On Biased Link Sampling in Data-driven Link Estimation and Routing in Low-power Wireless Networks (Invited Paper)

Hongwei Zhang
Department of Computer Science
Wayne State University, USA
hzhang@cs.wayne.edu

Lifeng Sang, Anish Arora
Department of Computer Sci. & Eng.
The Ohio State University, USA
{sangl, anish}@cse.ohio-state.edu

ABSTRACT

The wireless network community has become increasingly aware of the benefits of data-driven link estimation and routing as compared with beacon-based approaches, but the issue of *biased link sampling* (BLS) has not been well studied even though it affects routing convergence in the presence of network and environment dynamics. Focusing on traffic-induced dynamics, we examine the open, unexplored question of how serious the BLS issue is and how to effectively address it when the routing metric ETX is used. For a wide range of traffic patterns and network topologies and using both node-oriented and network-wide analysis and experimentation, we discover that the optimal routing structure remains quite stable even though the properties of individual links and routes vary significantly as traffic pattern changes. In cases where the optimal routing structure does change, data-driven link estimation and routing is either guaranteed to converge to the optimal structure or empirically shown to converge to a close-to-optimal structure. These findings provide the foundation for addressing the BLS issue in the presence of traffic-induced dynamics and suggest approaches other than existing ones. These findings also demonstrate that it is possible to maintain an optimal, stable routing structure despite the fact that the properties of individual links and paths vary in response to network dynamics.

Categories and Subject Descriptors

C.2.2 [**Network Protocols**]: Routing protocols

General Terms

algorithms, measurement, performance

Keywords

Low-power wireless networks, sensor networks, data-driven link estimation and routing, biased link sampling, convergence, stability

WICON'08 , November 17-19, 2008, Maui, Hawaii.

1. INTRODUCTION

Wireless communication assumes complex spatial and temporal dynamics [5, 18, 32, 33], thus estimating link properties is a basic element of routing in wireless networks. One commonly used link estimation method is letting neighbors exchange broadcast beacon packets, and then estimating link properties of unicast data transmissions via those of broadcast beacons. Nonetheless, there are significant differences between unicast and broadcast link properties [6, 22], and it is difficult to precisely estimate unicast link properties via those of broadcast due to temporal correlations in link properties and dynamic, unpredictable network traffic patterns [27, 29, 30]. To address the drawbacks of beacon-based link estimation, the method of data-driven link estimation has been proposed [13, 14, 17, 19, 20, 29, 30] and shown to significantly improve routing performance [30].

In data-driven link estimation, information about the properties of a link is provided by the MAC feedback for unicast data transmissions along the link. If a link is not currently used for data transmission, its current properties will most likely be unknown to the associated node (since the precise correlation among links associated with the same node tends to be complex and difficult to predict). This introduces the issue of *biased link sampling* (BLS) where properties of actively used links are constantly sampled and updated but properties of unused links are not sampled and unknown. BLS is not a problem if link properties are mostly static and do not change temporally. Nonetheless, temporal link dynamics is usually unavoidable due to dynamics in network traffic pattern and traffic-induced interference [29, 30], dynamics in environment [8, 21, 26], and/or node mobility. For instance, Figure 1 shows the network conditions in the presence of different traffic conditions, where network condition is represented by the unicast ETX (i.e., expected number of transmissions required to successfully deliver a unicast packet) for links associated with a randomly selected node in the Kansei testbed (*see* Section 2). We see that unicast ETX changes significantly (e.g., up to 32.44) as traffic pattern and thus co-channel interference varies [30]. Therefore, one may expect that, in the presence of temporal link dynamics, data-driven link estimation and routing may not converge to the optimal solution since, due to BLS, a node may be unable to discover the route that is not currently used but has become optimal.

Even though data-driven link estimation has been used in various forms, the severity that BLS affects routing optimality has not been well studied, and only ad hoc, if any, solutions have been proposed in existing data-driven link estimation and routing schemes. For instance, CARP [19], four-bit-estimation [13], and NADV [20] do not examine the BLS issue; LOF [29] and SPEED [14] exploratively sample alternative routes at randomized but high frequency (i.e., once every few and every single packet transmission respec-

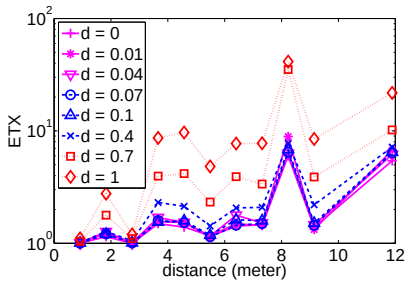


Figure 1: Link unicast ETX in the presence of different network traffic pattern. d denotes the probability that each node generates traffic at an arbitrary moment, and $d = 0$ denotes the case of no traffic in the network and thus zero co-channel interference. The data is for XSM motes (an enhanced version of MICA2 motes) and the B-MAC protocol, but similar phenomena are also observed when other MAC protocols (e.g., S-MAC) and radios (e.g., 802.15.4 and 802.11b radios) are used.

tively), which can reduce routing performance as we will show in Section 6; EAR [17] implicitly addresses the BLS issue by letting every node constantly overhear unicast transmissions around it, but overhearing is not energy-efficient in battery-powered sensor networks (since overhearing increases nodes' duty cycles), and it can lead to estimation errors since, due to MAC coordination mechanisms such as RTS-CTS handshake, the properties of overheard unicast transmissions may be different from those of unicast transmissions to a node itself. Thus, the lack of a thorough understanding of the BLS issue is an important problem since it affects the performance of a basic service in sensor networks — routing.

The objective of this paper is to study the open, unexplored question of how serious the BLS issue is and how to effectively address it in the presence of (potentially unpredictable) network dynamics. We focus on traffic-induced dynamics (i.e., varying network conditions due to changes in network traffic pattern) in this paper, and we relegate detailed study of other network dynamics (e.g., mobility, external interference from other wireless networks) to our future work. Therefore, we focus on mostly static deployment scenarios where environment conditions and nodes are mostly static, even though environment conditions may change slowly and nodes may fail or join the network. Not all sensor network deployments are mostly static, for instance, deployments where environment conditions may frequently change due to interference from other coexisting networks (e.g., 802.11 networks) or due to movement of persons or objects within the deployment space (e.g., a building), or deployments where sensor nodes themselves may be mobile. Nonetheless, mostly static deployment does represent a subclass of sensor network deployments, for instance, in applications where nodes are statically deployed in remote areas for environment monitoring. Moreover, traffic-induced dynamics are universally present in sensor networks, thus addressing the issue in mostly static deployment scenarios may shed light on how to address the issue in other deployment scenarios and how to address other network dynamics.

In studying the impact of BLS on routing optimality, we consider the routing metric ETX which is commonly used in wireless networks (e.g., sensor networks and mesh networks). Through mathematical analysis and testbed-based experimentation, we examine the stability of optimal routes and the severity of BLS. For a wide range of dynamic traffic scenarios (e.g., dynamic events, dynamic data collection, and their mix) and network setups (e.g., grid and random networks) we study, we find out that nodes' best forwarders and the optimal routing structure are rather stable even though the properties of individual links and routes may vary significantly as traffic pattern and network condition change. In cases where the optimal routing structure does change, we prove that data-driven link estimation and routing is guaranteed to converge to the optimal structure when network conditions worsen, and the convergence is quick (e.g., with a median sample size requirement of no more than 7); when network conditions improve, the optimal forwarder chosen for heavy traffic load tends to remain a good suboptimal forwarder for lighter traffic load, even though data-driven routing may not converge to the optimal structure.

These findings provide the foundation for addressing the BLS issue in the presence of traffic-induced dynamics. In contrast to existing approaches, for instance, these findings demonstrate the need to address the BLS issue, the drawbacks of frequent explorative sampling in mostly static networks, and the feasibility of an energy-efficient, light-weight approach to addressing the BLS issue. These findings also demonstrate that it is possible to maintain an effective, stable routing structure despite the fact that the properties of individual links and paths vary in response to network dynamics. Since routing stability enables consistent, predictable routing performance, these findings also suggest that we may regard stability as a basic evaluation criterion for routing metrics.

The rest of the paper is organized as follows. We briefly discuss in Section 2 the routing metric, the routing protocol, and the experimental facility we use in this study. We then analyze the convergence properties of data-driven link estimation and routing in Section 3. We study the dynamics of best forwarders and the routing stability in Section 4 and 5 respectively, and we discuss how to address the BLS issue in Section 6. We discuss related work in Section 7 and make concluding remarks in Section 8.

2. PRELIMINARIES

In this section, we discuss the routing metric, the routing protocol, and the experimental facility that we use in the analytical and/or experimental study of this paper.

Routing metric and protocol. We use the routing metric ETX (i.e., expected number of transmissions for delivering a data packet) [7, 28] in our study, and we use the data-driven link estimation and routing method L-ETX [30] for estimating the ETX metric for each link and path. In L-ETX, MAC feedback for unicast data transmissions are used to calculate the reliability PDR of individual unicast-physical-transmissions[1] along a link, then the ETX of this link is derived as $\frac{1}{\text{PDR}}$; the ETX metric of a path is the sum of the ETX values of the individual links along the path.

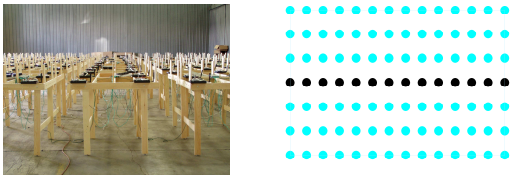
For the analysis of Sections 3 and 4, we also use a localized, geographic routing metric ETD (for *ETX per unit-distance to destination*) in evaluating the goodness of forwarder candidates. ETD is a geographic version of ETX, and it is defined as follows. Given a sender S, a neighbor R of S, and the destination D, the ETD via R is defined as

$$\begin{cases} \frac{ETX_{S,R}}{L_{S,D}-L_{R,D}} & \text{if } L_{S,D} > L_{R,D} \\ \infty & \text{otherwise} \end{cases} \quad (1)$$

[1]In many MAC protocols such as the B-MAC [24] and the IEEE 802.15.4 MAC, a unicast packet is (re)transmitted until being successfully delivered or until the number of transmissions exceeds a certain threshold value (e.g., 8). For convenience, we regard each individual transmission involved in transmitting a unicast packet as a unicast-physical-transmission.

where $ETX_{S,R}$ is the ETX of the link from S to R, $L_{S,D}$ denotes the distance from S to D, and $L_{R,D}$ denotes the distance from R to D. Zhang *et al.* [30] have shown that this local, geographic metric performs in a similar way as the global, distance-vector metric ETX for uniformly distributed networks; we will also show in Section 5 that phenomena observed through ETD based analysis and measurements in Sections 3 and 4 carry over to cases where the measurements are based on ETX.

Experimental facility. For the experimental study in Sections 3, 4, and 5, we use the publicly available sensor network testbed *Kansei* [12]. In an open warehouse with flat aluminum walls (see Figure 2(a)), Kansei deploys 98 XSM motes [11] in a 14×7 grid (as shown in Figure 2(b)) where the separation between neighboring grid points is 0.91 meter (i.e., 3 feet). The grid deployment pattern



(a) Kansei (b) 14×7 grid

Figure 2: Sensor network testbed *Kansei*

enables experimentation with regular, grid topologies, as well as random topologies (e.g., by randomly selecting nodes of the grid to participate in experiments). XSM is an enhanced version of Mica2 [2] mote, and each XSM is equipped with a Chipcon CC1000 [1] radio operating at 433 MHz. To form multihop networks, the transmission power of the CC1000 radios is set at -14dBm (i.e., power level 3) for the experiments of this paper unless otherwise stated. XSM uses TinyOS [4] as its operating system. For all the experiments in this paper, the default TinyOS MAC protocol B-MAC [24] is used; a unicast packet is retransmitted, upon transmission failure, at the MAC layer (more specifically, the TinyOS component QueuedSend) for up to 7 times until the transmission succeeds or until the 8 transmissions have all failed; a broadcast packet is transmitted only once at the MAC layer (without retransmission even if the transmission has failed).

3. BIASED LINK SAMPLING AND ROUTING CONVERGENCE

Taking the data-driven link estimation and routing method L-ETX [30] as an example, we analyze in this section the convergence properties of data-driven routing in the presence of biased link sampling (BLS) and traffic-induced dynamics (i.e., network dynamics introduced by varying network traffic patterns).

When network traffic pattern changes, the quality of a link may become worse (e.g., when receiver-side interference increases) or better (e..g., when receiver-side interference decreases). It turns out that these two types of link quality changes have different impact on data-driven protocols, as we show below.

PROPOSITION 1. *In the presence of biased link sampling and when an unused route becomes better than the currently used one, the convergence of data-driven routing depends on the relative change in the quality of the unused route; routing converges to the optimal if the quality of the unused route has deteriorated, otherwise routing does not converge.* □

Proof sketch: Consider a node S that is currently using a route P_0 through forwarder/neighbor R_0. Without loss of generality, let us consider another route P_1 through forwarder candidate R_1.

If the quality of P_1 becomes better than both its own earlier quality and the current quality of P_0, node S will not know, due to the issue of biased link sampling, that P_1 has become better than P_0 and will continue using the suboptimal route P_0 instead of the optimal route P_1. Therefore, data-driven routing does not converge to the optimal solution in this case.

On the other hand, if the quality of P_1 becomes worse than its own earlier quality but better than the current quality of P_0, the the current quality of P_0 will be worse than P_1's quality before the network condition change. Since S keeps in its routing table P_1's quality before the condition change, S will regard P_1 being a better route than P_0 and will change to P_1. Once S starts to use P_1, it resamples P_1 and link estimation will converge to the latest quality of P_1.

□

From Proposition 1, we can analyze the behavior of L-ETX in cases of improving network conditions and deteriorating network conditions separately. We first analyze the convergence speed when network condition deteriorates (i.e., link and route quality worsens). To this end, we first analyze the sample size requirement in L-ETX for identifying the best forwarder. We assume that the ETD metric (i.e., the geographic-version of the distance-vector protocol L-ETX) is used since it enables us to have a closed-form solution as shown below.

PROPOSITION 2. *Given a sender S, the destination D, and two of S's forwarder candidates K_1 and K_2 that are closer to D than S itself and whose corresponding unicast-physical-transmission reliability is P_1 and P_2 respectively, the sample size n that is sufficient to distinguish the relative goodness of K_1 and K_2 at 100(1-α)% confidence level is $(\frac{Z_{1-\alpha/2}(L_1\sqrt{P_1(1-P_1)}+L_2\sqrt{P_2(1-P_2)})}{L_1P_1-L_2P_2})^2$, where L_1 is the distance from S to D minus that from K_1 to D, L_2 is the distance from S to D minus that from K_2 to D, and $Z_{1-\alpha/2}$ is the (1-α/2)-quantile of the standard Gaussian variable N(0, 1).* □

Proof sketch: For a link with unicast-physical-transmission reliability P that is calculated based on n number of physical transmissions, the confidence interval (CI) for the packet delivery rate at significance level α (i.e., at 100(1-α)% confidence level) is $[P - Z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}, P + Z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}]$ [16]. Thus, for the two links with packet delivery rate P_1 and P_2 respectively, the corresponding CIs are as follows:

$$CI_1 = [P_1 - Z_{1-\alpha/2}\sqrt{\frac{P_1(1-P_1)}{n_1}}, P_1 + Z_{1-\alpha/2}\sqrt{\frac{P_1(1-P_1)}{n_1}}]$$
$$CI_2 = [P_2 - Z_{1-\alpha/2}\sqrt{\frac{P_2(1-P_2)}{n_2}}, P_2 + Z_{1-\alpha/2}\sqrt{\frac{P_2(1-P_2)}{n_2}}]$$

The CIs for the corresponding routing metric ETDs are therefore as follows:

$$CI_1' = [\frac{1}{L_1(P_1+Z_{1-\alpha/2}\sqrt{\frac{P_1(1-P_1)}{n_1}})}, \frac{1}{L_1(P_1-Z_{1-\alpha/2}\sqrt{\frac{P_1(1-P_1)}{n_1}})}]$$
$$CI_2' = [\frac{1}{L_2(P_2+Z_{1-\alpha/2}\sqrt{\frac{P_2(1-P_2)}{n_2}})}, \frac{1}{L_2(P_2-Z_{1-\alpha/2}\sqrt{\frac{P_2(1-P_2)}{n_2}})}]$$

Without loss of generality, we assume that we take equal number n of samples for both links (i.e., $n_1 = n_2$), and suppose that we want to calculate the required sample size n so that K_1 is no worse

a forwarder candidate than K_2. Then a sufficient condition [16] is as follows:

$$\frac{1}{L_1(P_1 - Z_{1-\alpha/2}\sqrt{\frac{P_1(1-P_1)}{n_1}})} \leq \frac{1}{L_2(P_2 + Z_{1-\alpha/2}\sqrt{\frac{P_2(1-P_2)}{n_2}})}$$

which implies that

$$n \geq (\frac{Z_{1-\alpha/2}(L_1\sqrt{P_1(1-P_1)} + L_2\sqrt{P_2(1-P_2)})}{L_1P_1 - L_2P_2})^2$$

Thus the minimum sample size required is

$$(\frac{Z_{1-\alpha/2}(L_1\sqrt{P_1(1-P_1)} + L_2\sqrt{P_2(1-P_2)})}{L_1P_1 - L_2P_2})^2$$

□

To get numerical results on the sample size requirement, we consider the case where the sender on the left end of the middle row of Figure 2(b) needs to select the best next-hop forwarder among the set of receivers in the middle row, and the destination is far away from the sender but in the direction extending from the sender along the middle row to the right. (Phenomena similar to what we will present have been observed for other sender-receiver pairs too.) To calculate the sample size required by the sender to identify the best forwarder, we need to measure the unicast-physical-transmission reliability from the sender to each receiver. To this end, we let the sender transmit 15,000 unicast packets to each of the receivers where each packet has a data payload of 30 bytes. Based on packet reception status (i.e., success or failure) at the receivers, we measure the unicast-physical-transmission reliability for each link. Using these data, we calculate the sample size required for comparing every two links, and then the sample size required to identify the best forwarder is the maximum of the sample size requirement for pair-wise comparison.

To understand the potential impact of traffic-induced interference on sample size requirement, we randomly select 42 motes out of the light-colored (of color cyan) 6 rows of Figure 2(b) as *interferer*s, with 7 interferers from each row on average. Each interferer transmits unicast packets (of payload length 30 bytes) to a destination randomly selected out of the other 41 interferers. (*Note* that, even though the overall traffic pattern in low-power wireless sensor networks tends to follow certain regular patterns, e.g., flowing from sources to a common sink, the local traffic pattern around the neighborhood of a node tends to be much more irregular. We will also show in Section 5 that the phenomena observed via the local, random traffic patterns carry over to experiments where sensor network specific traffic patterns are studied.) The load of the interfering traffic is controlled by letting interferers transmit packets with a certain probability d whenever the channel becomes available. In our experiments, we measure the unicast-physical-transmission reliability from the sender to its receivers when d is 0, 0.01, 0.04, 0.07, 0.1, 0.4, 0.7, and 1 respectively. Thus the interfering traffic pattern is controlled by d in this case. (Phenomena similar to what we will present have been observed for other interfering traffic patterns, for instance, with different spatial distribution and different number of interferers.)

Based on Proposition 2, we analyze the sample size requirements in the above interference scenarios, and Table 1 shows the median sample size required to identify the best forwarder at 95% confidence level. We see that the number of required physical-transmission-samples tends to be small; for instance, it may only take a very few number of unicasts to collect the required samples. This implies that data-driven link estimation tends to converge quickly. The quick convergence in link estimation implies that the routing structure in L-ETX can converge to the optimal one in a timely manner when network condition worsens (e.g., when network traffic load increases) to the degree that the optimal structure changes.

d	0	0.01	0.04	0.07	0.1	0.4	0.7	1
Median sample size	4	3	5	4	5	7	5	4

Table 1: Median sample size required to identify the best forwarder at 95% confidence level

From Proposition 1, we know that, due to BLS, L-ETX may not converge to the optimal solution when network condition improves. So, the questions are *how this issue of potential divergence affects routing optimality and how to address it*. We explore answers to these questions in the next section.

4. DYNAMICS OF BEST FORWARDERS

To provide guidelines on addressing the BLS issue in the presence of traffic-induced dynamics, we study in this section how the best forwarder of a node may change with traffic pattern. We first study the dynamics of best forwarder through mathematical analysis so that we can examine the issue in generic, different network setups, and then we verify the analytical results through testbed based experimentation.

4.1 Mathematical analysis of best forwarders

To get closed-form solutions, we use the ETD metric to evaluate the goodness of different forwarder candidates as we did in Section 3. We first present the analytical method and then the numerical results for different network setups.

Analytical method. To evaluate the goodness of a forwarder candidate using the ETD metric, we need to analyze the packet delivery rate (PDR) of the corresponding link in the presence of dynamic traffic/interference patterns. To this end, we need to analyze the interference at the forwarder candidate in different traffic scenarios so that we can calculate the signal-to-interference-and-noise-ratio (SINR) based on which we calculate the PDR.

To calculate the interference at a forwarder candidate (which is the packet receiver from the perspective of the sender), we adapt the interference model proposed by Qiu $et\ al.$ [25] to determine the concurrent transmissions (and thus the interference) in a network. In Qiu's model, the behaviors of IEEE 802.11 MAC in multi-hop networks are modeled using a Markov chain where the state i is the set S_i of nodes that are transmitting concurrently at a certain time moment. To adapt Qiu's model to the analysis of B-MAC, we need to adapt the probability $P_{01}(m|S_i)$ that a node m starts to transmit when the system is at state i. This is because TinyOS B-MAC [4] differs from 802.11 in how channel access is coordinated. Due to the limitation of space, we relegate to [31] the detailed derivation of the adapted model.

Using the adapted model, we can calculate the stationary probability π_i for each state i. Then, for each pair of transmitter t and receiver s, the interference that concurrent transmissions have at node s is

$$\sum_{i:t\in S_i} \sum_{j:j\in S_i, j\neq t} \pi_i Pow(j,s),$$

where $Pow(j,s)$ is the received signal strength at s for signals coming from j. $Pow(j,s)$ can be calculated using the log-normal path loss model as in [33]. Then, the SINR at receiver s, denoted

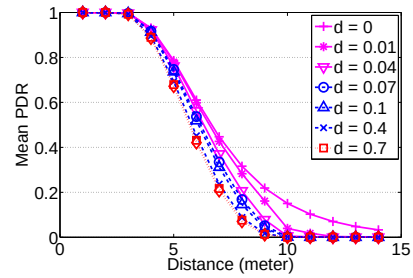
by $SINR(t,s)$, calculates as follows:

$$SINR(t,s) = \frac{Pow(t,s)}{N_0 + \sum_{i:t \in S_i} \sum_{j:j \in S_i, j \neq t} \pi_i Pow(j,s)}$$

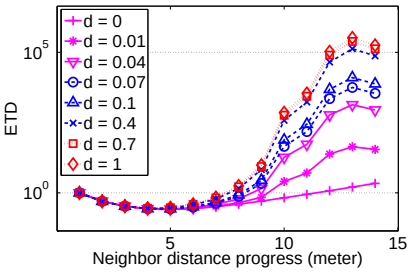
where N_0 is the background noise. Accordingly, we can calculate the packet delivery rate $PDR(t,s)$ from t to s as a function of $SINR(t,s)$, using the model proposed by Zuniga *et al.* [33], and thus we can calculate the corresponding ETD metric value. Having derived the ETD metrics for each forwarder candidate of a node, we can determine which is the best forwarder with the minimum ETD metric value.

Numerical results. Using the above models, we analyze the PDR and ETD in different scenarios, including randomly distributed and regularly distributed nodes, and for indoor and outdoor environments. In our network setups, radio transmission power is set as -14dBm, path loss exponent is set as 3.3 and 4.7 for indoor and outdoor environments respectively, and background noise is set as -105dBm and -100dBm for indoor and outdoor environments respectively. Given the high space complexity of Qiu's model [25], we can only run in Matlab the adapted model with no more than 44 transmitting nodes in our computer (which is a Dell Optiplex GX620 with 4GB memory). Thus we run the model in networks of around 40 transmitting nodes.

Due to the limitation of space, here we only discuss the case of indoor, randomly distributed interferers, and we refer interested readers to [31] for other cases where similar phenomena are observed. We consider a network setup that is the same as what have presented in Section 3 except that 1) the grid space is 15×7 and 2) the distance between any two closest grid points is 1 meter. Then, we let the node at the left end of the middle row serve as the sender, the rest nodes in the middle row serve as forwarder candidates, and the destination is far away from the sender in the direction extending from the sender to the forwarder candidates. Figure 3 shows



(a) PDR



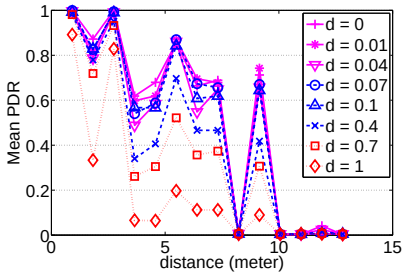
(b) ETD

Figure 3: PDR and ETD in an indoor environment with 42 randomly distributed interferers

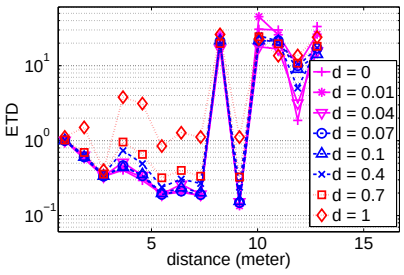
the PDR and ETD in this setup. We see that PDR and ETD change significantly with interference patterns, especially for links of lower PDR. Yet the best forwarder remains rather stable: it is node n_5 that is 5 meters away from the sender, except for the cases when $d = 0.7$ and $d = 1$ where the best forwarder is node n_4 that is 4 meters away from the sender. With other protocols (e.g., congestion control) in place, a network usually works under load much lighter than $d = 0.7$; in fact, Ng *et al.* [23] showed that the optimal traffic injection rate d is 0.245 in a regular linear topology, and the optimal d will be even lower in common, two-dimensional networks. Therefore, the optimal forwarder will not change if the network congestion level is well controlled (e.g., through congestion control). Moreover, the ETD value of n_4 is less than 5.51% more than that of n_5; this implies that, even though a node may be unable to find the optimal forwarder when network condition improves (i.e., interference level decreases), it may still be okay for a node to use the suboptimal forwarder since its performance is very close to the optimal.

4.2 Experimental analysis of best forwarders

To experimentally verify the analytical observations, we use the data collected in Section 3 for network conditions in the Kansei testbed. As in Section 3, we consider the case where the sender on the left end of the middle row of Figure 2(b) needs to select the best next-hop forwarder among the set of receivers in the middle row, and the destination is far away from the sender but in the direction extending from the sender along the middle row to the right. Figure 4 shows the PDR and ETD in different interference/traffic



(a) PDR



(b) ETD

Figure 4: PDR and ETD in the Kansei testbed with 42 randomly distributed interferers

scenarios. The results are more complex than in analysis in the sense that the PDR and ETD are not monotonic functions of the sender-forwarder-distance due to real-world factors such as hardware heterogeneity. Nonetheless, the dynamics of the best forwarder assumes a similar pattern: despite the huge variations in

PDR across different interference scenarios, the best forwarder is the node that is 9.15 meters (i.e., 30 feet) away from the sender in all the scenarios except for the case when $d = 1$; when $d = 1$, the best forwarder is 2.74 meters (i.e., 9 feet) away from the sender. This result is rather consistent with the indoor, analytical results as shown in Figure 3, even though there is slight differences due to differences in network setup and environment conditions. As we have discussed earlier, well-controlled traffic load in multi-hop wireless networks is usually much lighter than 0.7 and 1 [23], thus the best forwarder remains the same across different interference scenarios. We discuss exceptional scenarios of extreme traffic load in Section 6.

4.3 Summary

From the above mathematical and experimental analysis for random and grid topologies and for indoor and outdoor environments, we observe that the best forwarder remains quite stable even though PDR (and even ETD) changes with dynamic traffic patterns. Even though the best forwarder may change when traffic load dynamics passes through a threshold traffic load value, the best forwarder remains the same for a wide range of traffic scenarios. An intuitive explanation for this high stability in best forwarder, in contrast to the much more dynamic link reliability and routing metric value, are that there is usually a guard margin between the routing metric values of the best forwarder and other forwarders, and that, due to the positively correlated impact that each interferer's signal has on the best and other forwarders, it may take a significant change in traffic (and thus interference) pattern to overcome the guard margin as we have seen in our analysis. That is, the guard margin between the best forwarder and the other forwarders tends to mask the impact of traffic-induced dynamics.

In the varieties of scenarios we studied, moreover, the threshold value is either very low (e.g., less than 0.07) [31] or very high (e.g., greater than 0.7). When the threshold load is low, it does not matter much even if routing does not converge to the optimal forwarder when network traffic load decreases to pass the low threshold value; this is because the chosen suboptimal forwarder may well be close to optimal in performance, and optimality is less of a concern for light traffic load (when it is easy to ensure packet delivery reliability) [31]. When the threshold load value is high, it is rarely the case that we would expect to see network traffic load exceeding the threshold in practice when other protocols are in place to control the network congestion level, and thus the best forwarder tends to remain the same across different admissible traffic scenarios.

In the next section, we corroborate these observations by examining the behaviors of L-ETX in different dynamic traffic patterns and network setups. We discuss in Section 6 how to address the exceptional cases where best forwarders may change in a manner that significantly affect network performance.

5. ROUTING WITH DYNAMIC TRAFFIC PATTERNS

Having analyzed the convergence behaviors of data-driven routing and the dynamics of best forwarders, we experimentally evaluate the behaviors of L-ETX in the presence of three types of sensornet-specific dynamic traffic patterns: dynamic events, dynamic periodic data, and mixed dynamic events and periodic data. We also use both grid and random network topologies in this experimental study. Due to the limitation of space, here we only discuss the case of dynamic events and grid network topology, and we refer the readers to [31] for other cases (e.g., random network topology) where similar phenomena are observed.

We use a publicly available event traffic trace for a field sensor network deployment [3] to generate dynamic events in our study. Since the traffic trace is collected from 49 nodes that are deployed in a 7×7 grid, we randomly select and use a 7×7 subgrid of the Kansei testbed (as shown in Figure 2(b)) for experimentation. The mote at one corner of the subgrid serves as the base station, the other 48 motes generate data packets according to different traffic patterns, and the destination of all the data packets is the base station.

We use the event traffic trace mentioned above, but we control the set of nodes that actually generate source packets to control the event size, through which we generate dynamic events. More specifically, we study the following dynamic events which contain 7 event configurations:

$$1 \times 1 \rightarrow 3 \times 3 \rightarrow 5 \times 5 \rightarrow 7 \times 7 \rightarrow 5 \times 5 \rightarrow 3 \times 3 \rightarrow 1 \times 1$$

where each configuration specifies the subgrid of traffic sources. For instance, "3×3" specifies that the nodes in the farthest 3×3 subgrid from the base station generate event traffic. For each event configuration, we generate the associated event 40 times and measure the performance of L-ETX for this event configuration.

Examining the routes taken by packets from each node, we observe that there are very few route changes during the whole experiment. For instance, Table 2 shows the statistics of comparing the routes taken by every two consecutive packets from a same node: 99.98% of the time, consecutive packets use the same route, and only 0.02% of the time the route changes to be a longer one. The high stability of routes in the presence of dynamic traffic patterns are due to the following reasons: 1) estimation in L-ETX is very accurate and stable [30]; and 2) the best forwarder does not change much across different network traffic conditions as discussed in Section 4.

Consecutive routes	Same	Diff. route, same hop length	Increased hop length	Decreased hop length
Radio (%)	99.98	0	0	0.02

Table 2: Routing stability in the presence of dynamic events: grid network

Because of the stability in routing, packet delivery performance is rather consistent across similar network setups. Figure 5 shows the boxplot of event reliability for each event configuration, and Table 3 shows the median event reliability and its 95% confidence level confidence interval (CI) in different configurations. We see that, despite random variations, the event reliability for configurations #3 and #5 are similar to each other, and their CIs overlap with

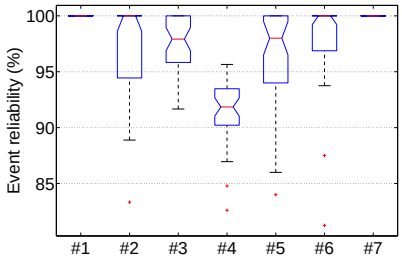


Figure 5: Event reliability for dynamic events: grid network

Config.	Median (%)	CI (%)
#1	100	(100, 100)
#2	100	(98.44, 100)
#3	97.92	(96.92, 98.62)
#4	91.85	(91.3, 92.39)
#5	98	(96, 99)
#6	100	(98.94, 100)
#7	100	(100, 100)

Table 3: The median event reliability and its 95% confidence level confidence interval for dynamic events: grid network

each other. A Wilcoxon Rank Sum [15] test shows that configurations #3 and #5 have equal median event reliability at the 95% confidence level. The same observation applies to other similar traffic patterns, that is, configurations #2 and #6, and configurations #1 and #7.

We also examine the detailed route information, for instance, the hop length and the end-to-end transmission count of routes. Using Wilcoxon Rank Sum tests, we find out that, at 95% confidence level, routes chosen by nodes equal distance away from the base station have equal median hop length and end-to-end transmission count in similar network setups (e.g., configurations #3 and #5). For instance, Figure 6 shows the (statistically) similar end-to-end

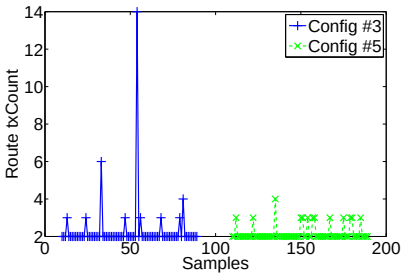


Figure 6: Time series of route transmission counts for a node 9 grid-hops away from the base station: dynamic events and grid network

route transmission counts for the routes taken by packets from a node 9 grid-hops away from the base station in configurations #3 and #5. The same observations apply to parameters such as the per-hop geographic distance and the per-hop transmission count of links used in similar network setups.

Verification. To corroborate the fact that the best forwarders are actually stable in the presence of the different dynamic traffic scenarios discussed above, we implement a variant of L-ETX, which we call *L-ETX-rcv*. L-ETX-rcv is the same as L-ETX except that the forwarder candidates of a node always try to overhear the unicast packet transmissions from the node. Note that the overhearing in L-ETX-rcv is similar to that in EAR [17], but unlike in EAR which studied 802.11b networks with the RTS-CTS mechanism, the B-MAC used in our study does not use RTS-CTS handshake, and thus overheard transmissions by a node f in L-ETX-rcv have the same properties as those of unicast transmissions destined for f itself. Therefore, a forwarder candidate f in L-ETX-rcv can determine, based on the overheard data transmissions, the latest link properties for unicast data transmissions from the sender s to itself, and then f can share this information with s so that s can make the right decision in choosing the optimal forwarder. Consequently, biased-link-sampling is not an issue in L-ETX-rcv due to the receiver-assisted, data-driven link estimation. We run L-ETX-rcv in the different dynamic traffic scenarios and network setups discussed earlier, and we find out that, similar to L-ETX, there is very little route changes and the best forwarders remain stable despite the traffic dynamics.

6. IMPLICATIONS FOR PROTOCOL DESIGN

We see from the findings of Sections 3, 4, and 5 that, despite BLS, L-ETX converges quickly when network condition deteriorates (e.g., due to increased traffic load). For the wide range of dynamic traffic scenarios and network setups we studied, we also see that even though 1) data-driven protocols may, theoretically speaking, not converge to the optimal solution when network condition improves, e.g., due to decreased traffic load, and 2) link properties do change significantly as traffic pattern changes, the best forwarders remain quite stable (in which case BLS is not a problem any more), or the optimal forwarder chosen for heavy traffic load may still be a very good suboptimal forwarder for lighter traffic load [31].

In our study, we have examined a wide spectrum of dynamic traffic scenarios (e.g., dynamic events, dynamic data collection, and their mix) and network setups (e.g., grid and random networks), but we understand that we have not covered all the scenarios that may exist in practice. For the mostly static deployment scenarios we studied, however, our findings on the high stability of the optimal routing structure in spite of dynamics of link properties are themselves not obvious and shed new light on how to address the BLS issue in mostly static networks such as those for remote environmental monitoring. On one hand, we are assured of the good performance of L-ETX in a variety of traffic conditions even if we do not design special mechanisms to address the BLS issue. On the other hand, to address the rare cases where improved network condition leads to significantly worse-than-optimal performance (e.g., due to slow but significant changes in environment conditions [21]), a sender can proactively sample unused links/routes, or the forwarder candidates can proactively overhear the sender's data transmissions to estimate the latest link quality; considering the quick convergence of L-ETX and the low probability or frequency that improved network condition may lead to significantly worse-than-optimal performance, however, the proactive sampling or forwarder-assisted receiver-side link estimation can be executed at very low frequency to reduce the overhead of proactive sampling or overhearing. This is in contrast to the existing approaches in LOF [29] and SPEED [14] where a node periodically samples unused links/routes by using them to deliver data packets, which leads to reduced routing performance due to frequent sampling of the links/routes that are not or not even close to be optimal. We have observed through experimentation that the periodic, probabilistic sampling in LOF and SPEED can lead to bad performance, especially when traffic load is high (e.g., the 7×7 event traffic trace [3]).

L-ETX can deal with dynamics such as node/link failure or node join in a straightforward manner. Node or link failure can be regarded as the case where the quality of the associated link(s) deteriorates (in fact, to be unusable), in which case L-ETX converges quickly; node join can be handled effectively through the initial link sampling procedure when a new node and the associated links first become up. Besides traffic-induced dynamics, our study in this paper has not focused on other network dynamics such as node mobility [9] and quickly changing environment conditions. Even

though we expect that the quick convergence and high stability of L-ETX routing may also help us design light-weight approaches to address these types of network dynamics, detailed study of this is beyond the scope of this paper, and we relegate it as a part of our future work.

7. RELATED WORK

Data-driven link estimation where MAC feedback for unicast data transmissions is used for estimating unicast link properties has been used in several sensor network routing protocols [13, 14, 17, 19, 20, 29, 30], and it has been shown that data-driven link estimation significantly improves estimation accuracy and routing performance as compared with beacon-based approach [30]. Nonetheless, the impact of biased link sampling (BLS) on routing optimality and the severity of the BLS issue in the presence of network dynamics are mostly unexplored. Lack of deep understanding of these issues has led to ad hoc approaches to explicitly or implicitly addressing the BLS issue. As a first step toward systematic treatment of the BLS issue in data-driven link estimation and routing, we have studied in this paper the routing convergence and optimality in the presence of traffic-induced dynamics, and the findings provide new insight into the BLS issue and suggest alternatives to existing approaches in data-driven link estimation and routing.

Ramachandran $et.\ al$ [26] studied routing stability (based on metric ETT [10]) in static wireless mesh networks. The study in [26], however, used broadcast-beacon based link estimation method, and it did not consider the errors in beacon-based link estimation. The study on routing stability in [26] was also based on link quality data collected in the absence of data traffic, and it did not consider the impact of network traffic pattern on link and path properties and thus not the impact of traffic-induced dynamics. Das $et.\ al$ [8] studied the stability of different routing metrics, but they did not focus on routing stability which we have shown to be different from the stability of individual routing metrics. It was not the focus of [8] to examine the BLS issue in data-driven link estimation and routing either.

Lin $et.\ al$ [21] proposed an adaptive transmission power control mechanism that controls radio transmission power level to ensure consistent link properties in the presence of environment dynamics. We have mainly focused on intra-network, traffic-induced dynamics in this paper, and we did not focus on environment dynamics. Nonetheless, the adaptive transmission power control mechanism of [21], if deployed, will make the findings of this paper applicable to a broader sensor network scenarios including those with quickly changing environment conditions.

8. CONCLUDING REMARKS

We have studied the open, unexplored issue of biased link sampling (BLS) in data-driven link estimation and routing. For a wide range of traffic patterns and network setups we studied, we discover that the optimal routing structure remains quite stable despite the significant variations in link properties and route metric values. For the rare cases where the optimal routing structure does change, we prove that, despite the BLS issue, data-driven link estimation and routing is guaranteed to quickly converge to the optimal structure when network conditions deteriorate; when network conditions improve, we empirically show that the optimal structure for heavy traffic load tends to remain a good suboptimal structure for lighter traffic load, even though data-driven routing may not converge to the optimal. These findings shed new light on the BLS issue and provide the foundation for a simple, light-weight mechanism of addressing the BLS issue in the presence of traffic-induced dynamics.

The highly stable routing structure in L-ETX provides a stable, consistent infrastructure for data transport and can help ensure predictable QoS in the presence of traffic dynamics; detailed study of this will be an interesting topic for future research. We have focused on traffic-induced dynamics in this paper, detailed study of how other network dynamics (such as node mobility and quickly changing environmental conditions) affect the stability of optimal routing and data-driven link estimation is also a part of our future work.

9. REFERENCES

[1] Chipcon CC100 RF transceiver. http://focus.ti.com/lit/ds/symlink/cc1000.pdf.

[2] Crossbow Mica2 motes. http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/MICA2_Datasheet.pdf.

[3] An event traffic trace for sensor networks. http://www.cs.wayne.edu/∼hzhang/group/publications/Lites-trace.txt.

[4] TinyOS. http://www.tinyos.net/.

[5] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris. Link-level measurements from an 802.11b mesh network. In *ACM SIGCOMM*, 2004.

[6] I. Chakeres and E. Belding-Royer. The utility of hello messages for determining link connectivity. In *WPMC*, 2002.

[7] D. S. J. D. Couto, D. Aguayo, J. Bicket, and R. Morris. A high-throughput path metric for multi-hop wireless routing. In *ACM MobiCom*, 2003.

[8] S. M. Das, H. Pucha, K. Papagiannaki, and Y. C. Hu. Studying wireless routing link metric dynamics. In *ACM IMC*, 2007.

[9] R. Draves, J. Padhye, and B. Zill. Comparison of routing metrics for static multi-hop wireless networks. In *ACM SIGCOMM*, 2004.

[10] R. Draves, J. Padhye, and B. Zill. Routing in multi-radio, multi-hop wireless mesh networks. In *ACM MobiCom*, 2004.

[11] P. Dutta, M. Grimmer, A. Arora, S. Bibyk, and D. Culler. Design of a wireless sensor network platform for detecting rare, random, and ephemeral events. In *IEEE/ACM IPSN/SPOTS*, 2005.

[12] E. Ertin, A. Arora, R. Ramnath, M. Nesterenko, V. Naik, S. Bapat, V. Kulathumani, M. Sridharan, H. Zhang, and H. Cao. Kansei: A testbed for sensing at scale. In *IEEE/ACM IPSN/SPOTS*, 2006.

[13] R. Fonseca, O. Gnawali, K. Jamieson, and P. Levis. Four-bit wireless link estimation. In *ACM HotNets*, 2007.

[14] T. He, J. Stankovic, C. Lu, and T. Abdelzaher. SPEED: A stateless protocol for real-time communication in sensor networks. In *IEEE ICDCS*, 2003.

[15] M. Hollander. *Nonparametric statistical methods*. Wiley, 1999.

[16] R. Jain. *The Art of Computer Systems Performance Analysis*. John Wiley & Sons, Inc., 1991.

[17] K.-H. Kim and K. G. Shin. On accurate measurement of link quality in multi-hop wireless mesh networks. In *ACM MobiCom*, 2006.

[18] D. Kotz, C. Newport, and C. Elliott. The mistaken axioms of wireless-network research. Technical Report TR2003-467, Dartmouth College, Computer Science, July 2003.

[19] R. Krishnan, A. Raniwala, and T. cker Chiueh. Design of a channel characteristics-aware routing protocol. In *IEEE INFOCOM miniconference*, 2008.

[20] S. Lee, B. Bhattacharjee, and S. Banerjee. Efficient geographic routing in multihop wireless networks. In *ACM MobiHoc*, 2005.

[21] S. Lin, J. Zhang, G. Zhou, L. Gu, T. He, and J. A. Stankovic. ATPC: Adaptive transmission power control for wireless sensor networks. In *ACM SenSys*, 2006.

[22] H. Lundgren, E. Nordstrom, and C. Tschudin. Coping with communication gray zones in ieee 802.11b based ad hoc networks. In *ACM WoWMoM*, 2002.

[23] P. C. Ng and S. C. Liew. Throughput analysis of IEEE 802.11 multi-hop ad hoc networks. *IEEE/ACM Transactions on Networking*, 15(2), 2007.

[24] J. Polastre, J. Hill, and D. Culler. Versatile low power media access for wireless sensor networks. In *ACM SenSys*, 2004.

[25] L. Qiu, Y. Zhang, F. Wang, M. K. Han, and R. Mahajan. A general model of wireless interference. In *ACM MobiCom*, 2007.

[26] K. Ramachandran, I. Sheriff, E. Belding, and K. Almeroth. Routing stability in static wireless mesh networks. In *PAM*, 2007.

[27] A. Willig. A new class of packet- and bit-level models for wireless channels. In *IEEE PIMRC*, 2002.

[28] A. Woo, T. Tong, and D. Culler. Taming the underlying challenges of reliable multihop routing in sensor networks. In *ACM SENSYS*, 2003.

[29] H. Zhang, A. Arora, and P. Sinha. Learn on the fly: Data-driven link estimation and routing in sensor network backbones. In *IEEE INFOCOM*, 2006.

[30] H. Zhang, L. Sang, and A. Arora. Data-driven link estimation in sensor networks: An accuracy perspective. Technical report, Wayne State University (http://www.cs.wayne.edu/∼hzhang/group/TR/DNC-TR-08-02.pdf), 2008.

[31] H. Zhang, L. Sang, and A. Arora. On the convergence and stability of data-driven link estimation and routing in sensor networks. Technical report, Wayne State University (http://www.cs.wayne.edu/∼hzhang/group/TR/DNC-TR-08-03.pdf), 2008.

[32] J. Zhao and R. Govindan. Understanding packet delivery performance in dense wireless sensor networks. In *ACM SenSys*, 2003.

[33] M. Zuniga and B. Krishnamachari. An analysis of unreliability and asymmetry in low-power wireless links. *ACM Transactions on Sensor Networks*, 3(2), 2007.