, with multiple priority queues to implement intra-node traffic prioritization. In this case, if a sender has multiple outgoing packets in queue, the packet with the tightest deadline requirements is scheduled first for transmission. However, a non-prioritized MAC with a priority queue cannot resolve inter-node traffic prioritization. When a number of senders within the same contention area try to send the packets with different deadline requirements, such MAC schemes cannot prioritize the transmission attempts of different senders, which leads to inter-node priority mismatches. Another type of prioritized MAC design has been proposed in [10,11]; this scheme uses a dynamic inter frame space (IFS) and backoff window (BW) extension-based CSMA/CA MAC scheme to resolve inter-node traffic prioritization. However, IFS/BW extension based MAC schemes may experience severe bandwidth under-utilization in multi-hop WSNs with fine-grained traffic classification. The limitations of these schemes are discussed in detail in Section 2.

In addition to medium access, routing is a major challenge for real-time communication provisioning in WSNs. To support high sustainable throughput with tight end-to-end deadline constraints, both the end-to-end hop count and the per-hop transmission delay must be considered in determining the routing metric so that the overall end-to-end communication delay can be minimized. To this end, most existing real-time routing schemes [10–12] use table-based geographical forwarding techniques. Each sensor node must maintain a routing table with location information of all the neighboring nodes and the average pairwise transmission delay. Based on this table, forwarding decisions are made to obtain the maximum forwarding speed. Table-based geographical forwarding techniques require each sensor node to have global localization capability and maintain the accurate one-hop connectivity information. Under dynamic wireless channel conditions, this translates into the frequent exchange of control packets, even when no event traffic exists on the network. The significant control overhead introduced by table-based geographic routing inevitably deteriorates the WSN lifetime in event-based, real-time applications.

Traffic admission control is another crucial component used to improve the bandwidth utilization and energy efficiency of a real-time communication scheme. By estimating the schedulability of packet transmissions, a proper admission control policy can be applied to the outgoing traffic in a per-hop manner. As a result, a packet transmission that is unlikely to meet the required latency constraints should be rejected at an early stage of the end-to-end transmission. However, most existing WSN real-time communication schemes do not consider admission control or simply drop packets only when the end-to-end transmission deadline is missed.

In this work, a novel service-differentiated real-time communication scheme (SDRCS) is proposed to provide soft, real-time guarantees for event-based traffic in WSNs using a cross-layer design. Compared with existing real-time communication schemes, the main contributions of this work are as follows.

1.1. Cross-layer real-time forwarding

SDRCS uses a dynamic forwarding technique to integrate routing functionality with a CSMA/CA-based prioritized MAC scheme. In this way, a receiver contention process is performed at each hop of packet forwarding based on the proposed real-time forwarding metric. Neighboring nodes with better forwarding distance, lower traffic load, and higher channel quality, i.e., those satisfying the real-time requirements, receive a higher priority to forward packets. No routing tables or neighboring node information need to be maintained or periodically exchanged for end-to-end communication; hence, the control overhead is mitigated. Since the forwarding decision is made on-demand, the SDRCS adapts well to network dynamics. More importantly, the fully distributed and on-demand forwarding design makes SDRCS suitable for duty-cycled WSNs.

1.2. Efficient prioritized MAC design

To provide better service-differentiation capability for diverse end-to-end deadline requirements in WSN applica-

Please cite this article in press as: Y. Xue et al., SDRCS: A service-differentiated real-time communication scheme for event sensing in wireless sensor networks, Comput. Netw. (2011), doi:10.1016/j.comnet.2011.06.018

tions, a novel polling contention period-based prioritized MAC is proposed in SDRCS, as an alternative to traditional IFS/BW extension-based MAC schemes. The proposed MAC design helps decrease the average IFS and BW sizes when the number of traffic priority categories is large, thus improving the overall bandwidth utilization of the end-to-end communication when four or more traffic priority categories are supported in the network.

1.3. Light-weight packet schedulability estimation

SDRCS includes a light-weight packet schedulability estimation mechanism utilizing a received signal strength (RSS)-based sensor node grouping technique and a uniform polling contention period design for traffic within any priority category. Based on packet schedulability estimation, proper admission control and early missed-deadline packet-dropping policies are designed to prevent unschedulable packets from being injected into the network and degrading bandwidth utilization.

The rest of the paper is organized as follows: In Section 2, the existing solutions for real-time communication in WSNs are discussed. An overview of SDRCS is provided in Section 3 and the design details and protocol operations are described. The results of extensive simulation evaluations are presented in Section 4 to evaluate the performance of SDRCS, and compare it with two existing protocols, RAP [11] and MMSpeed [10], using two important metrics: average end-to-end latency and on-time delivery rate. The paper is concluded in Section 5.

2. Related work

In this section, we discuss existing real-time communication schemes for wireless sensor networks. We point out the limitations of the existing solutions and the motivation for our cross-layer SDRCS design.

2.1. MAC layer solutions

A prevalent approach for achieving prioritized MAC in WSNs was recently developed [10,11], based on the IEEE 802.11e or 802.11 EDCA [13] standards. These MAC schemes are designed to dynamically adapt the inter frame space (IFS) and/or back-off window (BW) length according to different priority classes. A larger IFS is used to transmit packets with lower priority levels. When the number of packets with the same priority level increases, a larger BW is used to resolve the collision. These prioritized MAC approaches are generally referred to as *Dynamic IFS/BW Extension*-based approaches.

However, IFS/BW extension-based MAC designs do not scale well when the number of supporting priority levels increases under diverse end-to-end deadline requirements in WSN applications. In this case, the MAC design attempts to prioritize the medium access by increasing the IFS and BW sizes for low-priority traffic. Therefore, the average IFS and BW sizes allocated in end-to-end communication and the probability of priority reversion [14] dramatically increase, resulting in significantly degraded bandwidth utilization. Mainly for this reason, the revised version of 802.11 EDCA [13] limits the supported number of priority levels to four. However, supporting fewer priority levels in the network results in more traffic being classified into the same priority level. Such a situation leads not only to a degraded service differentiation ability but also an increased collision possibility in the medium access contention process. Accordingly, a higher collision possibility introduces a larger average BW size. Thus, the average communication throughput deteriorates. Based on the above observations, providing fine service differentiation capability and limited IFS/BW extension is vital in efficient real-time MAC design for WSN applications.

2.2. Routing layer solutions

The majority of existing real-time communication schemes [9–12] adopt traditional table-based pro-active routing approaches with different real-time routing metrics. In table-based pro-active routing, each sensor node maintains a routing table listing all its neighboring nodes. Based on specific real-time routing metrics, one neighboring node that satisfies the application-specific deadline requirement is selected as the next hop to complete packet forwarding. RAP [11] uses a greedy geographic forwarding metric, in which any outgoing packets are routed to the neighboring node with the shortest distance to the receiver. A major limitation of the RAP design is that greedy geographic forwarding does not consider local network conditions, such as load-balance, congestion level, and channel quality. Therefore, the RAP routing decision leads to unpredictable per-hop transmission delay in dynamic WSN environments, which affect not only communication throughput but also packet traversal speed estimation.

RPAR [9], SPEED [12], and MMSpeed [10] improve the real-time routing metric by considering both the geographic information and the average pairwise transmission delay of neighboring nodes. The pairwise transmission delay is usually affected by the local contention level, congestion level, and channel quality. Using the location and delay information, the sender can evaluate the spped of packet progress achieved by a neighboring node and thus make a forwarding decision to minimize end-to-end latency.

Table-based real-time routing techniques encounter common limitations in WSNs. First, to maintain the accuracy of the information listed in the routing table in dynamic WSNs, a number of control mesages must be exchanged periodically. This introduces significant control overhead, especially for event-based WSN applications. Second, table-based routing techniques are not suitable for duty-cycle design, which is vital for energy conservation in WSNs. In an unsynchronized WSN, the sensor nodes with duty cycle design randomly go into sleep mode to decrease their energy consumption. In this case, table-based routing techniques cannot properly identify the active next-hop candidate.

In contrast to table-based forwarding techniques, receiver-contention-based dynamic forwarding techniques have been proposed in recent studies [15–19]. Here, routing functionality is combined with a CSMA/CA-based

MAC design so that an adaptive receiver contention is performed at each hop. Sensor nodes with better forwarding distances than others, lower traffic loads, higher channel quality or higher residual energy levels receive higher priority to respond to the RTS packet with a CTS packet and thus, become the next hop. No routing tables or neighboring node information must be maintained or periodically exchanged. Since the forwarding decision is made on-demand, these schemes can easily adapt to a distributed duty-cycle design.

The existing dynamic forwarding techniques motivates the SDRCS design by allowing for an efficient cross-layer communication approach. However, since the existing dynamic forwarding approaches do not consider soft real-time provisioning in forwarding decision, new forwarding metrics based on prioritized MAC operations must be designed so that the application-specific deadline requirements can be enforced in end-to-end packet forwarding.

2.3. Other solutions

In addition to the aforementioned link- and network-layer solutions, physical- and transport-layer protocols have recently been developed to address energy conservation and reliable communication in delay-constrained WSN applications. In [20], it is pointed out that event detection probability and detection latency are functions of the duty cycles of the sensor nodes. Based on this observation, a distributed algorithm is proposed to regulate the probability of sensor nodes being active, such that an event that occurs anywhere in the network can be detected by the sink within a maximum detection latency and a minimum detection probability. In [21], it is shown that the end-to-end communication reliability and latency achieved in event-based WSNs can be regulated through transport-layer rate control. By observing the average end-to-end communication delay and the on-time delivery rate at the sink, proper control mechanisms for the event data rate are applied to the sensor nodes within the event area so that application-specific event transport reliability or latency requirements can be met at the sink. Since these designs are independent of the MAC and network layer operations, the proposed SDRCS operations can be complemented with these solutions.

3. SDRCS: a service-differentiated real-time communication scheme

In this section, the details of the SDRCS are given in terms of five components. The protocol operations executed in each component are described. The relationship among different components and the process by which a real-time packet is scheduled and forwarded in the SDRCS design are shown in Fig. 2. The void-avoidance capability of SDRCS is discussed at the end of the section.

3.1. Assumptions

We consider a static WSN with homogeneous sensor nodes, and a single sink (see Fig. 1). The nodes communicate through multihop wireless links, using a single channel and fixed transmission power. The sensor nodes are unsynchronized devices without location awareness. The sensor nodes are capable of measuring the received signal strength for each received packet. The above assumptions reflect the current hardware configurations of wireless sensor nodes [22].

We consider a mission-critical event sensing application [5], where the predefined events are detected by the nearby sensor nodes and the event information should be converge-casted [23] to the sink. According to the urgency of each event, data packets can be assigned different end-to-end deadline requirements. Only the packets delivered to the sink before the deadline are deemed useful. We also assume the networks to be connected, where at least one end-to-end forwarding path exists.

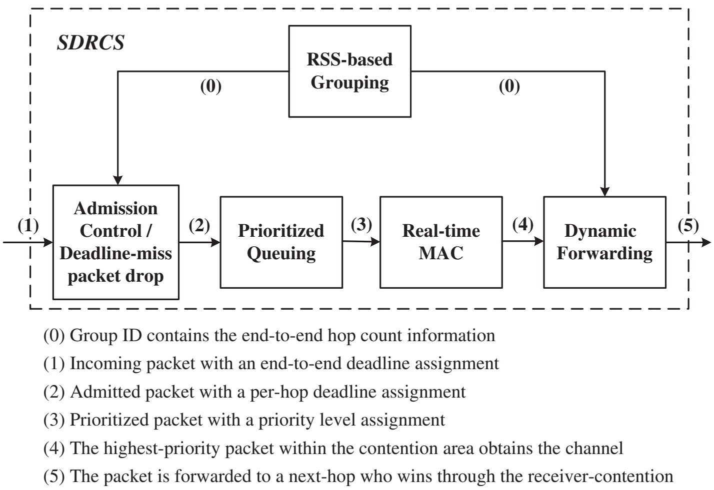


Fig. 2. SDRCS components and the packet forwarding steps.

Please cite this article in press as: Y. Xue et al., SDRCS: A service-differentiated real-time communication scheme for event sensing in wireless sensor networks, Comput. Netw. (2011), doi:10.1016/j.comnet.2011.06.018

3.2. RSS-based sensor node grouping

Many existing real-time communication protocols for WSNs assume precise location awareness at each sensor node [10,12], which requires GPS equipment or complex localization schemes. In the absence of such precise location awareness, we design a Received Signal Strength (RSS) based sensor node grouping method to roughly strip the sensing field into layers, as shown in Fig. 3. The layer information can be used to estimate the hop-distance from the node to the sink, which enables the estimation of packet traversal speed in the packet forwarding process. The accuracy of the hop distances resulting from the grouping can be controlled by grouping granularity, defined as *GRA*. The basic grouping operations are given below:

- Step 1: The sink initializes a *Grouping Message* broadcast with its group ID, where *G_ID* = 0.
- Step 2: Each sensor node, that receives a grouping message with received signal strength *RSS* higher than a pre-defined threshold RSS_{th} and does not have a group ID, is assigned a group ID $G\_ID = G\_ID_r + 1$, where G_ID_r is the group ID value contained in the received grouping message. It then sets its back-off window as $BW = [G\_ID*slot, (G\_ID + 1)*slot]$, and broadcasts a grouping message, which contains its own group ID, once.
- Step 3: Each sensor node, that receives a grouping message with received signal strength *RSS* lower than RSS_{th} and does not have a group ID, is assigned a temporal group ID as $G\_ID_{temp} = G\_ID_r + GRA$. It then sets a timer that expires in GRA * Broadcast period. If a grouping message is received with received signal strength *RSS* higher than RSS_{th} before the timer expires and verifies that $G\_ID_{temp} > G\_ID_r + 1$, a sensor node will assign its group ID as $G\_ID = G\_ID_r + 1$. It then cancels the timer, sets its back-off window as $BW = [G\_ID*slot, (G\_ID + 1)*slot]$, and broadcasts a grouping message, which contains its own group ID, once.
- Step 4: Each sensor node that has a G_ID_{temp} is assigned a group ID as $G\_ID = G\_ID_{temp}$ when the timer expires. It then sets its back-off window as $BW = [G\_ID*slot, (G\_ID + 1)*slot]$, and broadcasts a grouping message, which contains its own group ID, once.
- End: Each sensor node can broadcast the grouping message at most once based on Steps 2–4. Hence, the grouping process ends when all the nodes finish their broadcast.

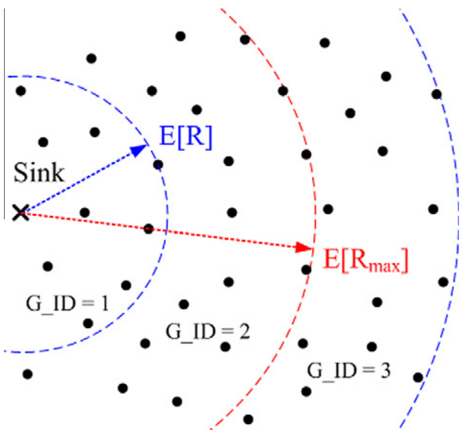


Fig. 3. Received signal strength (RSS) based sensor node grouping example with *GRA* = 2.

The performance of the proposed grouping process can be analyzed using log-normal shadow fading channel model [24]. First, the RSS value obtained by a receiver at a distance *R* from the transmitter is given by

$$RSS(R) = P_t - P_L(R_0) - 10\eta\log_{10}\left(\frac{R}{R_0}\right) + X_\sigma, \tag{1}$$

where P_t is the transmit power in dBm, $P_L(R_0)$ is the path loss at a reference distance R_0 in dBm, η is the path loss exponent, and X_σ is the shadow fading component, where $X_\sigma \sim N(0,\sigma)$. With $RSS(R) = RSS_{th}$, the expected transmission range $E[R]$ of a broadcast message is given by

$$E[R] = R_0 \cdot 10^{\frac{P_t - PL(R_0) - RSS_{th}}{10\eta}} \cdot E\left[10^{\frac{X_\sigma}{10\eta}}\right] \tag{2}$$

$$= R_0 \cdot 10^{\frac{P_t - PL(R_0) - RSS_{th}}{10\eta}} \cdot e^{\frac{\sigma}{10\eta} \cdot ln10}. \tag{3}$$

We also define the maximum transmission range $E[R_{max}]$ based on the noise power floor P_n, where

$$E[R_{max}] = R_0 \cdot 10^{\frac{P_t - PL(R_0) - P_n}{10\eta}} \cdot e^{\frac{\sigma}{10\eta} \cdot ln10}. \tag{4}$$

The grouping granularity, *GRA*, is then defined as

$$GRA = \frac{E[R_{max}]}{E[R]} = 10^{\frac{RSS_{th} - P_n}{10\eta}}. \tag{5}$$

By properly increasing RSS_{th}, the grouping granularity is increased as more layers are assigned to the network. This results in finer end-to-end hop-distance awareness at the sensor nodes. The design of the back-off window, *BW*, ensures that the sensor nodes with higher group IDs cannot interrupt the *Grouping Message* broadcast from a lower-group node; hence, the grouping process can propagate from the sink in a layered manner.

The node grouping should be done at the post-deployment stage. After the RSS-based grouping process, the sensor nodes can be mapped into strip-style groups, using $E[R]/GRA$ as the average width of the strip. The density of the WSN affects the grouping structure. With increasing node density, the result of this grouping would approach perfectly circular strips if the channel fading and noise components in the network are homogeneous [25]. The group ID can be used to estimate the hop distance from the node to the sink and the packet forwarding can be guided towards the sink without precise location information.

The main difference between using an RSS-based grouping technique and a traditional geographic localization technique in the design of real-time communication schemes lies in the definition of the node-to-sink distance. In a traditional geographic forwarding approach, the end-to-end distance is defined as the Euclidean distance, whereas in SDRCS, the distance is defined as the end-to-end hop count. As stated earlier, in a densely deployed WSN with homogeneous channel fading and noise components, the group ID can serve as a good indicator of node-to-sink geographic distance. However, in a sparsely deployed WSN or a WSN with dynamic channel fading and noise components, the group ID is more of an end-to-end hop count estimation, which may not be linearly related to geographic distance. Therefore, the proposed RSS-based grouping technique can help improve the accuracy of end-to-end hop-distance estimation in real WSN deployment, while avoiding the use of the expensive precise localization schemes or devices as in [10,11].

3.3. Per-hop deadline based prioritized queueing policy

In WSNs, an application-specific real-time requirement is usually presented as an end-to-end deadline, which indicates the maximum packet traversal time from the sender to the receiver [26]. However, in a multi-hop network, the end-to-end deadline is not the only criterion for determining the urgency of packet delivery. The end-to-end hop count also affects the packet delivery schedule. For example, if there are two schedulable packets with the same end-to-end deadline requirements competing for the channel, the one with a higher end-to-end hop count should be scheduled first. If we assume that each sensor node is able to predict the end-to-end hop-count to the sink, the end-to-end deadline requirement can be broken down into a per-hop deadline requirement, L_{hop}^{Req}, where

$$L_{hop}^{Req} = \frac{L_{e2e}}{HC_{e2e}}. \quad (6)$$

L_{e2e} is an application-specific parameter which reflects the required end-to-end delay for packet delivery. HC_{e2e} is the predicted hop-count value based on the *G_ID* of the sender, the *GRA* value and the forwarding strategy, which is discussed in Section 3.5. L_{hop}^{Req} reflects the per-hop traversal speed required to achieve the end-to-end real-time guarantees in a contention-based WSN. It can be used as an accurate indicator for packet delivery priority classification [26].

We use FIFO priority queues for packet scheduling at a node, as shown in Fig. 4. Since the prioritized MAC can only provide differentiated service for a limited number of priority classes, the per-hop deadline requirements are further mapped into N priority levels, where N is the number of the priority queues allocated at each sensor node.

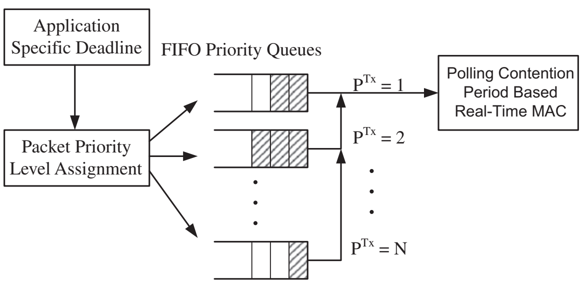


Fig. 4. Per-hop deadline based priority queues at each sensor node for intra-node real-time traffic classification.

In this paper, we use the following packet priority level assignment policy:

$$P^{Tx} = min\left(\left\lfloor \frac{L_{hop}^{Req}}{L_{hop}^{Min}} \right\rfloor, N\right), \quad (7)$$

where P^{Tx} is the assigned packet priority level in the queue and L_{hop}^{Min} is the minimum time required for one-hop packet forwarding, which depends on the MAC operations adopted in the dynamic forwarding design. The L_{hop}^{Min} value for SDRCS is given in Section 3.5. Since Early Deadline First (EDF) has been proven as the most efficient scheduling policy for channel access in wireless networks [27], the packet in a higher priority queue is scheduled earlier for transmission.

Note that the above priority level assignment policy works well when the application-specific L_{hop}^{Req} is uniformly distributed within its design space $\left[L_{hop}^{Min}, N * L_{hop}^{Min}\right]$. For different real-time applications with different L_{hop}^{Req} design spaces and distributions, different priority level assignment policies can be used such that the incoming packets with various L_{hop}^{Req} values can be classified properly into N priority classes and placed into an associated priority queue for transmission [10].

3.4. Polling contention period based real-time MAC

To better support the diverse end-to-end deadline requirements in WSN applications, we designed a polling contention period based real-time MAC to support prioritized channel access.

As mentioned in Section 2, Dynamic IFS/BW Extension is used by most existing real-time communication schemes for prioritized MAC support in WSNs. Such approaches employ extended arbitrary inter frame space (AIFS) and back-off window (BW) size adaptation for prioritized medium access contention. For a packet with priority level i, according to the IEEE 802.11 EDCA [13], the AIFS and BW values are derived as follows:

$$AIFS_i = SIFS + i * SLOT\_TIME, \quad (8)$$
$$BW_i = (BW_1 + 1) * i - 1, \quad (9)$$

where SIFS is the short inter frame space for controlling packet transmission contention. In the dynamic IFS/BW extension based MAC design, a higher number of priority levels supported in the network results in higher average AIFS and back-off window values and a lower average throughput.

In SDRCS, a fixed number of polling slots are used for prioritized packet transmission contention instead of variable inter frame space and back-off window sizes. This design was inspired by the bus access control mechanisms used in computer systems. The basic MAC operation adopted by SDRCS is shown in Fig. 5.

For any packet transmission, a sender first senses the medium. If the medium is idle, the sender waits for the

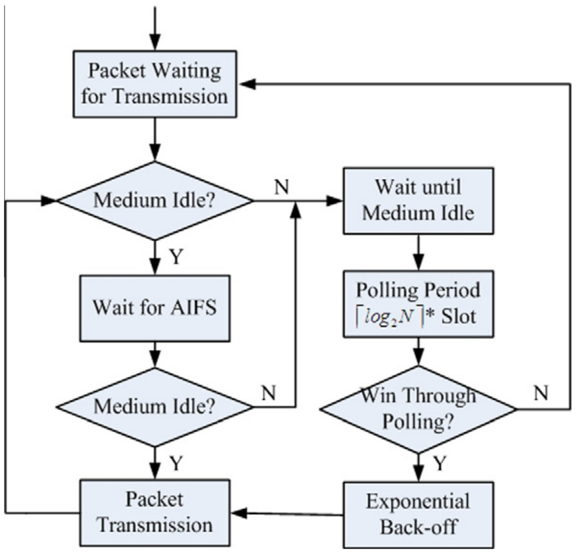


Fig. 5. Polling period-based transmission contention in SDRCS Real-time MAC.

Table 1
Polling-slot design for maximum priority level = 7.

Priority level	Slot 1	Slot 2	Slot 3
1	Active	Active	Active
2	Active	Active	Inactive
3	Active	Inactive	Active
4	Active	Inactive	Inactive
5	Inactive	Active	Active
6	Inactive	Active	Inactive
7	Inactive	Inactive	Active

AIFS period of time and senses the medium again. If the medium remains idle, the sender attempts to initiate packet transmission by sending out an RTS packet. Since several nodes within the interference range may have been waiting for this chance to transmit, all these nodes enter the polling period to compete for the transmission of an RTS packet based on the priority level associated with the outgoing packet. The entire polling period consists of $\lceil \log_2 N \rceil$ polling slots for contention entities with N priority levels. For example, if 7 priority levels are supported in SDRCS, 3 polling slots are required for medium access contention among all possible competitors within the interference area. According to Table 1, any sensor node with an outgoing packet at priority level i will transmit a burst signal in its active polling slots and remain silent in its inactive polling slots. Any node that senses a burst in its inactive polling slots will be suppressed in the following transmission period. In this manner, only the nodes with the highest priority level among all competitors can survive the polling period. Note that more than one node may survive the polling contention period because they have the outgoing packets with the same priority or they are located in a hidden-terminal scenario. Therefore, an extra back-off period is used after the polling period to handle the possible collision (*Exponential Back-off* state at the bottom-left in Fig. 5). Since much fewer number of competing nodes can enter the back-off period after the polling contention period, the BW can be set to a much smaller size compared to that used in dynamic IFS/BW extension based prioritized MAC design. Assuming each priority level has the same amount of traffic load, the proposed polling-contention mechanism provides better overall throughput when the number of priority levels satisfies $N \geqslant 4$.

3.5. Receiver contention based dynamic forwarding

Motivated by the cross-layer forwarding design discussed in Section 2, we developed a receiver contention-based dynamic forwarding for converge-cast packet routing. This approach is combined with the RTS/CTS exchange period of the proposed real-time MAC design. In the real-time MAC design in Section 3.4, if a sender i wins in a polling contention period and gains access to the medium after the exponential back-off period, it will initiate an RTS broadcast containing its own group ID, G_i. All the neighboring nodes that overhear this RTS message enter the receiver contention period, in which only sensor nodes with group IDs equal to or lower than G_i, become the qualified next-hop candidates. Therefore, the packet can only be forwarded towards the sink and gain a non-negative packet traversal speed. The unqualified nodes enter the NAV (Network Allocation Vector) period. Each qualified next-hop candidate is required to evaluate its capability for maximizing the packet traversal speed for this transmission. This capability is classified into M priority levels for receiver contention.

Based on the aforementioned forwarding process and the grouping mechanism described in Section 3.2, the grouping granularity *GRA* gives the maximum number of groups that a packet can traverse within one hop. L_{hop}^{Min} gives the minimum time for one-hop packet transmission. Therefore, the maximum packet traversal speed achievable by a forwarding decision without queuing delay is given by

$$Speed_{max} = \frac{GRA}{L_{hop}^{Min}}. \tag{10}$$

Regarding a specific next-hop candidate j, the average traversal speed $Speed_j$ by forwarding the packets to j is derived from its average pairwise packet transmission time $t_{i,j}^{Avg}$, the queue length L_Q, the average per-packet queuing delay t_Q^{Avg}, and its group ID G_j, where

$$Speed_j = \frac{G_i - G_j}{t_{i,j}^{Avg} + L_Q * t_Q^{Avg}}. \tag{11}$$

In (11), the packet progresses by a forwarding decision based on the group ID difference between the sender and receiver, i.e., $G_i - G_j$. The per-hop packet delivery delay consists of two parts, the packet transmission delay and packet queuing delay. The packet transmission delay, $t_{i,j}^{Avg}$, is calculated using a Exponentially Weighted Moving Average (EWMA) algorithm [28]:

$$t_{i,j}^{Avg} = \alpha t_{i,j} + (1 - \alpha) t_{i,j}^{Avg}, \tag{12}$$

where α is the moving average coefficient, $0 < \alpha < 1$, which represents the degree of weighting decrease. A higher α discounts the older delay value faster, and makes the recent delay value more important in determining the average delay. With low traffic rates and highly unreliable time-varying channel conditions, a larger moving average coefficient is generally used. In (12), $t_{i,j}$ is the instantaneous packet transmission time, which is measured as the time between an RTS transmission and the corresponding ACK is receipt. If the packet is dropped because it exceeds the maximum retransmission time, N_{Re_Trans}, then we have,

$$t_{i,j} = L_{hop}^{Min} * N_{Re\_Trans}.$$

In (12), $t_{i,j}^{Avg}$ is a good indicator of the link quality (packet error rate) of a potential receiver. A higher $t_{i,j}^{Avg}$ than L_{hop}^{Min} indicates the possible retransmission for a packet delivery between node i and j.

The average per-packet queuing delay, t_Q^{Avg}, reflects the local contention level for a particular next-hop candidate and is again calculated using an EWMA:

$$t_Q^{Avg} = \beta t_Q + (1-\beta) t_Q^{Avg},$$

where β is the moving average coefficient, $0 < \beta < 1$, and t_Q is the instantaneous per-packet queuing delay, which is measured as the time between two consecutive packets dequeued from a priority queue. A larger queuing length and per-packet queuing delay indicate a lower packet traversal speed for a given forwarding decision.

Based on (10) and (11), the contention priority for a next-hop candidate j is given as

$$P_j^{Rx} = min\left(M - \left\lfloor \frac{Speed_j * M}{Speed_{max}} \right\rfloor, M\right). \quad (13)$$

The above receiver priority assignment guarantees the following:

- The next-hop candidate with the lowest group ID receives the highest priority for transmitting its CTS packet.
- For multiple next-hop candidates with the same group ID, the one with a better channel quality and lower traffic load gets a higher priority for transmitting its CTS packet.
- The sensor node with the maximum the packet traversal speed is assigned the highest priority.
- A packet is forwarded only to achieve non-negative traversal speed.

Upon receiving the RTS broadcast and evaluating its forwarding priority based on (13), each receiver candidate competes to reply with a CTS packet based on its determined forwarding priority. The same prioritized MAC mechanism is used for receiver contention, as described in Section 3.4. The nodes with the highest forwarding priority among all candidates capture the channel through the $\lceil \log_2 M \rceil$ polling period. After an extra back-off period BW_{CTS}, the winning receiver notifies the sender by a CTS packet with its node ID. Accordingly, the sender will unicast the data packet to the winner, wait for an acknowledgment and finish the one-hop packet forwarding.

A complete prioritized packet transmission contention and receiver contention period for real-time MAC is shown in Fig. 6. According to our receiver-contention based dynamic forwarding operation, two important parameters are determined for (6) and (10). The minimum per-hop latency for the packet transmission with any priority level assignment, L_{hop}^{min}, is given as

$$L_{hop}^{min} = AIFS + t_{Polling}^{RTS} + \frac{1}{2} BW_{RTS,min} + t_{RTS} + SIFS + t_{Data} + SIFS + t_{Polling}^{CTS} + \frac{1}{2} BW_{CTS,min} + t_{CTS} + SIFS + t_{ACK}, \quad (14)$$

where $AIFS$ and $SIFS$ are arbitrary and short IFSs, $t_{Polling}^{RTS}$ and $t_{Polling}^{CTS}$ are the fixed times of the polling periods for RTS and CTS packets, $BW_{RTS,min}$ and $BW_{CTS,min}$ are the minimum back-off window values, and t_{RTS}, t_{CTS},t_{Data} and t_{ACK} are the RTS, CTS, Data and ACK packet transmission times, respectively. The end-to-end hop count estimation for sender i is given by:

$$HC_{e2e} = \frac{G_i}{Avg(G_{Fw}^i)}, \quad (15)$$

where $Avg(G_{Fw}^i)$ is the moving average of the number of groups a packet can traverse within a single hop transmission from node i and $0 \leqslant Avg(G_{Fw}^i) \leqslant GRA$. The receiver's group ID is obtained by i for each transmission by piggybacking on the CTS packets. The initial value of $Avg(G_{Fw}^i)$ is set to GRA. The $Avg(G_{Fw}^i)$ value depends on the local network density, the channel quality and the network congestion level at node i.

3.6. Admission control and early deadline-miss packet drop

Admission control is important in real-time provisioning for WSNs. A well-designed admission control policy can prevent unschedulable traffic from entering the network, thus improving bandwidth utilization and energy efficiency. According to the receiver-contention based dynamic forwarding design, as explained in Section 3.5, the minimum end-to-end deadline requirement for a packet initiated at sender i can be derived from (10) as

$$L_{e2e}^{min} = \frac{G_i * L_{hop}^{min}}{GRA}. \quad (16)$$

Using (15) and (16) in (6) and (7), any schedulable end-to-end requirement can be mapped to a priority level i, where $1 \leqslant i \leqslant N$. Therefore, a simple admission control policy can be adopted at the sender, where packets with priority levels greater than N are not admitted to the network.

At the relay-node, SDRCS employs an early-deadline-miss (EDM) drop policy for relaying nodes. For any relaying

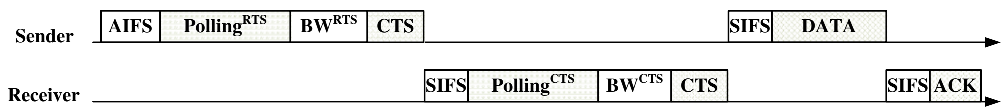


Fig. 6. A complete prioritized packet transmission contention and receiver contention period for real-time MAC.

node k, the cumulative packet transmission time is recorded as t_A and the remaining deadline for a packet, L_r, is calculated as:

$$L_r = L_{e2e} - t_A.$$

From (6) and (15), the updated per-hop deadline is calculated at each relay node as

$$L_{hop}^{Req} = \frac{L_r}{G_k / Avg(G_{Fw}^k)}. \quad (17)$$

If the updated L_{hop}^{Req} is mapped into a priority level greater than N, the packet will be dropped because it is unlikely to be delivered to the sink on time based on the end-to-end hop count estimation at node k. In contrast to the packet drop policies adopted by [10] or [12], which depend on periodically updated per-hop delay information stored in the neighbor list, this early drop policy better adapts to the dynamic channel and load conditions encountered in WSNs, thus avoiding false packet drops due to outdated per-hop pairwise delay information.

3.7. Void avoidance

SDRCS relies on a greedy forwarding strategy at every hop to transmit a data packet to a locally optimal next-hop node ensuring positive progress toward the sink. However, this may not always be possible. For example, in a situation where all the neighboring nodes of a sender are associated with higher groupIDs, the sender will fail to locate a qualified next-hop node that has a positive progress toward the sink. This undesirable phenomenon is usually called a communication void [29]. The presence of communication voids is a challenging problem for any greedy forwarding approach. Although a dense deployment of wireless nodes reduces the likelihood of the occurrence of a void in the network, it is still possible for some packets to encounter voids that are caused by the presence of dead nodes or the boundaries of a wireless network. These packets must be discarded only when a single greedy-forwarding strategy is used, even though a topologically valid path to the destination node may still exist. Thus, it is imperative to provide an effective and efficient void-handling approach in SDRCS.

MMSpeed [10] and SPEED [12] use passive participation to deal with communication voids. The idea of passive participation was introduced in [15], and it exploits a self-healing property of the network topology itself. Once a node identifies itself as a void node, it simply discards the data packet and keeps itself from forwarding any subsequent data packets toward the destination. The node may periodically check whether it can locate a neighboring node guaranteeing positive progress to participate in packet forwarding at a later time. This simple strategy has a reverse-propagation effect, which eventually informs other intermediate nodes to explore other possible paths in the network, such that nodes leading to a broken route can be avoided on routing paths. However, passive participation is not always effective. For instance, as shown in Fig. 7, source node S wants to deliver a sequence of data packets to sink D. The first data packet is greedily forwarded to node V at the first hop. However, node V cannot continue to greedily forward the data packet. Node V drops the data packet and will not participate in forwarding subsequent data packets for destination D. It appears to S that node V no longer exists in the topology. However, no other node in its neighborhood capable of positive progress can help forward the subsequent data packets. Thus, data packets must be discarded, even hough a topologically valid path does exist from S to D: $S-V-A-B-C-E-D$. It was argued in [30] that passive participation is not effective in a randomly deployed wireless network with low density.

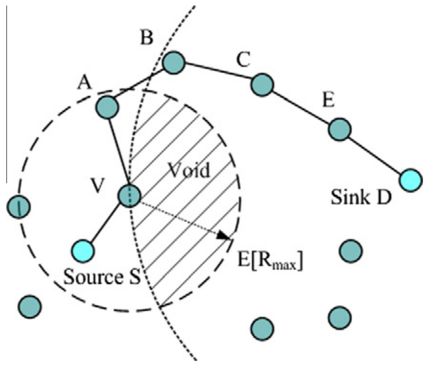


Fig. 7. An example topology with communication voids.

In SDRCS, a communication void is handled inherently by grouping ID assignments and the design of forwarding metrics. First, note that RSS-based grouping is a limited broadcast process initiated at the sink, as described in Section 3.2. Any node can be reached by the broadcast grouping message and assigned a group ID while the network is connected. In addition, any node with a group ID assignment must be able to reach the sink through the reversed broadcast path, if symmetric links are assumed between each pair of connected nodes. Then, considering the receiver-contention based dynamic forwarding operation described in Section 3.5, packet forwarding fails to find a next hop only if one of the following conditions is true:

- There is no node within the transmission range of the sender.
- Any node within the transmission range has a higher group ID than the sender.

However, RSS-based grouping design prevents either condition from occurring in SDRCS operation. First, since the network is assumed to be connected, there must be at least one node within the transmission range of the sender. Second, any sender must have at least one neighboring node with a lower group ID, from which the grouping message is received. As a result, the SDRCS operation guarantees that a packet can always be forwarded from the sender to a node with lower group ID and finally reach the sink, whose group ID equals 0. In Section 4.4, we discuss how grouping results adapt to the communication voids.

Table 2
Simulation parameters.

Sensing field dimensions	(500 × 500) m
Sink location	(25,25)
Number of sensor nodes	100
Node placement	Random uniform
Packet length	128 bytes
Radio bandwidth	250 kbps
Channel model	log-normal shadow fading
Path loss exponent	4
Shadow fading variance	6
Transmission power	1 dBm
Noise power floor	−95 dBm
Maximum transmission range	125 m
Reference distance	0.3 m
Moving average coefficient	0.5

4. Performance evaluation

The performance of our real-time communication scheme, SDRCS, was analyzed in GlomoSim [31] using the simulation parameters shown in Table 2. We chose a log-normal shadow fading channel model [24] to reflect the channel dynamics in real WSN deployments and implemented the model in the simulator. The node related parameters were also carefully chosen to reflect typical MicaZ node capabilities [22]. We explored extensive simulation scenarios for SDRCS and compared its performance with the existing service differentiated real-time communication schemes, RAP [11] and MMSpeed [10]. The MAC operation parameters for dynamic IFS/BW extension based prioritized MAC design (used by RAP and MMSpeed) and real-time MAC (used by SDRCS) are listed in Table 3. For RAP and MMSpeed, AIFS[i], the $BW_{RTS}^{Min}[i]$ and $BW_{RTS}^{Max}[i]$ values are defined based on the simulation settings in [10] and (8), where i is the data packet transmission priority level. For SDRCS, the L_{hop}^{min} value was derived according to (14).

Two important end-to-end metrics were measured for real-time performance evaluation in our simulations:

- *End-to-End On-Time Packet Delivery Rate*: The ratio of the number of unique packets received at the sink with end-to-end latency less than or equal to the end-to-end deadline requirement, to the total number of packets sent by the source node. This metric reveals the end-to-end real-time capacity of the network achieved by a given communication scheme.
- *Average End-to-End Packet Transmission Latency*: Average end-to-end transmission time for all on-time delivered packets. The packets that are dropped enroute due to missed deadlines are not included in the average end-to-end latency calculation. This metric gauges the service-differentiation capability of a given communication scheme.

Table 3
Dynamic IFS/BW extension based prioritized MAC and real-time MAC parameters.

	MMSpeed & RAP	SDRCS
Retransmission Limit	7	7
Number of Priority Classes	7	7
SIFS	10 µs	10 µs
Time Slot	20 µs	20 µs
AIFS[1]	30 µs	80 µs
$BW_{RTS}^{Min}[1]$	15 Slots	10 Slots
$BW_{RTS}^{Max}[1]$	255 Slots	200 Slots
BW_{CTS}	N/A	4 Slots
$L_{per-hop}^{min}$	N/A	2200 µs
GRA	N/A	2

4.1. RSS-based grouping with varying grouping granularity

In this simulation scenario, we examined the performance of the RSS-based grouping scheme. In Fig. 8(a) and (b), the RSS-based geographic grouping results with $GRA = 1$ and $GRA = 2$ are shown, respectively, on a sample network topology generated for the simulation. It can be observed that in the sample network topology with homogeneous channel fading and noise components, RSS-based geographic grouping can properly divide the sensing field into near circular strips. Since the node distribution, and the channel fading and noise components are homogeneous in the network, the node group ID assignment shows a nearly perfect linear relation with the group-to-sink distance.

Next, we examined how the grouping granularity GRA affects the end-to-end performance of the SDRCS. Two sets of simulations were performed with different source nodes and end-to-end deadline requirements. Using the network topology shown in Fig. 8, we chose the source nodes located in the left-bottom corner to maximize the possible end-to-end hop count. In the first simulation set, one constant-bit-rate (CBR) event data flow CBR1 is generated from a node located at (459,411), with end-to-end deadline requirement $L_{e2e}^{Req} = 30$ ms. In the second set, another event data flow CBR2 is generated from a node located at (402,451), with $L_{e2e}^{Req} = 60$ ms. Accordingly, $Priority_{CBR1}^{Tx} = 2$ and $Priority_{CBR2}^{Tx} = 5$. For each set, 2000 real-time packets were sent from the source node. The simulation was conducted 10 times with different random seeds, and the average value is shown.

In Fig. 9(a), the average on-time delivery rates of $CBR1$ and $CBR2$ are shown for GRA values ranging from 1 to 4. It can be observed that increasing the grouping granularity, GRA, to some extent helps improve the end-to-end real-time performance at both priority levels. In this case, for $GRA = 2$ and node degree of 15, up to a 30% improvement in the end-to-end on-time delivery rate is achieved with both traffic flows. However, increasing GRA without considering the network density leads to groups with uneven node distributions or empty groups. In these cases, the packet traversal speed cannot be estimated properly and the end-to-end real-time performance is degraded. In addition, a larger GRA introduces higher control overhead (due to more temporary group ID updates) and longer grouping times (due to the larger back-off window).

To further investigate the relationship between network density and optimal GRA value, we conducted simulations with varying node degree values (15, 22 and 30) and show the resulting average on-time delivery rate of

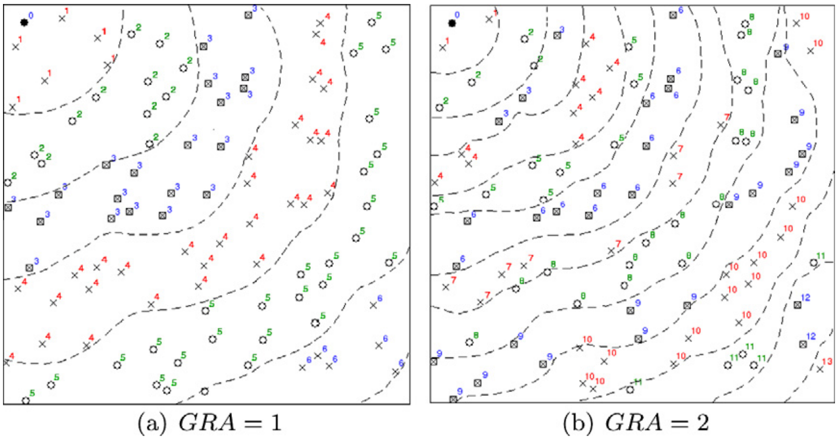


Fig. 8. A sample simulation network topology with a node degree of 15 for grouping granularities of *GRA* = 1 and *GRA* = 2. The numbers shown next to each node are the resulting group IDs.

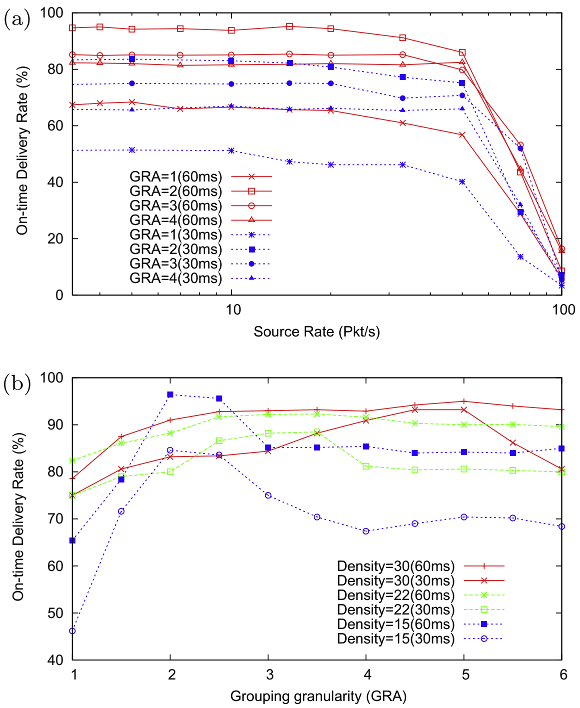


Fig. 9. On time delivery rate of CBR1 and CBR2 traffic.

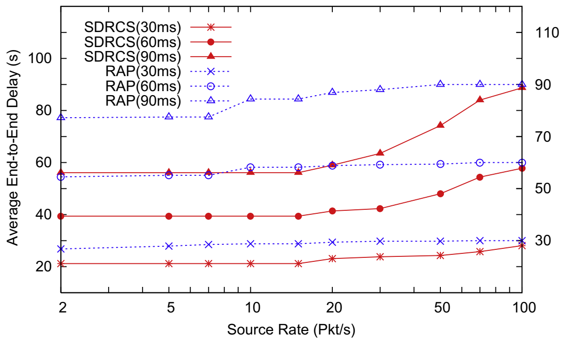
*CBR*1 and *CBR*2 in Fig. 9(b). The simulation results confirm that a larger *GRA* value can achieve better localization information for higher network density. For node degrees of 15, 22 and 30, the best *GRA* values are found to be 2, 3.5, and 5, respectively. The optimal *GRA*s in a WSN may vary with time due to node failures. Such situations require a network regrouping based on a different *GRA*. We will address dynamic GRA adjustment strategies in a future work.

4.2. Performance comparison

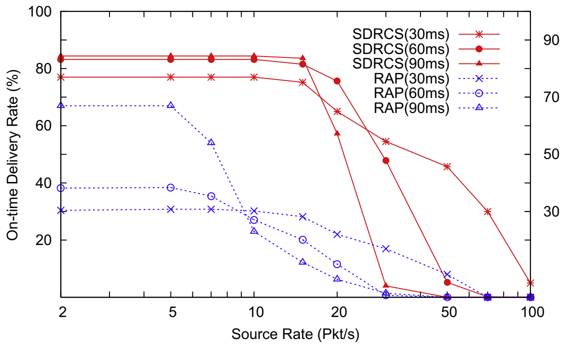
In this simulation scenario, we compare the real-time performance of SDRCS with the existing service-differentiated real-time communication schemes, RAP [11] and MMSpeed [10]. To fully test the protocol performance, we randomly generate 10 network topologies with node degrees of 15. For each network topology, we chose three nodes located in the lower-right corner as the event data sources for maximizing the possible end-to-end hop count.

In this scenario, three constant bit rate (CBR) event data flows CBR1, CBR2, and CBR3, each with different end-to-end deadline requirements are generated simultaneously at the source nodes. The end-to-end deadline requirements of the three flows are 30 ms, 60 ms, 90 ms, respectively. According to the sample packet priority assignment policy given in (7), $P^{Tx}_{CBR1} = 2$ $P^{Tx}_{CBR2} = 4$, and $P^{Tx}_{CBR3} = 5$. Note that, the P^{Tx} values listed here are the initial packet priority level assignment at the sender. The priority level of each packet is updated at each hop according to (7) and (17). For each CBR flow, 2,000 packets are generated and sent to the sink.

The average end-to-end transmission latencies and average on-time delivery rates achieved by SDRCS and RAP are shown in Fig. 10. In Fig. 10(a), it can be observed that both RAP and SDRCS provide service differentiation for traffic flows with different end-to-end deadline requirement in terms of different average end-to-end transmission latencies, because both designs provide prioritized queuing and MAC support. However, RAP always results in a higher end-to-end latency. In our simulated environment setting, a log-normal shadow fading channel model is used to reflect the channel dynamics in real WSN deployments. The channel quality enroute affects the end-to-end delay in terms of per-hop retransmission. The tradeoff between larger per-hop forwarding distance and shorter per-hop latency plays an important role in determining the end-to-end real-time performance. Since RAP assumes a perfect channel model, it simply chooses the next-hop to maximize the per-hop forwarding distance but fails to consider this tradeoff in its forwarding metric design. As a result, the average end-to-end delay is significantly higher for RAP thatn for SDRCS. Moreover, beacuse RAP does not have a dynamic packet traversal speed-estimation strategy in its protocol design, only a baseline missed-deadline packet drop policy is used in

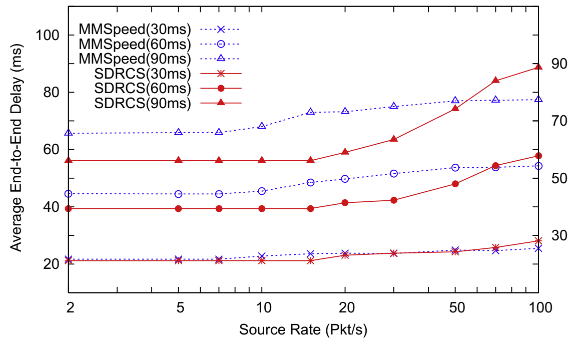


(a) End-to-end latency (SDRCS vs. RAP).

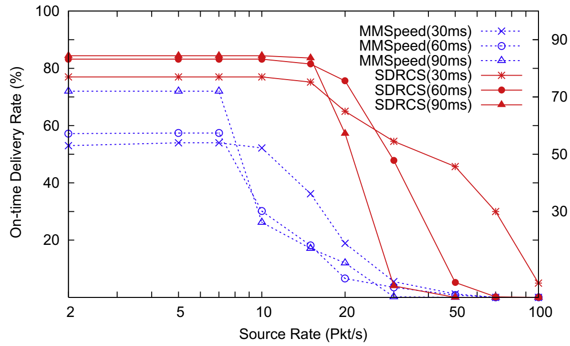


(b) On-time delivery rate (SDRCS vs. RAP).

Fig. 10. End-to-end real-time performance of SDRCS and RAP [11].



(a) End-to-end latency (SDRCS vs. MMSpeed)



(b) On-time delivery rate (SDRCS vs. MMSpeed)

Fig. 11. End-to-end real-time performance of SDRCS and MMSpeed [10].

end-to-end transmission. Therefore, unnecessary packet forwarding cannot be eliminated to improve bandwidth utilization. The long transmission latency and the low bandwidth utilization also affect RAP performance in terms of the on-time delivery rate. Fig. 9(a) showing the on-time delivery rates, reveals that SDRCS provides up to 40% higher average on-time delivery rate for 2,000 packet transmissions and maintains steady on-time delivery rates for much higher source rates without congestion.

In Fig. 11, the comparison of the real-time performance of SDRCS and MMSpeed are shown, in terms of average end-to-end transmission latency and the average on-time delivery rate. In Fig. 11(a), it can be observed that both schemes provide a relatively low average end-to-end latency in achieving soft real-time guarantees. This result indicates that, in contrast to purely geographic forwarding, utilization of both geographic information and channel quality is necessary delay-sensitive communication. In Fig. 11(b), it can be observed that, compared with MMSpeed, SDRCS improves the on-time delivery rate for traffic of any priority level by approximately 20% when the event source rate is less than 5 Pkt/s. When the event source rate reaches 5 Pkt/s, MMSpeed starts to experience network congestion with a significantly decreased on-time delivery rate. In contrast, SDRCS maintains a steady on-time delivery rate for high event source rates up to 20 Pkt/s. The better real-time performance of SDRCS can be attributed to the following factors.

First, *SDRCS provides better overall per-hop transmission latency for traffic with low priority levels.* According to the MAC operations described in Section 3.4, the average per-hop latency for SDRCS is 2200 μ s regardless of the priority levels associated with the traffic. According to (8), this value is approximately the same as the minimum per-hop latency for a dynamic IFS/BW extension-based approach with a priority level equal to 2. However, for MMSpeed, traffic with a lower priority level experiences longer IFS and back-off times in each packet transmission attempt, which results in significantly increased transmission delays. Therefore, the dynamic IFS/BW extension based MAC design adopted by MMSpeed leads to lower transmission throughput and a lower on-time delivery rate, especially for low priority-level traffic. Accordingly, as shown in Fig. 11(b), traffic with looser end-to-end deadline requirements experiences congestion earlier in MMSpeed; whereas in SDRCS, traffic of all priority levels of suffers from congestion at approximately the same source rate. In addition, SDRCS achieves a better on-time delivery rate at the same packet source rate.

4.3. Energy efficiency analysis

In this section, we discuss the energy efficiency of the SDRCS in terms of packet forwarding, packet duplication, and packet dropping. In Fig. 12, the average number of data packet forwarded per unique end-to-end packet delivery (forwardings per delivery; FPD) is shown for SDRCS and MMSpeed under varying end-to-end deadline requirements and source rates. FPD reflects the energy efficiency of the protocol in terms of average end-to-end hop count,

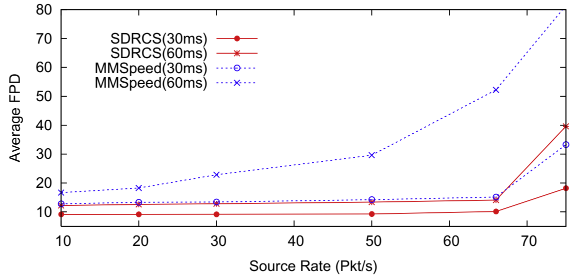


Fig. 12. Energy efficiency performance of SDRCS and MMSpeed, in terms of the average number of packets forwarded for each unique end-to-end on-time packet delivery.

average retransmission time, percentage of dropped packets and duplicated packet rate. When the data rate is low, and the end-to-end deadline requirement is loose, most packets reach the sink on time. Therefore, FPD captures the product of the average end-to-end hop count and the average number of per-hop retransmissions introduced by channel fading, which accounts for the majority of energy consumption in WSNs. The simulation results in Fig. 12 show that SDRCS always results in a lower FPD than MMSpeed under similar deadline and source rate requirements, thus leading to better energy efficiency.

When the data rate is high and the end-to-end deadline requirement is tight, the network contention and congestion levels are increased, resulting in higher packet error rates and longer per-hop transmission delays. Accordingly, the percentage of unschedulable packets or duplicated packet transmissions significantly increase. As a result, large amount of energy is wasted on the delivery of these packets. Under such circumstances, the energy efficiency of the system depends on properly designed admission control and packet-dropping policies.

Compared with MMSpeed, SDRCS adopts an improved policy based on a more accurate and adaptive packet traversal speed estimation method. MMSpeed uses table-based forwarding, where the neighbor list and the average transmission time between a pair of neighboring nodes are periodically exchanged. The packet traversal speed achieved by a certain neighbor is determined by the sender based on this periodically exchanged information. For an end-to-end packet delivery process, if any intermediate node cannot find a neighbor satisfying the required packet traversal speed, the packet is dropped. Such a rigid packet-dropping policy requires an accurate end-to-end hop-count estimation and frequent neighbor information exchange. Under dynamic network conditions, this drop policy may experience a large percentage of false missed-deadline packet drops.

In contrast, with SDRCS, both the packet traversal speed and the required per-hop deadline are estimated based on the information instantaneously updated at the receiver, which helps improve the accuracy of the schedulability estimation. In Fig. 11(a), it can be observed that the average end-to-end latency for SDRCS increases with increasing source rate until it approaches the end-to-end deadline requirement. This indicates that the EDM drop policy correctly estimates most packet drops. Therefore, the average end-to-end transmission latency for each on-time packet delivery increases with an increasing network queuing delay. However, for MMSpeed, the average end-to-end latency does not increase proportionally with increasing source rate, even when the network is fully congested. This situation indicates that MMSpeed's tight drop policy results in a large portion of schedulable packets being dropped because the end-to-end transmission latency is close to the deadline requirement.

The energy efficiency of a communication protocol is also affected by the number of duplicate packets generated. To analyze this metric, two kinds of traffic flows were set up: a single flow sent from one lower-right node with an end-to-end deadline of 30 ms and a mixed flow sent from two lower-right nodes with end-to-end deadlines of 30 ms and 60 ms. The duplicated packet transmission rates of the SDRCS and MMSpeed are shown in Fig. 13, where SDRCS reduces duplicate packet transmissions in the network by more than 70% in comparison with MMSpeed. This difference is mainly due to MMSpeed's probabilistic per-hop multicast mechanism, where the number of multicast receivers in each hop is determined by an end-to-end link error rate estimation. However, due to channel dynamics, such a link error rate estimation is not accurate. Improper per-hop multicasting results in duplicated packets being delivered to the sink and greatly reduces the bandwidth utilization. In contrast, SDRCS results in a much better bandwidth utilization and a higher end-to-end on-time delivery rate.

Based on above analysis, we conclude that SDRCS yields higher energy efficiency for both low-data-rate, loose-deadline and high-data-rate, tight-deadline conditions.

4.4. Void avoidance performance

In this section, we discuss the void-avoidance performance of SDRCS and compare its end-to-end real-time performance with MMSpeed. A network topology was manually generated by removing 30 nodes from the network to create two communication voids located in the lower-left and top-right corner as shown in Fig. 14 with the resulting RSS-based group formation. As described in Section 3.7, each node can obtain a group ID assignment in the sample network topology because the network remains connected. The group ID assignment adapts to the

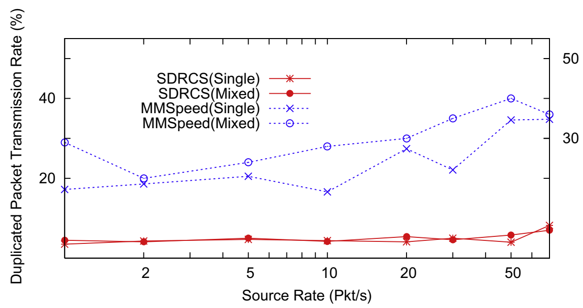


Fig. 13. Percentage of duplicated packets received at the sink for all on-time delivered packets with single or mixed priority CBR traffic flows (SDRCS vs. MMSpeed).

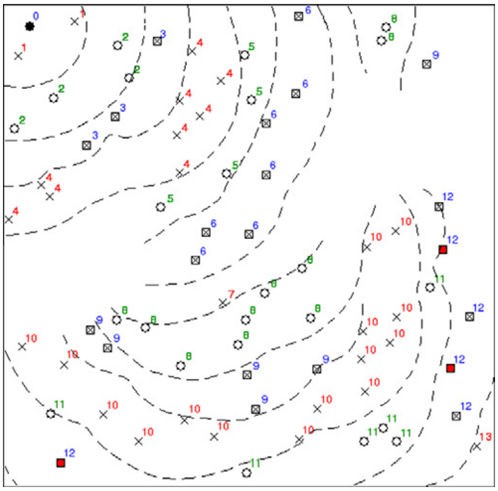


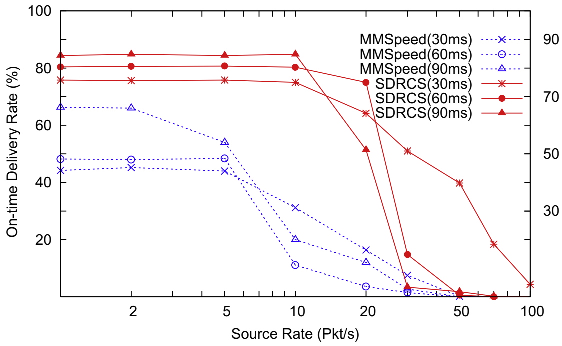
Fig. 14. A sample network topology with two communication voids and the RSS-based group formation results with *GRA* = 2. The source nodes are marked with solid squares.

network void and reflects the end-to-end hop-count estimation instead of a distance estimation.

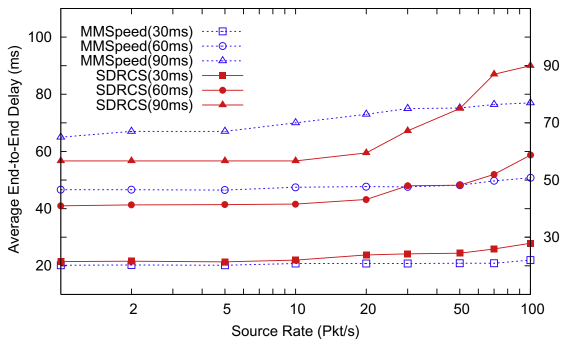
We repeated the simulation scenario described in Section 4.2 using three source nodes marked with solid squares in Fig. 14. The average end-to-end transmission latency and on-time delivery rate are observed for both SDRCS and MMSpeed. The simulation results are shown in Fig. 15. Compared with the real-time performance based on a void-free topology, as shown in Fig. 11, SDRCS did not suffer from significant service degradation, whereas MMSpeed suffered severe degradation.

These results are attributed to two majors factors. First, SDRCS guarantees a route from the sender to the sink, irrespective of the route taken by the dynamic forwarding process. Therefore, a communication void can only slightly increase the end-to-end hop count according to the grouping results. Second, the network capacity bottleneck is located around the sink, where all the packets converge. Therefore, even though fewer parallel forwarding paths can be taken from the sender to the sink, the end-to-end throughput cannot be affected dramatically.

For MMSpeed, the on-time delivery rates dropped by 10% for traffic flows at all priority levels. Since MMSpeed relies on negative participation for void avoidance, some of the forwarding decisions may lead to broken routes. As a result, relaying nodes drop packets and stop participating in further transmissions. Since it takes time for the negative participation to propagate in the reverse direction to the upstream nodes and trigger new route exploration, a number of packets are dropped due to communication voids. This situation becomes severe when the total number of packet transmissions is small.



(a) Average on-time delivery rate.



(b) Average end-to-end transmission latency

Fig. 15. End-to-end real-time performance of SDRCS and MMSpeed based on the sample network topology with communication voids.

5. Conclusions

In this paper, we developed a service-differentiated real-time communication scheme (SDRCS) for multi-hop communication in WSNs. Using RSS-based grouping, SDRCS enables end-to-end hop-distance awareness for sensor nodes with a low control overhead. The hop-distance estimation accuracy can be controlled by adjusting the grouping granularity parameter to meet various application requirements. Along with this grouping approach, a channel-aware dynamic forwarding approach is utilized, with a polling contention-period based prioritized MAC for inter-node traffic differentiation. Compared with the commonly adopted dynamic IFS/BW extension based MAC approaches, the developed MAC design features better service differentiation capability with better bandwidth utilization when the number of priority levels in the network is greater than 4. By including a receiver contention process in the MAC operation, the forwarding decision is locally performed to maximize packet traversal speed. Based on this MAC operation, we also designed a per-hop deadline based prioritized queueing policy for intra-node traffic differentiation.

SDRCS requires no extra hardware for localization, transmission power adaptation, or multiple channel transmission support. It also adapts to network dynamics, such as varying channel quality, local congestion and communication voids. Our analysis showed that SDRCS achieves a better on-time delivery ratio and a higher throughput, with better energy efficiency, than existing approaches such as RAP [11] or MMSpeed [10]. The dynamic forwarding technique eliminates the neighbor exchange packets found in recent approaches. This difference is especially beneficial in high density WSNs, where each node has a high degree. Consequently, SDRCS provides a complete design for real-time communication in WSNs.

Please cite this article in press as: Y. Xue et al., SDRCS: A service-differentiated real-time communication scheme for event sensing in wireless sensor networks, Comput. Netw. (2011), doi:10.1016/j.comnet.2011.06.018

References

[1] Y. Xue, B. Ramamurthy, M.C. Vuran, A service-differentiated real-time communication scheme for wireless sensor networks, in: Proceedings of SenseApp '08, Montreal, Canada, 2008.
[2] S. Oh, L. Schenato, P. Chen, S. Sastry, Tracking and coordination of multiple agents using sensor networks: system design, algorithms and experiments, Proceedings of IEEE 95 (1) (2007) 234–254.
[3] J.R. Moyne, D.M. Tilbury, The emergence of industrial control networks for manufacturing control, diagnostics, and safety data, Proceedings of IEEE 95 (1) (2007) 29–47.
[4] S.N. Pakzad, G.L. Fenves, S. Kim, D.E. Culler, Design and implementation of scalable wireless sensor network for structural monitoring, Journal of Infrastructure Engineering 14 (1) (2008) 89–101.
[5] T. He, V.P., T. Yan, L. Luo, L. Gu, G. Zhou, Achieving real-time target tracking using wireless sensor networks, in: Proceedings of the IEEE RTAS'06, San Jose, California, 2006, pp. 37–48.
[6] I.F. Akyildiz, T. Melodia, K.R. Chowdhury, A survey on wireless multimedia sensor networks, Computer Networks 51 (4) (2007) 921–960.
[7] Y. Wang, M.C. Vuran, S. Goddard, Cross-layer analysis of the end-to-end delay distribution in wireless sensor networks, in: Proceedings of IEEE Real-Time Systems Symposium (RTSS '09), Washington, D.C., 2009, pp. 138–147.
[8] Y. Wang, M.C. Vuran, S. Goddard, Analysis of event detection delay in wireless sensor networks, in: Proceedings of IEEE International Conference on Computer Communications (INFOCOM '11), in press.
[9] O. Chipara, Z. He, G. Xing, Q. Chen, X. Wang, C. Lu, J. Stankovic, T. Abdelzaher, Real-time power-aware routing in sensor networks, in: Proceedings of 14th IEEE International Workshop on Quality of Service: IWQoS 2006, New Haven, CT, 2006, pp. 83–92.
[10] E. Felemban, C. Lee, E. Ekici, MMSPEED: multipath multi-speed protocol for QOS guarantee of reliability and timeliness in wireless sensor networks, IEEE Transactions on Mobile Computing 5 (6) (2006) 738–754.
[11] C. Lu, B. Blum, T. Abdelzaher, J. Stankovic, T. He, RAP: A real-time communication architecture for large-scale wireless sensor networks, in: Proceedings of IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS 2003), Washington, DC, 2003.
[12] T. He, J. Stankovic, C. Lu, T. Abdelzaher, A spatiotemporal communication protocol for wireless sensor networks, IEEE Transactions on Parallel and Distributed Systems 16 (2005) 995–1006.
[13] IEEE802.11WG, Draft supplement to IEEE standard 802.11-1999: Medium access control (MAC) enhancements for quality of service (QoS).
[14] I. Aad, C. Castelluccia, Differentiation mechanisms for IEEE 802.11, in: Proceedings of IEEE INFOCOM 2001, vol. 1, Anchorage, AK, 2001, pp. 209–218.
[15] M. Zorzi, R. Rao, Geographic random forwarding (GeRaF) for ad hoc and sensor networks: multihop performance, IEEE Transactions on Mobile Computing 2 (4) (2003) 337–348.
[16] M.C. Vuran, I.F. Akyildiz, XLP: a cross layer protocol for efficient communication in wireless sensor networks, IEEE Transactions on Mobile Computing 9 (11) (2010) 1578–1591.
[17] P. Casari, M. Nati, C. Petrioli, et al., Efficient non-planar routing around dead ends in sparse topologies using random forwarding, in: Proceedings of IEEE ICC'07, Glasgow, UK, 2007.
[18] L. Galluccio, A. Leonardi, G. Morabito, S. Palazzo, A MAC/routing cross-layer approach to geographic forwarding in wireless sensor networks, Ad Hoc Networks 5 (6) (2007) 872–884.
[19] Y. Xue, M.C. Vuran, B. Ramamurthy, Cost-efficiency of Anycast-based forwarding in duty-cycled WSNs with lossy channel, in: IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), Boston, MA, 2010, pp. 1–9.
[20] Y. Zhu, N.L.M., Probabilistic approach to provisioning guaranteed qos for distributed event detection, in: Proceedings of the IEEE INFOCOM'08, Phoenix, AZ, 2008, pp. 592–600.
[21] V.C. Gungor, O.B. Akan, I.F. Akyildiz, A real-time and reliable transport (RT) 2 protocol for wireless sensor and actor networks, IEEE/ACM Transaction Networks 16 (2) (2008) 359–370.
[22] MicaZ, MicaZ wireless module, <http://www.cmt-gmbh.de/MICAz.pdf>.
[23] H. Zhang, A. Arora, Y. Choi, et al., Reliable bursty convergecast in wireless sensor networks, Computer Communication 30 (13) (2007) 2560–2576.
[24] M. Zuniga, B. Krishnamachari, An analysis of unreliability and asymmetry in low-power wireless links, ACM Transaction on Sensor Networks 3 (2) (2007) 7–31.
[25] S. Vural, E. Ekici, Hop-distance based addressing and routing for dense sensor networks without location information, Ad Hoc Networks 5 (4) (2007) 486–503.
[26] T.F. Abdelzaher, S. Prabh, R. Kiran, On real-time capacity limits of multihop wireless sensor networks, in: Proceedings of IEEE RTSS 2004, Lisbon, Portugal, 2004.
[27] Y. Tian, A. Dogan, F. Özgüner, An urgency-based prioritized MAC layer protocol for real-time traffic in ad-hoc wireless networks, in: Proceedings of the 17th International Symposium on Parallel and Distributed Processing, Washington, USA, 2003.
[28] D.G. Kleinbaum, L.L. Kupper, K.E. Muller (Eds.), Applied Regression Analysis and Other Multivariable Methods, PWS Publishing Co., Boston, MA, USA, 1988.
[29] D. Chen, P.K. Varshney, A survey of void handling techniques for geographic routing in wireless networks, IEEE Communications Surveys and Tutorials 9 (2007) 50–67.
[30] D. Chen, J. Deng, P.K. Varshney, On the forwardingarea of contention-based geographic forwarding for ad hoc and sensor networks, in: Proceedings of IEEE SECON 2005, Santa Clara, CA, 2005.
[31] X. Zeng, R. Bagrodia, M. Gerla, GloMoSim: A library for parallel simulation of large-scale wireless networks, in: Proceedings of Workshop on Parallel and Distributed Simulation, Alberta, Canada, 1998, pp. 154–161.

Yuyan Xue received the BS and MS degrees in control engineering from Nankai University, China, in 2002 and 2005, respectively. She received the Ph.D. degree in computer engineering from University of Nebraska, Lincoln, under the supervision of Dr. Byrav Ramamurthy and Dr. Mehmet C. Vuran, in 2010. Her research interests include real-time and secure communication, cross-layer design and analysis in wireless sensor networks.

Byrav Ramamurthy is currently an Associate Professor in the Department of Computer Science and Engineering at the University of Nebraska-Lincoln (UNL). He is the author of the book "Design of Optical WDM Networks – LAN, MAN and WAN Architectures" and a co-author of the book "Secure Group Communications over Data Networks" published by Kluwer Academic Publishers/Springer in 2000 and 2004 respectively. He serves as the Chair of the IEEE Communication Society's Optical Networking Technical Committee (ONTC). He serves as the IEEE INFOCOM 2011 TPC Co-Chair. His research areas include optical and wireless networks, peer-to-peer networks for multimedia streaming, network security and telecommunications. His research work is supported by the US National Science Foundation, US Department of Energy, US Department of Agriculture, AT& T Corporation, Agilent Tech., Ciena, HP and OPNET Inc.

Mehmet C. Vuran received his B.Sc. degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 2002. He received his M.S. and Ph.D. degrees in electrical and computer engineering from Broadband Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, under the supervision of Prof. Ian F. Akyildiz in 2004 and 2007, respectively. Currently, he is an Assistant Professor in the Department of Computer Science and Engineering at the University of Nebraska-Lincoln and director of Cyber-Physical Networking Laboratory. Dr. Vuran received the NSF CAREER award in 2010. He has received numerous academic honors, including the

2010 Maude Hammond Fling Faculty Interdisciplinary Research Fellowship from the UNL Research Council and the 2007 ECE Graduate Research Assistant Excellence Award from Georgia Tech's School of Electrical and Computer Engineering. He is an associate editor of Computer Networks Journal and Journal of Sensors. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and the IEEE Communication Society. His current research interests include cross-layer design and analysis, wireless sensor networks, underground sensor networks, cognitive radio networks, and deep space communication networks.

Please cite this article in press as: Y. Xue et al., SDRCS: A service-differentiated real-time communication scheme for event sensing in wireless sensor networks, Comput. Netw. (2011), doi:10.1016/j.comnet.2011.06.018