
Long Cheng*, Jianwei Niu*, Yu Gu†, Tian He‡, Qingquan Zhang§
* State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China
† IBM Research - Austin, USA
‡ University of Minnesota, Twin Cities, USA
§ University of Maryland, Baltimore County, USA

Abstract—Radio duty cycling is a commonly employed mechanism to support long-term sustainable operations of WSNs. Combined with the effect of unreliable wireless links, many challenges arise for ensuring delay bounded data delivery with reliability constraint. However, research on energy-efficient data forwarding with statistical delay bound in duty-cycled WSNs still remains unaddressed. This paper proposes EDGE, a novel opportunistic forwarding technique tailored for duty-cycled WSNs with unreliable wireless links. The key idea is to exploit the available path diversity to minimize the transmission cost while providing statistical delay guarantees. Delay quantiles are derived at each node in a distributed manner and are used as the guidelines in forwarding decision making, so that an early arriving packet will be opportunistically switched to the energy-optimal path for communication cost minimization. Extensive evaluation results show that EDGE effectively reduces the transmission cost by 12%–25% with statistical delay guarantees under various network settings.

Index Terms—Wireless Sensor Networks; Statistical Delay Guarantee; Duty-Cycled Networks

I. INTRODUCTION

Energy-efficient wireless communication is critical for many sustainable applications such as environment and infrastructure monitoring [1], [2] in wireless sensor networks (WSNs). On one hand, to support long-term operations with limited energy supplies, a sensor network has to be operated at a duty-cycled (even less than 5%) mode, where nodes sample wireless channels very briefly and stay in dormant state most of the time for energy saving [3]–[6]. Consequently, a sender may experience certain sleep latency in each hop, which is defined as the time that a sender spends on waiting for a receiver to wake up. On the other hand, data forwarding in WSNs could be delay-sensitive with certain reliability constraint [7], such as for mission-critical applications [8], or data collection with mobile sink where limited sojourn time is assigned at each rendezvous point and data packets must be delivered to the rendezvous point before the mobile sink passing by [9], [10].

Wireless medium is inherently unreliable, due to many factors such as interference, attenuation, and channel fading, especially for low-cost sensor nodes [11]. The inherent uncertainties and dynamics in wireless link properties pose several challenges for data forwarding with service guarantee (e.g., the delivery reliability guarantee within a specified time frame) in duty-cycled WSNs. Firstly, the unreliable links inevitably incur data retransmissions and the interval between consecutive retransmissions would be large due to the sleep latency, subsequently resulting in large end-to-end (E2E) delivery latency [4]. Secondly, link unreliability makes it challenging to accurately estimate the E2E delay due to the large variance of in-situ delays. Under highly unreliable communication links, routing schemes based on the expected E2E delay (EED) lack data delivery performance guarantee in general due to the instability of the wireless medium [12]. Thirdly, it is difficult to achieve energy-efficiency and delivery latency simultaneously since the expected-energy-optimal routing path normally varies from the expected-delay-optimal path considering the combined effects of radio duty cycling operation and unreliable links.

Statistical delay guarantee is considered as a practical QoS (Quality of Service) requirement by defining the constraint in terms of the delay bound violation probability, since deterministic delay bounds are normally prohibitively expensive to guarantee in WSNs [13], [14]. Despite pioneering works have been proposed for delay-efficient data forwarding in duty-cycled WSNs [4], [15], [16], existing approaches offer no explicit statistical guarantee of delay bounded data delivery. In this work, we focus our investigation on enabling energy-efficient data forwarding with probabilistic delay bound guarantees, which will be of increasing importance for many delay sensitive WSN applications [17]. We propose the energy-efficient statistical delay guarantee (EDGE) protocol that addresses the aforementioned challenges in duty-cycled WSNs with unreliable links. The key novelty of EDGE lies in the derived delay quantiles and dynamic forwarding decision making with statistical delay guarantees, in which a node switches to the energy-optimal path to forward a data packet if the packet arrives opportunistic earlier than a statistical delay threshold. Specifically, the major contributions of this work are listed as follows:

• To our knowledge, this is the first distributed data forwarding protocol that explicitly supports diverse reliability guarantees with delay bounds in duty-cycled WSNs. A novel opportunistic forwarding technique, i.e., EDGE, over duty-cycled networks is proposed to improve the energy efficiency without compromising the probabilistic delay guarantee.
• We propose a recursive and distributed method to com-
compute the delay quantiles at each node during the routing path construction phase, which is then used as the guideline in forwarding decision making for opportunistic energy saving in the process of data forwarding.

- Through comprehensive performance comparisons, we demonstrate the effectiveness of EDGE for energy saving during data forwarding with guaranteeing probabilistic delay bounds in duty-cycled WSNs. Specifically, it reduces the transmission cost by 12%~25% with statistical delay guarantees under various network settings.

The rest of the paper is organized as follows: Section II describes the preliminaries, including network model, assumptions, and motivations behind this work. Section III elaborates the design of EDGE in details. Simulation results are provided in Section IV. Section V surveys the related work. Finally, conclusions are drawn in Section VI.

II. PRELIMINARIES

A. Radio Duty Cycling Model

We consider a connected stationary multi-hop WSN. For the sake of simplicity, we assume a single always-awake sink node is deployed for data collection purpose. However, our solution can be easily extended to scenarios with multiple sinks. Each sensor node operates at a duty-cycle mode. All nodes have their own periodic duty-cycle schedules. These schedules are shared with one-hop neighboring nodes, so that a sender knows the rendezvous time for exchanging data with an intended receiver. A dormant node switches to the active state when (i) it is scheduled to switch to the active state, or (ii) it has some packets to transmit to a receiver that is active at that time. The working cycle $T$ is equally divided by fixed length short time slot $\tau$, where the length of $\tau$ is set to be enough to complete a round-trip packet transmission (including data and ACK). Let $\{t_A(1), t_A(2), \ldots, t_A(k)\}$ represent the sleep/wakeup schedule of node A in one working cycle, where $t_A(k)$ represents the $k_{th}$ active slot in one working cycle.

B. Assumptions

1) Unreliable wireless links: Due to many factors such as interference, attenuation, and fading, wireless communication links between neighboring nodes are unreliable [18]. There has been a lot of existing works on how to efficiently and accurately measure wireless link quality [19]. The MAC layer normally provides the link quality estimation service [20]. It is worth noting that link qualities may change over time. However, empirical studies [21], [22] have shown that the changing rate is slow. Consequently, the link quality updating is normally needed at a low frequency.

2) Local time synchronization: Our design only needs local synchronization among nodes in one-hop neighborhood to ensure the communication rendezvous between a pair of nodes, and thus robust to time synchronization errors. Low-cost and accurate time synchronization for WSNs has been extensively studied in the literature. For example, GLOSSY [23] can achieve synchronization error below $0.4 \mu s$, which is more than sufficient for local synchronization.

C. Observations

This section demonstrates two sets of experimental results collected from our indoor TinyOS/Iris testbed. We uniformly deployed 30 Iris nodes in indoor environment and set the minimal transmission power level at each node so that a 5-hop network is organized. In this experiment we set the length of one time slot $\tau$ as 50ms and set the working cycle length $T$ equal to $200\tau$ for each node [18]. Each node randomly generates a specified duty-cycle working schedule. We first demonstrate the available path diversity in duty-cycled WSNs. Then, we identify the need for capturing the statistical characteristics of E2E delay to guarantee a certain delivery reliability with delay bound.

![Fig. 1. Trade-off between delay and energy in duty-cycled WSNs](image)

1) Energy-Latency Trade-off: In a duty-cycled network, a neighboring node with high link quality may suffer large sleep latency since it has to wait for a certain time until the receiver wakes up. In this case, delay-optimal routing and energy-optimal routing are likely following different paths, which we refer to the path diversity depending on different routing metrics. We compare the performance between two routing schemes in terms of the E2E delay and transmission cost in Fig. 1 by varying the duty-cycles. As shown in Fig. 1(a), the delay-optimal routing scheme increases more than 25% transmission cost compared with the energy-optimal routing. Fig. 1(b) shows that the energy-optimal routing incurs larger E2E delay than the delay-optimal routing. We observe that it is difficult to achieve optimal energy efficiency and E2E delay simultaneously in duty-cycled WSNs.

![Fig. 2. Large E2E delay variance in duty-cycled WSNs](image)

2) Variance and Uncertainty of E2E Delay: Due to the randomness nature of communication links, it has been revealed that the path delay exhibits high variance and instability in WSNs [12], [24]–[26]. To confirm this empirically, we
conducted experiments and obtained statistics of the E2E delays in our testbed, where 1000 packets were routed from the source node to the sink along the delay-optimal path with 1% duty-cycle. As shown in Fig. 2, the actual E2E delays vary significantly, ranging from 175 time slots to 1000 time slots. We also observe that the radio duty cycling (especially low-duty-cycle) operation actually amplifies the variation and the uncertainty of E2E delays. This is because, once encountering a transmission failure, a sender will have to wait for a long time until the receiver wakes up again.

The above observations indicate that, without capturing the statistical characteristics of E2E delay in a duty-cycled network, it is challenging to provide an explicit guarantee to achieve given QoS requirements due to the highly-varying distribution of E2E delays. To efficiently support delay-sensitive data forwarding with delay bound and reliability requirements, it is desirable that we identify not only whether a given delay bound is achievable or not, but also the probability to achieve the performance.

III. MAIN DESIGN

The objective of our design is to explore the routing path diversity to opportunistically minimize transmission cost and statistically guarantee the delay bound for data delivery. This section first presents an overview of the EDGE design, followed by the detailed description of each component.

A. Design Overview

In our network model, it is expected that the sensory data is to be received at the sink node before a designated deadline with a certain reliability requirement. Intuitively, to guarantee the delay bound required by applications, a packet will be transmitted along the delay-optimal path in default. The key idea of EDGE is to deviate an early arriving packet from the default routing path for transmission cost reduction without sacrificing the probabilistic delay guarantee. Overall, the design of EDGE consists of the following three major components, as illustrated in Fig. 3.

* Energy/delay optimal path construction. To provide delay-efficient data forwarding, the delay-optimal path from any node to the sink is built in a recursive and distributed manner at the initialization phase. As a side benefit, the energy-optimal path can be also constructed simultaneously without introducing extra overhead. We call a routing path the delay-optimal or energy-optimal path if the expected E2E delay or energy cost along this path is minimal. Due to the existence of path diversity in duty-cycled WSNs, as shown in Fig. 3(a), normally different nodes are involved in these two optimal paths.

* Computation of delay quantile. Due to the unreliable wireless links, the E2E delay for data forwarding is a random variable. In our design as shown in Fig. 3(b), for any node v, we first derive its delay probability mass function (pmf) along the delay-optimal routing path. Then, we calculate the required delay threshold to guarantee that the sink receives a packet with probability p from node v, defined as p-quantile delay \( D_v(p) \). As shown in Fig. 3(c), each node obtains its p-quantile delay as the statistical delay threshold and shares the pmf with its children nodes for dynamic forwarding decision making.

* Opportunistic forwarding decision making. Given a delivery ratio constraint p and an achievable delay bound (e.g., the maximal \( D_v(p) \) of a network), a packet is forwarded along the delay-optimal path to achieve the routing constraints. Due to the randomness nature of wireless communication, a packet may opportunistically arrive early at a node (e.g., after several consecutive successful transmissions), as shown in Fig. 3(d). In this case, we explore the path diversity in duty-cycled WSNs to minimize communication cost under the condition that the statistical delay bound is satisfied. For this purpose, a node makes a local forwarding decision to switch to the energy-optimal path if meeting certain conditions.

B. Optimal Path Construction

1) EEC and EED Modeling: To facilitate the analysis in duty-cycled WSNs, we use the time-expanded network model to represent a network with radio duty cycling operations [4]. For a node A with k active slots in one working cycle, it can be represented by k time-expanded nodes \( \{A_1, A_2, \ldots, A_k\} \), where \( A_i \) only wakes up at the \( i_{th} \) active slot to receive data. Let \( EEC(A_i) \) denote the expected energy consumption (EEC), and \( EED(A_i) \) denote the expected E2E delay (EED) for a packet that is received by node A at its \( i_{th} \) active slot and finally transmitted to the sink, respectively. Suppose A is sending a packet to a nexthop node B with n active slots in one working cycle. Therefore, A has up to n chances to forward a packet to B in one working cycle, which are sorted in the
order of B’s wakeup times since A’s first trial of forwarding the packet.

Given the EEC and EED values of node B, the sleep/wakeup schedules of A and B, and the link quality between A and B, we can model the data forwarding process using the Markov model, as shown in Fig. 4. Let \( q \) denote the probability of a successful transmission for each trial (i.e., the bi-directional link quality between A and B). After \( n \) consecutive unsuccessful trials, it goes back to the initial state. We calculate \( EEC(A_i) \) as follow:

\[
EEC(A_i) = \sum_{j=1}^{n} q(1-q)^{(j-1)}(j + EEC(B_j)) + (1-q)^n(n + EEC(A_i))
\]

where \( q(1-q)^{(j-1)} \) is the probability that A successfully transmits the packet to B after \( (j-1) \) failed attempts, and \( (j + EEC(B_j)) \) is the corresponding expected transmission cost for the \( j_{th} \) trial of transmission. Thus, we have,

\[
EEC(A_i) = \sum_{j=1}^{n} q(1-q)^{(j-1)}(j + EEC(B_j)) + (1-q)^n(n + EEC(A_i))
\]

Similarly, we can derive \( EED(A_i) \) as follow:

\[
EED(A_i) = \sum_{j=1}^{n} q(1-q)^{(j-1)}(d_j + EED(B_j)) + (1-q)^n
\]

where \( T \) is the length of one working cycle, and \( d_j \) is the sleep latency from A having a packet at its \( i_{th} \) active slot to B receiving the packet at its \( j_{th} \) active slot.

2) Energy/Delay Optimal Path Construction: Considering the minimum EEC and minimum EED as two different forwarding metrics in the route construction, we can build the energy-optimal path and delay-optimal path from any node to the sink [4]. Note that EEC and EED metrics measure the end-to-end expectations, which are different from simply choosing a neighbor that has a smaller hopcount and the best link quality or lowest sleep latency as the nexthop node. The process of building such optimal routing structures starts from the sink, and propagates throughout the network iteratively, which is similar to the distributed Bellman-Ford algorithm. After the sink broadcasting its EEC and EED values (The sink’s EEC and EED values are always 0), nodes that have received the message calculate and then rebroadcast their EEC and EED values to downstream neighbors. This process continues at a node until it receives no updated expectation values from neighbors. It is worth noting that an individual node may receive the broadcasted EEC and EED values from neighbouring nodes multiple times, it only updates and rebroadcasts the expectation values if its own EEC or EED values can be reduced by choosing the message sender as the nexthop.

For a node with multiple active slots, the broadcasted message contains EED values of each active slot individually since different active slots normally have different expectation values due to different sleep latencies. During the recursive calculation of EEC and EED, each node obtains the neighbors that can lead to minimum EEC or EED, and thus builds the minimum EEC and minimum EED paths towards the sink, which is referred to as the energy-optimal path and delay-optimal path, respectively. Since the process of building the energy-optimal path is same as the construction of the delay-optimal path, these two paths will be obtained simultaneously. As shown in Section II-C, although the minimum expected energy and the minimum expected delivery delay are achieved by two different routing paths, respectively, normally neither of them can achieve the optimal delay and energy simultaneously due to the existence of path diversity in low-duty-cycle WSNs.

C. Computation of Delay Quantile

1) Packet Delivery Delay Distribution: Given a routing path, we propose a distributed method to calculate the probability mass distribution (pmf) of packet delivery delay for each node on the path in duty-cycled WSNs. We first specify a time frame \( R \), where the end of \( R \) is considered as the deadline of data packet arrival to the sink. Time slots in \( R \) are indexed from 1 to \( r \) backwardly, where \( r \) is the number of time slots in \( R \), as shown in Fig. 5. Let \( t_B(j) \) denote the time slot index for B’s \( j_{th} \) active time slot within \( R \), where the physical meaning is the remaining time to deadline when B starts to send a packet towards the sink at that time. The packet delay distribution of node B can be denoted by a set of tuples \((t_B(j), p_B(j))\), where \( p_B(j) \) is the corresponding probability of the sink finally receives the packet given the remaining time of \( t_B(j) \). Suppose B is node A’s nexthop node along the routing path. We calculate A’s pmf as follows:

\[
p_A(i) = \sum_{\forall j: \ t_A(i) > t_A(j)} p_B(j) \cdot q \cdot (1-q)^{n_{ij}}
\]

where \( p_A(i) \) is the packet delivery probability to the sink when node A has a packet to send at its \( i_{th} \) active slot within \( R \). \( q \) is the corresponding bi-directional link quality between the sender and receiver. \( n_{ij} \) is the number of B’s active slots.
between $t_A(i)$ and $t_B(j)$ with the exclusion of $t_B(j)$. The term $p_B(j) \cdot q \cdot (1 - q)^n_{ij}$ denotes the probability that node B exactly receives the packet at its $j_{th}$ active slot within $R$.

Similar to the optimal path construction in Section III-B2, the process of pmf computation starts from the sink, and propagates to the rest of the network until all nodes are covered. The initial value at the sink is $(i, 100\%)$ for any time slot $i$ in $R$. With the initial values, according to Eq. 4, we calculate the pmf of sink’s neighboring nodes. Similarly, all the nodes within the network compute their pmf as long as their nexthops’ pmf becomes available. Fig. 5 shows an example to explain how the pmf is computed. The sink node first broadcasts its initial values. Then, node B computes its pmf based on the link quality (i.e., 0.5) towards the sink backwardly. Suppose B receives a packet at $t_B(1)$. Since the sink is always awake (i.e., the corresponding probability value for any time slot in $R$ is 100%), after $t_B(1)$, B has only 2 chances to forward the packet to sink before the deadline. Thus, the probability that sink finally receives the packet is $p_B(1) = 1 - 0.5^2 = 0.75$. Accordingly, for $t_B(2)$ where the slot index is 8 in $R$, the probability $p_B(2) = 1 - 0.5^7 = 0.992$. For B’s $3_{rd}$ active slot, $p_B(3) = 1 - 0.5^{12} \approx 1.0$. Node A derives its own pmf based on B’s pmf and the link quality between A and B according to Eq. 4. For example, the probability of delivering a packet to the sink for $t_A(3)$ becomes $0.8 \times 1.0 + 0.2 \times 0.8 \times 0.992 + 0.2^2 \times 0.8 \times 0.75 = 0.993$.

From Eq. 4, the number of tuples to be calculated in delay distribution equals the number of time slots within $R$ at each node. The pmf computation directly based on Eq. 4 can be accomplished with $O(M^2)$, where $M$ is the number of time slots within the time frame. From Fig. 5, we observe that, $p_A(i)$ is a non-decreasing function of the remaining time, the increased probability becomes marginal as increasing the remaining time, and $p_A(i)$ can not be larger than $p_B(j)$ for $\max\{t_B(j) \in R | t_A(i) > t_B(j)\}$. For example, $p_A(2)$ is smaller than $p_B(2)$, as shown in Fig. 5. Therefore, we can approximate the pmf by reducing the number of calculated entries within $R$, which greatly improves the computation and storage efficiency in implementation considering the intrinsic resource constraints in WSNs. Given the range of possible delivery reliability requirements, e.g., $p \in [50\%, 99\%]$, we only need to calculate limited number of tuples within $R$, from the start point where the first entry’s probability is larger than 50% to the end point where the cumulative probability of the backward remaining entries is less than 1%. Take node A in Fig. 5 for example, we just need to calculate the pmf from $t_A(1)$ to $t_A(3)$ since its cumulative probability of remaining entries is less than 1% if the length of $R$ is larger than 15 time slots. In this way, although $R$ can be any large time frame, the number of considered entries is actually very limited.

From the perspective of delay-efficient data forwarding, a packet is transmitted along the delay-optimal path in default to minimize the expected E2E delay. Therefore, in our design, we calculate the pmf of packet delivery delay based on the delay-optimal path. Given a delay bound with certain delivery ratio constraint, we know whether the desired performance is statistically achievable or not. Since the computation of pmf can be embedded in the process of delay-optimal path construction, there is very little overhead incurred by the computation of delay quantile.

2) Deriving the p-quantile Delay Bound: From the pmf calculation of the delay-optimal routing path, it is easy to find the minimum achievable delay bound for any node $v$ to guarantee that the delivery ratio is not less than $p$, which is denoted as p-quantile delay $D_v(p)$, using the discrete quantile function:

$$D_v(p) = \min \{ t_v(i) \in R | p_v(i) \geq p \}$$

where $t_v(i)$ is the slot index corresponding to $v$’s $i_{th}$ active slot within $R$. For example in Fig. 5, we derive the 0.90-quantile delay of node A in order to guarantee the delivery ratio constraint 0.90 by searching the first entry that has a probability larger than 0.9 backwardly from its pmf. Thus, we have $D_A(0.90) = 10$ (time slot $\tau$ as the unit).

We note that there is a tradeoff between the delivery ratio requirement $p$ and the length of the minimum achievable delay bound. Larger $p$ normally leads to a larger $p$-quantile delay bound. In our design, $D_v(p)$ is used as the guideline in local forwarding decision making to reduce the transmission cost opportunistically. Therefore, during the initialization, nodes will share their pmf values to children nodes. In addition, deriving the $p$-quantile delay bound can also guide the application layer to avoid the selection of unachievable delay bounds.

D. Opportunistic Forwarding Decision

Given an achievable delay bound and the elapsed time since a packet was sent from the source node, the current forwarding node can obtain the remaining time to the deadline of delivering the packet towards the sink. Suppose each packet contains information about the remaining time and the required delivery reliability requirement $p$. We call a packet early arriving when the remaining time is larger than its $p$-quantile delay. Note that for a source node sending a packet to the sink with a smaller $p$-quantile delay threshold than the designated delay bound, we also consider such packet as an early arriving packet. As discussed in Section III-A, a packet will be forwarded opportunistically via the links outside of the default delay-optimal path to improve the energy efficiency if it arrives early. In this case, a forwarding decision will be made in a purely distributed manner at the sending node based on three inputs: (i) the remaining time to delivery deadline of the packet, (ii) the link quality and sleep latency between itself and the nexthop node along the energy-optimal path, and (iii) the $p$-quantile delay of that nexthop node.

Suppose node A receives an early arriving packet at its $i_{th}$ active slot in one working cycle, and intends to make a forwarding decision. After the optimal path construction in Section III-B2, we know each node maintains two nexthop nodes depending on energy or latency routing metrics. Assume B and C is A’s nexthop along the delay-optimal path and energy-optimal path towards the sink, respectively, as shown
in Fig. 6. Node A first computes the expected one-hop delay to node C, denoted as $EOD(A_i, C)$. Given the working schedule of node C, i.e., $n$ active slots in one working cycle, we sort these $n$ active slots in the order of wakeup times from the time A receives the packet (i.e., A’s $i_{th}$ active slot). Based on the analysis model in Section III-B2, we calculate $EOD(A_i, C)$ using the following equation:

$$EOD(A_i, C) = \sum_{j=1}^{n} q^{(1-q)(j-1)} d_{ij} + T(1-q)^n,$$

(6)

where $d_{ij}$ is the sleep latency when sender A first receives a packet at its $i_{th}$ active slot and then waits for B’s $j_{th}$ active slot to transmit the packet. $T$ is the length of one working cycle.

Then, node A will make a decision to forward the packet to node C when meeting the following conditions: (i) node C’s $p$-quantile delay $DC(p)$ plus $EOD(A_i, C)$ is smaller than the packet’s remaining time; and (ii) the link quality between AC is better than the link quality between AB, where B is the nexthop node on the delay-optimal path. Otherwise, the packet will be propagated along the default delay optimal path towards the sink. For example in Fig. 6, assume the remaining time is 200 (time slot $\tau$ as the unit), $DA(p) = 100$, $DC(p) = 150$, $EOD(A_i, C) = 30$. Since $DC(p) + EOD(A_i, C) < 200$, and link quality between AC (i.e., 0.9) is better than link quality between AB (i.e., 0.8), node A will dynamically choose C as its nexthop.

Note that after switching to the energy-optimal path, whenever it does not meet the above two conditions, the packet will switch back to the default routing path until an early arriving case happens again. We can see that such a dynamic forwarding design helps to reduce the transmission cost thus improve the energy efficiency without compromising the statistical delay guarantee. Another distinctive feature of EDGE lies in that it supports diverse statistical delay guarantees (i.e., any reasonable $p$ with an achievable delay bound specified by applications), since each node maintains the pmf locally and the forwarding decision is made in a distributed manner.

Due to the dynamic changes of environment, link qualities may change over time [19]. Therefore, link quality information should be updated periodically [20]. Meanwhile, each node will also update its packet delivery delay distribution and share the pmf with neighboring nodes. Like existing distributed link quality aware routing protocols, the overhead incurred by updating routing information is expected to be amortized over a reasonably long period of network operation with improved routing performance.

IV. Simulations

In this section, we evaluate our EDGE design with various network settings using the ns-2 [27] simulator. To show the efficiency of our design, we compare the performance of EDGE with two statistically optimal performance bounds [4]. (i) MinEED: packets are persistently forwarded via the delay-optimal path, which is constructed using the minimum expected E2E delay (EED) metric. (ii) MinEEC: packets are always forwarded via the energy-optimal path, which is built using the minimum expected number of transmissions (EEC) metric.

A. Simulation Setup

We deploy 300 nodes in a 300$m \times 300$m square field and randomly generate the network topology. The node transmission range is set as 40$m$. Link qualities among nodes are derived from the Nakagami fading model, where PRR is related with the distance between two neighboring nodes. In all simulations, the sink node is positioned at (0$m,0$m). The source node and the sink are at opposite positions to create as large routing hops as possible [12]. We set the working cycle length $T$ equal to 200$\tau$ for each node. By default, the required data delivery reliability $p$ is set to 95%, and the deadline is set as the derived 0.95-quantile delay (i.e., the least achievable delay bound) unless otherwise specified. We have studied the impact of these system parameters. Statistical results have been averaged over 300 runs, and the related standard deviations are provided as error bars.

B. Performance Metrics

We use the following main metrics for comparison. Due to space limit, system insights, e.g., the ratio of routing along the energy-optimal path during data forwarding in EDGE, have been interwoven with the descriptions of the main performance comparison.

- In duty-cycled WSNs, the receiver-side energy consumption is mainly determined by the predefined working schedules, which has no correlation with different forwarding schemes. Therefore, we use the sender-side energy consumption as the performance metric to reflect the energy efficiency of a forwarding protocol. Energy Consumption is measured by the total number of transmissions for a single packet delivery in each simulation round.
- We measure the Deadline Missing Ratio as the ratio of the number of packets received by the sink after their delay bounds to the total number of packets sent by source nodes. This metric depends on the value of the specified deadline.
- Due to the unreliable nature of wireless transmissions, the path delay exhibits certain randomness and can be considered a random variable. To demonstrate the delay distribution, we measure the E2E Delay and its variance, where the E2E delay is the time elapsed from a packet being sent out by the source node to the sink receives it.
C. Simulation Results

1) Impact of Duty-Cycle: We first evaluate the impact of duty cycles on the performance of different protocols, by varying duty-cycle from 1% to 3%. In order to be applicable to diverse duty-cycle schedules, we do not assume any specific working schedule in our simulation. Each node randomly selects active slots within the range of duty-cycle setting.

Fig. 7(a), Fig. 7(b), and Fig. 7(c) plot the energy cost, deadline missing ratio and E2E delay performance under different duty-cycles. From Fig. 7(a), EDGE reduces the energy cost by around 13% on average compared with the MinEED as the duty cycle varies. We observe that EDGE incurs a little higher energy cost than MinEEC. One reason is that we set the deadline as the least feasible delay bound. At the beginning of the data forwarding, EDGE forwards data packets along the MinEED path. Only after one or more successful transmissions (i.e., an early arriving packet is received), EDGE will switch to the MinEEC path for energy saving. It is expected that EDGE will always route data packets along the energy-optimal path from the beginning when the delay bound becomes very loose. Since MinEEC serves as the lower bound of the number of data transmissions, the energy saving achieved by EDGE is actually remarkable. In Fig. 7(b), our design effectively ensures the required 95% delivery reliability requirement. While deadline missing ratio in MinEEC is always higher than 15%, since it does not consider the statistical delay bound guarantee. From Fig. 7(c), the E2E delay decreases significantly as the duty-cycle increases. EDGE yields larger E2E delay than MinEED since switching to the energy-optimal path inevitably increases the sleep latency. From Fig. 7, we can see that EDGE opportunistically optimizes the energy-latency trade-off with ensuring statistical delay bound requirements.

To have a better understanding of different routing schemes, Fig. 8 shows the CDF curves of the E2E delay with different duty-cycles. From the figures, we can see that the delays are distributed in a very large range due to the combined effects of radio duty cycling operation and unreliable links. This observation of highly-varying delay distribution highlights our motivation to investigate the guideline for delay-sensitive reliable data forwarding guarantee, instead of naively using the EED routing metric. Note that the delay bound is always changing since we set it as the derived 0.95-quantile delay, which depends on the specific duty-cycle setting in each simulation round.

An interesting observation regarding Fig. 8 is that, EDGE guarantees the statistical delay bound through shortening the “long tails” of delay distribution compared with the MinEED. Through examination of the simulation traces, we found that with the increase of duty-cycle, path diversity also increases, providing larger room for performance improvement. Correspondingly, the ratio of routing along the energy-optimal path during data forwarding in EDGE increases from 81%, then 86%, to 91% when increasing the duty-cycle from 1% to 3%.

2) Impact of Link Quality: In Fig. 9, we study the impact of link quality on performance for three forwarding schemes by varying the average link quality from 0.5 to 0.8. The duty-cycle is set to 2%. It is easy to understand that both the energy cost and E2E delay decrease as the link quality improves, as illustrated in Fig. 9(a) and Fig. 9(c). However, we observe from Fig. 9(a), EDGE obviously improves the energy efficiency even with increasing the link quality settings. For example, when the average link quality is 0.8, EDGE provides more than 18% energy saving than MinEED. The reason behind this result is plausible, unreliable wireless communication is just one of the two factors that can affect the cause of path diversity, where the other one is the radio duty cycling operation. As long as with less-than-perfect link qualities under duty-cycled mode, essentially, there exist path diversities, which provides room for EDGE to improve the energy efficiency with service guarantee. Another interesting observation regarding
Fig. 9(b) is that, as the link quality becomes better, the deadline missing ratio in MinEEC surprisingly increases from 16% to 38%. This is because, as the link quality increases, the derived $p$-quantile delay bound will be remarkably decreased. Consequently, the risk of missing the deadline in MinEEC is significantly increased. Our EDGE design improves the energy efficiency by 12% ~ 18% while guaranteeing the deadline missing ratio lower than 5%.

3) Impact of Delivery Ratio Requirement: In Fig. 10, we compare the routing performance when varying the delivery ratio requirement $p$ within the corresponding derived $p$-quantile delay bound. The duty-cycle is set to 2% in this test. The performance of both MinEED and MinEEC are not affected by the delivery ratio requirement. However, as shown in Fig. 10(a), when the delivery ratio is increased from 80% to 95%, there is also a consequential increase of energy saving achieved by EDGE. The reason is that, the $p$-quantile delay bound is a nonlinear increasing function of $p$. Larger $p$ indicates more chances for EDGE to switch to the energy-optimal path. The ratio of routing along the energy-optimal path during data forwarding in EDGE increases from 77% to 86% when $p$ is changed from 0.8 to 0.95. Similarly, we can observe the energy-latency trade-off from Fig. 10(a) and Fig. 10(c). Fig. 10(b) plots the corresponding deadline missing ratio. EDGE guarantees the bounded delay with required delivery ratio. While MinEEC suffers a very high deadline missing ratio up to 50%.

V. RELATED WORK

There have been many approaches proposed for minimizing the communication delay for always-active WSNs in the literature [28]–[30]. Recently, more attention has been paid on capturing the statistical characteristics of delay in WSNs due to the unreliable nature of wireless transmissions. A multtimescale estimation (MTE) method is proposed in [12] to derive the probabilistic path delay bounds using probability inequalities based on the estimation of the mean and variance of packet transmission time. Wang et al. [24] present a comprehensive delay performance measurement based on an operational large-scale WSN. Several key characteristics of delay distribution are observed in [24]: (i) delay distribution exhibits randomness; (ii) the median of the delay is relative stable; and (iii) there exist many large delays for most nodes. In [25], the authors develop a comprehensive cross-layer analysis framework to model the distribution of the E2E delay and capture the heterogeneous effects of multihop WSNs.

For many long-term WSN applications, delay-sensitive data forwarding under radio duty cycling operation is of great importance for timely interaction with the physical environment. In the context of duty-cycled WSNs, Gu et al. [4] introduce a dynamic switch-based forwarding (DSF) to minimize the impact of sleep latency due to the low duty-cycle operation. Xiong et al. [31] investigates the multiple task scheduling to alleviate congestion and reduce data loss in duty-cycled WSNs. The authors in [32] propose to reduce data forwarding delay and increases delivery ratio by synchronizing radio activity with available energy. Guo et al. [18] propose a tree-based opportunistic flooding protocol for low duty-cycle WSNs with unreliable links. ADB [33] optimizes the broadcast for asynchronous duty-cycle WSNs through transmission task delegation, so as to avoid retransmissions over poor links by switching to better links when encountering a transmission failure. Olaf et al. [34] propose an opportunistic routing protocol to reduce delay and energy consumption for duty-cycled WSNs.

However, existing delay-sensitive data forwarding works in duty-cycled WSNs either overlook the unreliable property of wireless links (by assuming perfect link qualities), or minimize communication latency without explicit reliability guarantee of delay bounded data delivery since their design objectives are mainly to minimize the expected E2E delay in a best-effort
manner. In addition, the utilization of the path diversity in duty-cycled WSNs has not been properly investigated yet in the literature. In this work, we advance the state-of-the-art by designing an energy-efficient data forwarding scheme, which is able to effectively provide diverse probabilistic delay bound guarantees under an emerging and practical sensor network model with radio duty cycling operations.

VI. CONCLUSION

In this work, we propose EDGE, a novel design tailored for duty-cycled networks to achieve energy-efficient data forwarding with statistical delay bound guarantee. To explicitly ensure deadline sensitive probabilistic data delivery, we capture the statistical characteristics of E2E delay through the distributed delay quantile calculation at each node. With the derived $p$-quantile delay threshold, potential energy saving chances can be accurately identified during data forwarding, which is based on the actual packet reception results at a per packet basis. By exploring the routing path diversity in duty-cycled WSNs, transmission cost can be effectively reduced in an opportunistic manner. To evaluate our work, extensive simulation results show that EDGE achieves remarkable performance improvements, approaching the energy cost lower bound with statistical delay guarantee.

ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China (61170296, 61190125 and 61300174), 973 Program (2013CB035503), China Postdoctoral Science Foundation (2013M530511, 2014T70026), and Open Foundation of State key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications, SKLNST-2013-1-02).

REFERENCES