Design of a QoS-aware Routing Mechanism for Wireless Multimedia Sensor Networks

Md. Abdul Hamid, Muhammad Mahbub Alam, and Choong Seon Hong
Networking Lab, Department of Computer Engineering, Kyung Hee University, Korea 446-701
Email: {hamid, mahbub}@networking.khu.ac.kr

Abstract—In wireless sensor networks, majority of routing protocols considered energy efficiency as the main objective and assumed data traffic with unconstrained delivery requirements. However, the introduction of image and video sensors demands certain quality of service (QoS) from the routing protocols and underlying networks. Managing real-time data requires both energy efficiency and QoS assurance in order to ensure efficient usage of sensor resources and correctness of the collected information. In this paper, we present a novel QoS-aware routing protocol to support high data rate for wireless multimedia sensor networks. Being multi-channel multi-path the foundation, the routing decision is made according to the dynamic adjustment of the required bandwidth and path-length-based proportional delay differentiation for real-time data. The proposed protocol works in a distributed manner to ensure bandwidth and end-to-end delay requirements of real-time data. At the same time, the throughput of non-real-time data is maximized by adjusting the service rate of real-time and non-real-time data. Results evaluated in simulation demonstrate a significant performance improvement in terms of average delay, average lifetime and network throughput.

I. INTRODUCTION

This paper aims at designing a novel quality of service (QoS) aware packet delivery technique to support high data rate and delay requirements for wireless multimedia sensor networks (WMSNs). The promising pace of technological growth has led to the design of sensors capable of sensing and producing multimedia data. However, as the multimedia data contain images, video, audio and scalar data, each merits a different metric. In order to accommodate high data rate, the design of an efficient routing protocol is of primary interest. The essence of such a protocol emerges due to few challenging and motivating reasons. First, existing data rates of about 40 kbit/s and 250 kbit/s supported by the MICA2 and MICAz sensor motes are not geared to support multimedia traffic [1]. Beside the hardware improvement and associated cost, an alternate approach is to efficiently utilize the available bandwidth by using multiple channels in a spatially overlapped manner to support multimedia applications. Second, the use of multi-path leverages two things: (1) load may be balanced for not to overwhelm the limited buffers at the intermediate sensor nodes, and (2) one path condition may not permit high data rate for the entire duration of the event being monitored. Hence, by allowing multiple paths, the effective data rate at each path gets reduced and the application can be supported. This paper presents a protocol that targets the application of WMSNs where sensors produce multimedia contents form the deployed area to deal with both critical and general data. Applications may include monitoring of a volcano explosion, toxic gases or a forest fire, sniper or enemy detection, detection of the location of survivors for rescue services. Once a node detects an important event, fast and reliable delivery is required since late or failed delivery may cause severe disaster. In real-time applications, such as multi-media streaming, delivered data can become useless by only a few milliseconds.

Though current sensor networks use single channel, a significant number of present sensor node prototypes use radio modules capable of transmitting on multiple channels. For example, radio capabilities of MICAz mote [2] can communicate on multiple channels as specified in the 802.15.4 standard. The idea of using multiple channels in wireless networks is not new. [3] studies how the capacity of a static multi-channel network scales as the number of nodes increases in the network. Authors show that it may be possible to build capacity-optimal multi-channel networks with as few as one interface per node. Authors in [4] present a multi-channel defense mechanism against jamming attacks in wireless sensor networks by automatically assigning different channels to nodes in the jammed area in order to defeat an attacker. [5] introduces a control-theoretic approach to maximize throughput in multi-channel sensor networks by choosing node’s communication frequencies.

Classical multi-path routing has been explored for two reasons: (1) load-balancing in which traffic is split across multiple disjoint paths, and (2) reliable data delivery in which multiple copies of a packet are sent along different paths. While a plethora of techniques have been developed for sensor networks, all protocols featured either the concept of multi-path or multi-channel assumption. QoS provisioned routing protocol with the efficient use of both multi-path and multi-channel for WMSNs has not been addressed. We devise a packet delivery mechanism over multi-path multi-channel provisioned WMSN in which sensors ubiquitously retrieve multimedia contents from the environment. Our major goal is to support high data rate while keeping the attainable delay so that packets can be delivered to the destination with their bandwidth and delay requirements. More specifically, multi-path multi-channel lay the foundation on which routing decisions for real-time and non-real-time traffics are taken using the dynamic bandwidth adjustment and path-length-based proportional delay differentiation (PPDD) techniques. To meet the bandwidth requirements, the proposed technique...
provides network-wide dynamic bandwidth adjustment options for the nodes in a distributed manner. To meet the delay requirement, the proposed technique provides PPDD options extending the idea of proportional delay differentiation (PPD) mechanism in [6].

The rest of the paper is organized as follows. Section II describes an overview of existing works. Section III provides network model and assumptions. Section IV presents the QoS-aware routing protocol in details. Section V presents performance evaluation. Finally, Section VI concludes this paper with summary and directions for future works.

II. RELATED WORKS

In [1] and [7], surveys of the research challenges and the current status of the literature in WMSNs are presented. More specifically, factors influencing multimedia delivery over WSN are pointed out along with their shortcomings and open research issues. Author in [8] gives a short overview of the hot topics in multimedia surveillance systems and introduces research activities currently engaged in the world. Multi-flow Real-time Transport Protocol (MRTP) [9] is suited for real-time streaming of multimedia content by splitting packets over different flows. However, MRTP does not specifically address energy efficiency considerations in WMSNs. In [10], a wakeup scheme is proposed to balance the energy and delay constraints. In [11], the interesting feature of the proposed protocol is to establish multiple paths (optimal and sub-optimal) with different energy metrics and assigned probabilities. Hence, it is inherently a multi-path protocol with QoS measurements and a good fit for routing of multimedia streams in WSN. In [12], a Multi-Path and Multi-SPEED routing protocol is proposed for WSN to provide QoS differentiation in timeliness and reliability. A class-based queuing model is employed in [13] in order to support both best effort and real-time traffic at the same time. The bandwidth ratio, \( r \), is defined as an initial value set by the gateway. This \( r \)-value represents the amount of bandwidth to be dedicated both to the real-time and non-real-time traffic on a particular outgoing link in case of a congestion. As a consequence, the throughput for normal data does not diminish by properly adjusting such \( r \) value. However, the same \( r \)-value is set initially for all nodes to satisfy the least hop node’s delay requirement, which does not provide flexible adjustment of bandwidth sharing for different links. Also the average delay increases with higher real-time data rate. The protocol is extended in [14] by assigning a different \( r \)-value for each node in order to achieve a better utilization of the links. But finding the \( r \) values and sending these to the particular node is not only an overhead but energy consuming as well, since the \( r \) values have to be unicasted to every single node from the gateway. Moreover, whenever a route changes, a set of new \( r \) values for the new routes has to be calculated and transmitted to all the nodes. In our protocol, each node locally adjusts the bandwidth and delay requirement based on the path-length and incoming traffic.

III. NETWORK MODEL AND ASSUMPTIONS

We consider a static wireless network containing homogeneous multimedia sensor nodes capable of performing all possible application tasks (e.g., capable of sensing video, audio, scalar data). The network is based on a flat, homogeneous architecture in which every node has the same physical capabilities and can interact with neighbor sensors [1]. A subset of the deployed sensors has higher processing capabilities termed as multimedia processing hubs [1] as shown in Fig. 1. The processing hubs are responsible for in-network processing (i.e., data aggregation, discard of redundant data) in a distributed fashion [1]. Both the multimedia sensors and processing hubs are equipped with single radio interface and multi-channels and are capable of transmitting or receiving data on one channel at a given time. The task of the sensors is to dynamically serve the need of multimedia data from the target area to the sink.

IV. QoS-AWARE ROUTING PROTOCOL

The proposed QoS-aware routing protocol is of two-fold significance: (1) multi-path multi-channel provisioned network topology construction and (2) QoS-aware packet scheduling technique. We describe each of them in the sequel.

A. Multi-path Multi-channel Network Topology Construction

We exploit the existing techniques to realize the network with multi-path multi-channel provisions based on the multi-path construction mechanism in [15] and multi-channel assignment technique in [4]. The outcome of these techniques is to assign each network nodes with the knowledge of available paths and channels to transmit/receive data packets. Fig. 2 shows one possible multi-path construction termed as localized disjoined multi-paths that use localized information alone and not relying on global topology. As shown in Fig. 2(a), some path request packets have initially been flooded throughout the network by the source nodes. The sink then has some empirical information about which of its neighbors can provide it with the highest quality data (lowest loss or lowest delay). To this most preferred neighbor, it sends out a primary-path (P2) reinforcement as shown in Fig. 2(b). As with the basic directed diffusion scheme, that neighbor then locally determines its most preferred neighbor in the direction of the source, and so on. Accordingly, alternative paths P1 and P2 are constructed. For details, please refer to [15].

Next task is to assign each node its transmission activities to efficiently utilize the bandwidth using multiple frequencies.
Mutually Orthogonal Latin Square (MOLS) based scheduling is applied to assign transmission/reception activities presented in [4] as described below.

**Definition 1.** A $p \times q$ rectangular array formed by the symbols $1, 2, \ldots, k$, where $k \geq p$ and $k \geq q$, is called a latin rectangle if every symbol from the symbol set appears at most once in each column and once in each row.

**Definition 2.** A latin square of order $p$ is a $p \times p$ matrix with entries from a set of $p$ distinct symbols such that each row and column contains every element exactly once. The symbol in the $i^{th}$ row and the $j^{th}$ column is written as $a_{i,j}$.

**Definition 3.** Two distinct $p \times p$ latin squares $A$ and $B$, where $(a_{i,j})$ and $(b_{i,j}) \in \{1, 2, \ldots, p\}$, are said to be orthogonal if the $p^2$ ordered pairs $(a_{i,j}, b_{i,j})$ are all different.

$$S = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 4 & 5 & 6 & 0 \\ 2 & 3 & 4 & 5 & 6 & 0 & 1 \\ 3 & 4 & 5 & 6 & 0 & 1 & 2 \\ 4 & 5 & 6 & 0 & 1 & 2 & 3 \\ 5 & 6 & 0 & 1 & 2 & 3 & 4 \\ 6 & 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix} \quad \quad R = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 0 \\ 2 & 3 & 4 & 5 & 6 & 0 & 1 \\ 3 & 4 & 5 & 6 & 0 & 1 & 2 \\ 4 & 5 & 6 & 0 & 1 & 2 & 3 \\ 5 & 6 & 0 & 1 & 2 & 3 & 4 \\ 6 & 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

The square matrices $S$ and $R$ are examples of latin squares of order 7. In latin square based scheduling, channels correspond to the rows and time slots correspond to the columns [16]. According to definition 3, the two $7 \times 7$ latin squares $S$ and $R$ are orthogonal.

**Lemma 1.** If two nodes are assigned two symbols from two different orthogonal latin squares, then there is at most one collision for these two nodes in every time frame (proof is given in [16]).

$$SR = \begin{bmatrix} (0,1) & (1,2) & (2,3) & (3,4) & (4,5) & (5,0) \\ (1,2) & (2,3) & (3,4) & (4,5) & (5,6) & (6,1) \\ (2,3) & (3,4) & (4,5) & (5,6) & (6,0) & (0,2) \\ (3,4) & (4,5) & (5,6) & (6,0) & (0,1) & (1,3) \\ (4,5) & (5,6) & (6,0) & (0,1) & (1,2) & (2,3) \\ (5,6) & (6,0) & (0,1) & (1,2) & (2,3) & (3,4) \\ (6,0) & (0,1) & (1,2) & (2,3) & (3,4) & (4,5) \end{bmatrix}$$

During the network initialization phase, a distributed distance-2 vertex coloring algorithm [17] is performed. This approach requires only local information from immediate neighbors to assign the vertex color to the network node. The algorithm outputs different vertex colors to all nodes within interference range of each other (the 2-hop distance is a good approximation of the carrier sensing range in ad hoc networks, and node activation scheduling usually requires all neighbors of a node within 2-hops to be silent when the node transmits [4]). Therefore, the problem of assigning square symbols to nodes can be modeled as a distance-2 graph coloring problem such that each node can directly use its assigned vertex-color as its square symbol. In addition to vertex-coloring, MOLS matrices are generated during the network initialization phase. The orthogonality of the squares corresponds to there being exactly one time/channel assignment for every pair of nodes in different squares. In this way, a node can decide to be a sender or a receiver by picking the appropriate square. For example, let the entries in the square labeled $S$ represent the set of sender nodes, and entries in square $R$ represent the set of receiving nodes. Combining the 2 squares together will result in a unique time/channel assignment for each pair of sender/receiver as shown in $SR$ matrix. The uniqueness of each pair assignment is guaranteed by the orthogonality of the 2 squares. As shown in Fig. 3, there are 9 nodes numbered from 1 to 9 and each gets one of 7 different colors numbered from 0 to 6. Any pair of communicating nodes may select appropriate symbols according to their vertex colors to be a sender/receiver pair for a collision free transmission/reception.

With these multi-path multi-channel provisions, nodes in the network may decide to forward packets to the appropriate path/channel depending on the desired QoS requirements (i.e., required bandwidth and delay described later).

### B. QoS-aware Packet Scheduling

1) Multimedia Traffic Class and Queuing Model: Sensor data may originate from various types of events that have different levels of importance as depicted in table I. Hence, packet scheduling policy should consider different priorities (importance) for different type of traffic classes. For example, time-critical packets may be assigned with high priority compared to non-time-critical packets to meet the deadlines. Fig. 4 shows the queuing model for a sensor considering different traffic classes described in table I. On each node, there is a classifier that checks the type of the incoming packets and sends to appropriate queues, and a scheduler that schedules the packets according to the delay and bandwidth requirements.
2) QoS Assurance: To meet the QoS requirements for a packet from source to destination along a path, let us derive a path specific condition. Suppose, a packet $p_i$ on path $P$ originated at source node at time $t_i$ and has to be reached to the destination by $t_i + T_i$, where $T_i$ is the deadline of packet $p_i$. The arrival time of the packet at hop $j$ denotes the time it is inserted into the queue at that node. The departure time of $p_i$ from hop $j$ denotes the time the transmission of $p_i$ is completed. The arrival time of $p_i$ at hop $j + 1$ is equal to its departure time from hop $j$ plus propagation delay. Let the time this packet $p_i$ spends at hop $j$ be $d_j$, which is the interval between its arrival time and departure time at hop $j$. And let $s_j$ denote the switching delay from one channel to another at each hop. So, the packet $p_i$ will reach the destination preserving the delay bound if

$$
\sum_{j=1}^{H} (d_j + \rho_j + s_j) \leq T_i \tag{1}
$$

where, $\rho_j$ is the propagation delay for each hop $j$ and $H$ is the total number of hops a packet travels. $\rho_j$ in (1) can be neglected since propagation occurs at the speed of light. Considering the delay for a specific path, sink (network designer) may determine the required bandwidth consumed by different traffic classes. We denote $B$ as the required bandwidth. Initially, the sink will determine the value of $B$ based on the observed delay for time-critical packets and broadcasts this value. After receiving the value, all nodes will dynamically calculate (locally) their own value of $B$

$$
B_H = \frac{B}{H + \alpha} \tag{2}
$$

where, $H$ is hop count of the node from the sink, $\alpha$ is the adjusting factor and the node will set this value based on the incoming traffic. When the sink observes that the end-to-end delay is increased, it increases the value of $B$ to allocate more bandwidth to real-time traffic and vice versa. Also, the sink tries to make the value of $B$ as minimum as possible without violating the QoS requirement to maximize the bandwidth use of non-real-time traffic.

Path-length-based Proportional Delay Differentiation (PPDD): PPDD is a DiffServ-based service which is an extension of Proportional Delay Differentiation (PDD) [7] defined for wired network. The PPDD scheduler services packets in classes and realizes proportional average per-hop queuing delays among them locally at each node along the path. At node $k$, packets from class $i$ experience smaller delay than class $j$ for all $i > j$, and $i, j \in S_{b,k}$, where $S_{b,k}$ is the set of backlogged classes at node $k$. Usually, end-to-end delay of a packet is proportional to the number of hops. For example, for a path, a packet $H$ hops away from the sink experiences smaller end-to-end delay than a packet that is more than $H$ hops away. As shown in Fig. 5, a packet from node 1 will experience more end-to-end delay than that of a packet from node 4. The spacing between delays is tuned by the sink based on observed delay with a set of class differentiation parameters. As its name suggests, the model not only holds at each node, it also holds across all nodes in a path. The PPDD service model is defined as follows.

![Fig. 4. Queuing model on a multimedia sensor node.](image)

![Fig. 5. Dynamic bandwidth adjustment and PPDD calculation.](image)
Let $1 = \delta_1 > \delta_2 > \cdots > \delta_H > 0$ be delay differentiation parameters which define that a packet smaller hops away from the sink may allow higher delay than a packet that arrives from a node more hops away. Let $d_H^k$ denote the average queuing delay of a packet at node $k$ that is $H$ hops away from the sink. Then, PPDD requirement is given according to (3).

$$\frac{d_H^k}{d_{H+1}^k} = \frac{\delta_H}{\delta_{H+1}} \quad (3)$$

Then, with WTP, each class is serviced with a separate first-in-first-out queue. The head-of-line packet of a class is assigned a WTP based on the service class and waiting time of the packet and the scheduler always schedules the highest priority head-of-line packet for transmission.

3) Routing: The packets are routed through the nodes along the path from the source to the destination in which nodes choose paths/channels that meet the bandwidth and delay requirements. Each node knows available path options and collision-free channel assignment within 2-hops and adjusts bandwidth and delay according to (2) and (3), respectively, to relay traffic along the path. Packets that do not meet the deadline (i.e., QoS requirements) are discarded. Best effort traffics are routed through the alternative paths to balance the distribution of remaining traffic. Redundant data are aggregated by the processing hubs to reduce the network traffic.

V. PERFORMANCE EVALUATION

The effectiveness of the proposed QoS-aware routing approach is judged through simulation in $n$s-$2$ [18]. We consider three important performance metrics: average delay per packet, average lifetime of a node, and network throughput. Parameters those affect these metrics are real-time data generation rate, buffer size, and packet discard probability. We compare our results with single-$r$ and multi-$r$ mechanisms [14]. We assume a multi-hop network where 100 nodes are uniformly distributed and randomly placed in a 500m x 500m square area. The sink is positioned within the region boundary at [500, 250] location. Twenty five (25%) sensors have higher processing capabilities (multimedia processing hubs) and are responsible for in-network processing in a distributed fashion. We use 7 channels and take 250 $\mu$s as the channel switching delay. Other simulation parameters are taken similar to [14].

First, we consider the impact of real-time data rate on the average delay per packet for both real and non-real-time data. Average delay is defined as the average time a packet takes to reach the sink from a source. We observed that both multi-$r$ and single-$r$ mechanisms have higher delay compared to proposed protocol as shown in Fig. 6(a). Multi-$r$ mechanism performs better than single-$r$ mechanism as expected. Because, every node adjusts its $r$-value based on the resource it has. This is more efficient than the single-$r$ case in which a unique $r$-value is imposed by the sink for all the nodes. Intuitively, average delay per packet for the proposed protocol is less than single-$r$ mechanism, since forwarding nodes locally adjust the bandwidth (value of $B$) proportional to the expected load instead of a single-$r$ value for all forwarding nodes. Also, even though the value of $B$ is not exactly set like multi-$r$ mechanism, the average delay is less than multi-$r$ mechanism since it uses unicast message to deliver the individual value to the nodes. The proposed mechanism requires very less control packets since a single $B$ value is sent to all nodes and increases the forwarding rate of the data packets. Fig. 6(b) shows the effect of real-time data rate on average delay per non-real-time packet. The delay increases with the rate since packets incur more queuing delay and share the same amount of bandwidth. Note that the average delay for non-real-time packets in multi-$r$ is more than that of single-$r$ mechanism. In multi-$r$, the increase in the throughput of non-real-time packets causes extra queuing delay on the nodes leading non-real-time packets to experience end-to-end delay [14]. Our protocol has less average delay compared to multi-$r$ protocol since nodes can schedule non-real-time packets exploiting multiple paths.

Next, we consider network throughput which is measured as the total number of packets received at the sink divided by the simulation time. Fig. 7(a) shows that proposed protocol outperforms existing protocols due to efficient utilization of wireless spectrum. When the number of real-time packets increases, it gets more difficult to satisfy increasing number of QoS paths. Hence, this can cause rejection of paths or packet drops for non-real-time packets leading to throughput degradation. Fig. 7(b) shows the average lifetime of the nodes. Our model consumes less energy since it does not require multiple unicast transmission of the $r$ value like multi-$r$ mechanism. Also, the processing hubs reduces redundant data and channel assignment technique results in less collisions.
mechanism provides a significant performance improvement in terms of average delay, average lifetime and network throughput. However, we realize that there are many challenges that need to be resolved. We plan to run the experiment on a test bed to evaluate the performance. To minimize switching delay, we like to focus on the selection of routes that do not require frequent switching. Finally, we intend to delve into the performance issues for both mobile and multiple sinks.

ACKNOWLEDGMENT

This research was supported by the MKE under the ITRC support program supervised by the IITA (IITA-2008-(C1090-0801-0002)). Dr. CS Hong is the corresponding author.

REFERENCES


VI. CONCLUSION

We have designed a QoS-aware routing mechanism to meet the challenges posed by WMSNs and shown that the proposed