This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications

A Unified Metric for Correlated Diversity in Wireless Networks

Shuai Wang, Anas Basalamah, Song Min Kim, Guang Tan,
Yunhuai Liu, and Tian He

***Abstract*—Recent pioneer work has shown that packet receptions on adjacent links are correlated, which contradicts the long held assumption that wireless links are statistically independent. Since wireless link correlation affects a wide range of protocol designs, it is essential to quantify the impact generically. Specifically, this work focuses on a unified transmission cost metric for diversity based routing schemes including opportunistic routing, network coding, and hybrid routing. Our study covers both unicast and broadcast. Compared with the legacy metrics, our metric provides a direct and accurate estimation of the transmission cost in the presence of link correlation. The new metric helps a wide range of routing algorithms determine when they can benefit from reception diversity, and how to maximize the benefit, at negligible costs. We evaluate the metric on one 802.11 testbed and three 802.15.4 testbeds running TelosB, MICAz, and GreenOrbs nodes. The experimental results show that our metric (i) reduces 92% and 94% of the estimation error of the transmission cost in unicast and broadcast, and (ii) outperforms the link independent metric in both unicast and broadcast applications.**

***Index Terms*—Link Correlation, Opportunistic Routing, Network coding, Hybrid Routing, IEEE 802.11, IEEE 802.15.4.**

I. Introduction

Given the broadcast nature of the wireless medium and the randomness of packet losses, packet reception diversity – a phenomenon in which packets from a node are received by multiple receiving nodes following different reception patterns, often exists. This phenomenon creates great opportunities for performance improvement. A radically new routing category, known as diversity based routing, including opportunistic routing [1], [2], network coding [3], [4], and hybrid routing [5]–[7], thus has been proposed to exploit the potential benefit brought by this diversity. Up to now, more than 1,000 research articles have been published under this category according to the ACM digital library.

Traditional diversity based routing designs implicitly or explicitly assume that the packet receptions of multiple receiving nodes are independent when they exploit the diversity benefit. Recent studies [8]–[10], however, show clear evidences that indicate otherwise: reception patterns at multiple receivers can indeed be correlated. The simplified link independence assumption will likely generate overly optimistic estimates of reception diversity. This misconception may lead to a suboptimal forwarder selection in routing designs, hence reducing network performance.

Consider an extreme situation where links possess perfect positive correlation, that is, the receivers have exactly the same reception patterns. In this case, reception diversity no longer exists. However, a traditional network coding or opportunistic routing algorithm will continue to make significant efforts (e.g., by packet encoding/decoding) to pursue the diversity benefit unproductively.

In this paper, instead of offering yet another new diversity aware routing design, we provide a unified metric to capture the diversity benefit among the packet receptions. This metric supports a wide range of diversity based routing designs and offers a more accurate estimate on transmission cost than provided by the link independence metric. It can therefore help a large number of routing designs fully exploit the diversity benefit and avoid unnecessary overhead. The implementation of our metric introduces nearly zero overhead.

Since packet reception diversity varies in degree across different network environments, we extensively check the accuracy of our metric on an 802.11 testbed with nine laptops located at our department building, and three 802.15.4 testbeds – a 30 TelosB node platform in a noisy open office, a 20 GreenOrb platform in an outdoor environment, and a 30 MICAz node platform in a dedicated lab. The experiment results confirm the accuracy and efficacy of our design.

Drawing on the accurate and general metric, we examine two applications, one using unicast and the other broadcast. In the unicast application, our metric helps two hybrid routing designs predict how much benefit it can obtain from reception diversity, which leads to an optimal solution in forwarder selection. In the broadcast application, we use our metric to predict the performance of a network coding protocol named Rateless Deluge [11]. It helps Rateless Deluge determine whether it can enjoy the benefit of reception diversity so that it can encode packets only when necessary. The experimental results show that our link correlation metric outperforms the link independent metric and κ factor [8] in both unicast and broadcast applications.

The remainder of the paper is structured as follows. Section II reviews related work. Section III presents the motivation. Section IV introduces the metric, followed by its implementation and application in Section V. Section VI

S. Wang and T. He are with the Department of Computer Science, University of Minnesota, Minneapolis, MN 55455, USA (e-mail: {shuaiw, tianhe}@cs.umn.edu).

A. Basalamah is with the Department of Computer Engineering, Umm Al Qura University, Saudi Arabia (e-mail: anasbasalamah@gmail.com).

S. Kim is with the Department of Computer Science, George Mason University, Fairfax, VA 22030, USA (e-mail: kim.songmin81@gmail.com).

G. Tan is with SIAT, Chinese Academy of Sciences, China (e-mail: guang.tan@siat.ac.cn).

Y. Liu is with Third Research Institute of the Ministry of Public Security, China (e-mail: yunhuai.liu@gmail.com).

1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications

evaluates the metric accuracy and its performance on unicast and broadcast. Finally, Section VII concludes the paper.

II. Related Work

This section reviews prior works on diversity based routing and link correlation.

A. Diversity based routing

• **Opportunistic Routing (OR):** Taking advantage of reception diversity, opportunistic routing protocols (e.g., ExOR [1]) defer the selection of the next hop for a packet until the protocol acquires knowledge about the set of forwarding nodes that have actually received that packet. In ExOR [1], the authors propose the primary opportunistic routing protocol, which uses single path ETX to select forwarding nodes, so that only forwarders with a lower ETX than the sender's are selected and incrementally prioritized. Using single path ETX as a metric for forwarder selection is an approximation since it does not capture all the potential opportunistic paths from the forwarder to the destination. To take the potential multipaths into account, anypath ETX or EAX [12], is used to estimate a path's cost more accurately.

• **Network Coding (NC):** By exploiting the nature of wireless broadcast and equipping the relaying nodes with the ability to perform coding operations, network coding allows the nodes to combine several packets for transmission together. Two coding strategies – inter-flow opportunistic coding [3] and intra-flow random linear network coding (RLNC) [4] are widely used. In opportunistic coding [3], the coding strategy finds coding opportunities in advance. Then, the relay node encodes multiple packets together and broadcasts in a single transmission. On the contrary, RLNC [4] directly makes a random linear combination of incoming packets. Once receivers collect enough linearly independent combinations, they are able to decode and retrieve the original packets.

• **Hybrid Routing (HR):** Both opportunistic routing and network coding exploit the reception diversity of wireless links. Recent studies [5]–[7] find that they are complementary and combining them into a hybrid paradigm offers an elegant solution to address their limitations. In MORE [6], the authors propose the first practical system to combine intra-flow RLNC with ExOR-style opportunistic routing. One potential problem of MORE is that it sends out the next coding segment until the source receives the ACK from the destination. Lin et al. thus propose SlideOR [7], which encodes packets in overlapping sliding windows such that coded packets from one window position may benefit another.

The previous diversity based routing schemes implicitly or explicitly assume that wireless links are independent which may lead to under-utilized diversity benefit or extra cost. This work proposes a more precise link correlation metric to help diversity based routing schemes fully exploit the diversity benefit. The previous link independent metric is a special case in our unified model.

B. Link correlation

Most work assumes that links are independent when they exploit the diversity gain of the wireless broadcast medium. Recently, Srinivasan and Zhu et al. reveal the link correlation phenomenon [8], [9]. Zhu et al. [9] propose a probabilistic flooding algorithm to reduce energy consumption in transmission by using implicit ACKs inferred from link correlation. Kim and Zhao et al. study the underlying causes of link correlation [13], [14]. Khreishah et al. [15] develop a fully distributed multicast algorithm using network coding based opportunistic routing under arbitrary correlations among the links. Kim et al. propose a novel metric which incorporates both spatial and temporal correlations for better wireless routing [16]. In application aspects, researchers propose efficient data dissemination protocols [17]–[19] and data collection protocols [20]–[22] by exploiting link correlation. The closest work to ours is the κ factor [8], where the authors explore a metric that captures the degree of packet reception correlation on different links. According to the correlation degree, the κ factor can indirectly infer the performance of network protocols.

Different from the κ factor which is an indicator of correlation degree, we propose a metric which directly captures the transmission cost, i.e., the expected number of transmissions, under inter-link correlation. The physical attribute of our metric eliminates the need to interpret indirect metrics (e.g., the κ factor), making it straightforward for network designers to guide the behavior of a protocol (e.g., selecting forwarder sets and deciding to apply coding technologies or not) and thus fully exploit the diversity benefit of the broadcast medium. In addition, the highlight of this work is its broad application. We provide the first extensive study to build a unified metric for diversity based routing schemes, including opportunistic routing, network coding, and hybrid routing, covering both unicast and broadcast. The derivation of the metric is further used to infer the throughput, delay, and energy consumption.

III. Motivation

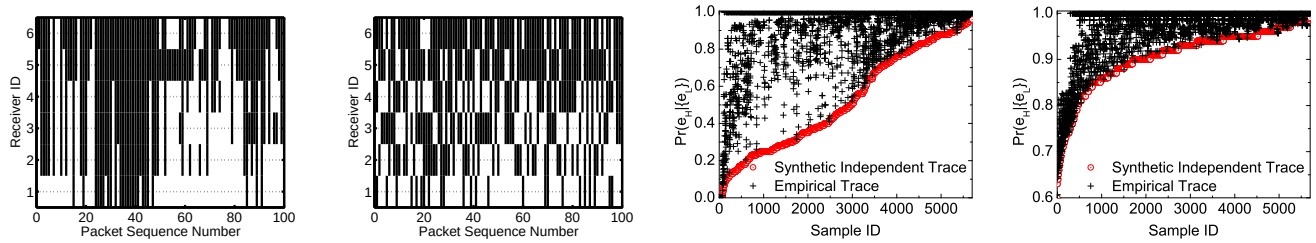
In this section, we first demonstrate the existence of link correlation. Then, we show the metric bias in existing work and illustrate the impact of link correlation on the effectiveness of diversity based routing.

A. Existence of Link Correlation

To verify the existence of link correlation, we conduct a small experiment on an 802.11 testbed located at the department building in the University of Minnesota. In the experiment, one AP and six laptops are deployed. The AP broadcasts 100 packets to the receivers under channel 6. Packets are identified by sequence number to record the missing packets. Figure 1(a) shows the packet receptions at the six receivers from empirical measurements. The lost packets are marked by black bands. Long vertical black bands indicate that packets are lost at multiple receivers. As a comparison, Figure 1(b) plots the independent synthetically generated traces with the same PRR where few multiple simultaneous losses are observed. The comparison of the sample traces indicates that

1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications



(a) Empirical Trace (b) Synthetic Independent Trace (c) 802.11: Statistic Results (d) 802.15.4: Statistic Results

Figure 1. Packet receptions at six receivers when a single transmitter broadcasts 100 packets; (a) empirical trace; (b) synthetic trace with the independent links under the same packet reception ratio. Packet losses are marked by black bands. Compared with synthetic independent trace, empirical trace has more long black bands, indicating more correlated losses; (c) Statistics of Receiving Probability on 802.11 and 802.15.4 testbeds: the $Pr(e_H|\{e_L\})$ value calculated with the empirical trace is usually larger than that with the synthetic independent trace, indicating the existence of positive correlation.

the packet receptions at multiple receivers in Figure 1(a) are correlated.

The statistic results on both 802.11 and 802.15.4 testbeds further confirm that wireless links are indeed correlated, as shown in Figure 1(c) and Figure 1(d). Here $Pr(e_H|\{e_L\})$ denotes the probability that the higher quality link e_H receives a packet given the condition that the lower quality link set $\{e_L\}$ has received the packet. From both figures, we can see that the $Pr(e_H|\{e_L\})$ value calculated with the empirical trace (denoted by the red circles) is usually larger than that with the link independent trace (denoted by the black crosses), which indicates the existence of positive correlation in this experiment environment.

Root cause of link correlation: Link correlation is caused by (i) cross-network interference under shared medium and (ii) correlated fading introduced by highly dynamic environments [10], [23]. With the increasing popularity of wireless technologies, the wireless spectrum becomes more and more crowded. For example, IEEE 802.11b/g/n and IEEE 802.15.4 use the 2.4 GHz ISM band, which leads to possible cross-network interference. Traffic from high-power communication (e.g., IEEE 802.11b/g/n) could introduce destructive noise in other low-power communication (e.g., IEEE 802.15.4), causing correlated packet loss at multiple receivers simultaneously. Non-network appliances such as microwave ovens can also introduce high-power interference in this crowded 2.4 GHz ISM band. On the other hand, wireless signals suffer shadow fading caused by the presence of obstacles in the propagation path of the radio waves, which leads to correlated packet receptions among receivers that are closely located.

B. Metric Bias

The existing diversity based routing schemes implicitly or explicitly assume that wireless links are independent when they exploit the intrinsic characteristics of wireless broadcast medium. For example, in MORE [6], the authors assume that "wireless receptions at different nodes are independent" when they design the transmission cost metric to decide how many packets a forwarder should send. We find that this assumption may cause serious estimation errors which will further lead to under-utilized diversity benefits or extra costs.

1) The Unicast Case: Consider the scenario in Figure 2(a) where u is the source and v_4 is the destination. The source u has a low delivery probability to its neighbor set $\{v_1, v_2, v_3\}$. When the node v_1 fails to receive a packet, candidate v_2 or v_3 probably has received the packet. Opportunistic routing significantly increases the packet receiving probability since it only requires that one of the three neighbors correctly receives it. In other words, the probability that all forwarders lose a packet becomes low because of the links' diversity.

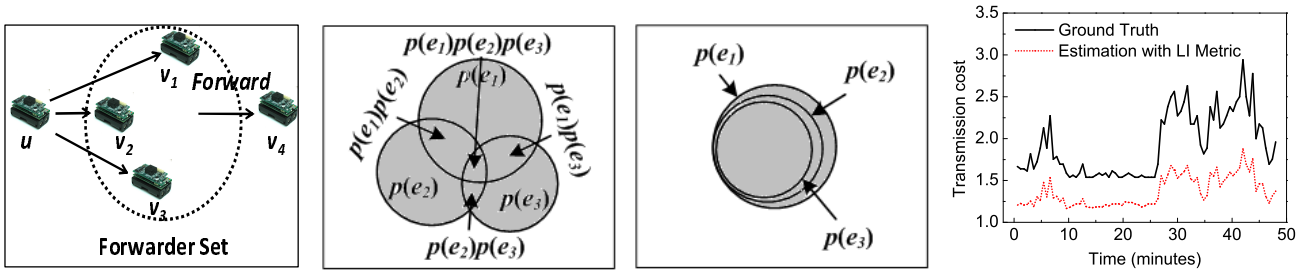
Figure 2(b) plots a Venn diagram representing the link independent case that the three receivers successfully receive transmissions from the sender. In the figure, we use e_i to denote link uv_i and $p(e_i) \in (0, 1]$ represents the probability that a source node can directly deliver a packet via link e_i. If we consider a case in which wireless links e_1, e_2, and e_3 are perfectly positive correlated, as shown in Figure 2(c), the probability that the forwarder set receives a transmission (i.e., the shadow area in Figure 2(c)) degenerates to the best link e_1's quality, which is much less than that in the link independent case (i.e., the shadow area in Figure 2(b)). In the perfectly positive correlated case in Figure 2(c), the received packets with the forwarder set is the same as that with the best link and opportunistic routing cannot provide benefits in performance but introduces additional control costs.

The statistic results from the testbed experiment in UMN confirms that there exists a gap between the estimated receptions with link independent metric and the ground truth. The experiment scenario is as shown in Figure 2(a). Figure 2(d) plots the expected unicast transmission cost with the link independence (LI) metric and the ground truth which is calculated using the number of sent packets divided by the number of received packets during a period of 30s. From Figure 2(d), we find that the estimates are 30% lower than the real values. This is because the link independence metric underestimates the expected transmission cost when links are positive correlated.

2) The Broadcast Case: In Figure 3(a), the source broadcasts four packets to three destinations. Assume that each destination receives the packets indicated in Figure 3(a), i.e., the first destination receives p_1 and p_2, the second p_2 and p_3, and the last, p_3 and p_4. Without coding, the sender has to retransmit all the four packets. In contrast, with network coding, it is sufficient to transmit two randomly coded packets. For example, the sender may send $p_1^{'} = p_1 + p_2 + p_3 + p_4$ and $p_2^{'} = p_1 + 2p_2 + 3p_3 + 4p_4$. Though they lose different packets, all three destinations can retrieve the four original

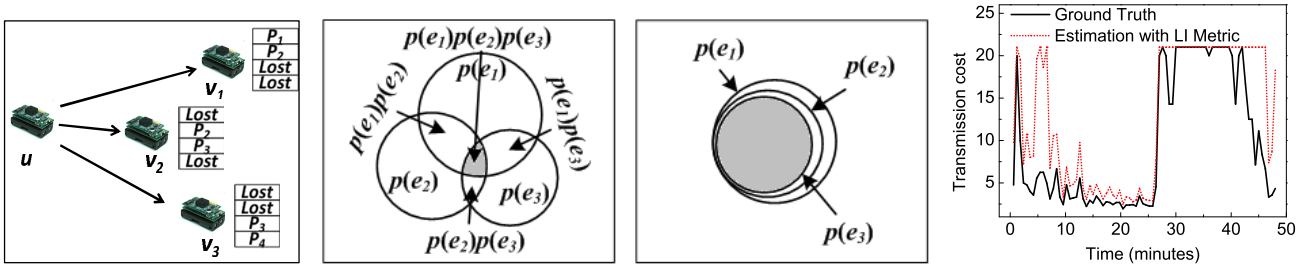
1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications



(a) Unicast example (b) Independent case (c) Positive correlated case (d) Estimate vs. truth

Figure 2. Metric bias for unicast: (a) a unicast scenario with three nodes in the forwarder set. (b) The Venn diagram of each receiver's packet reception in the link independent case; The dark area represents the event that the forwarder set receives a packet. (c) The Venn diagram in the positive correlated case. (d) Experimental result showing that the estimated number of received packets by the link independence (LI) metric is greater than the real one.



(a) Broadcast example (b) Independent case (c) Positive correlated case (d) Estimate vs. truth

Figure 3. Metric bias for broadcast: (a) a broadcast scenario with three receivers. (b) The shadow area represents the event that all the three nodes receive a packet in the link independence case. (c) The link correlation case. (d) Experimental result showing that the estimated number of received packets by the link independence (LI) metric in broadcast is always smaller than the real one.

packets using the two coded packets. Thus, in this example, network coding can reduce the needed retransmissions from four packets to two.

It has been shown that with network coding, we can attain the maximum possible information flow in a network. However, the maximum information flow provided by network coding is based on the assumption of link/path independence. When this is the case, the probability that all the nodes receive the same transmission is low. However, if the receptions were to be perfectly correlated, as shown in Figure 3(c), the total number of transmissions would be the same as in a traditional scheme since we only need to retransmit the packets lost by the worst-quality link.

Our observation from the testbed also reflects the inaccuracy of the link independent metric. We deploy one source node and three receivers as shown in Figure 3(a). The maximum number of retransmission for a packet is set to seven times the number of one-hop receivers for broadcast, i.e., 21 in this scenario. Figure 3(d) shows the gap between the ground truth and the expected broadcast transmission cost with the link independence (LI) metric. Broadcast needs few transmissions when links are correlated since the receivers have a high chance to lose the same packets. The estimation with the link independence metric increases the expected transmission cost and is much higher than the real one.

IV. The Link Correlation Metric

Indicated by our empirical study in Section III, the legacy assumption of link independence leads to biased errors in predicting the performance. Instead of addressing such errors with point solutions, the rest of this paper aims to design an accurate and general metric that can systematically rectify a wide range of diversity-based protocols.

A. Definitions

We now estimate the transmission cost for diversity based routing to reliably unicast or broadcast packets to a set of nodes. There are several ways to enable reliable delivery of packets, for example, ARQ and coded-feedback [24]. In ARQ, if a sender does not receive an ACK before timeout, it retransmits the packet until it receives an ACK. In coded-feedback, each node uses a coded feedback vector to represent the entire space of the received packet. When the upstream node receives the coded ACK, it extracts the original space and learns how many packets still need to be sent.

For a given node u, we use $N(u)$ to denote u's one-hop neighbor set $\{v_1, v_2, \ldots, v_M\}$, where M is u's out-degree. Let the link quality between u and the potential receiver v_j be $p(e_j)$, $j = 1, 2, \ldots, M$. Without loss of generality, we assume $p(e_1) \geq p(e_2) \geq \ldots \geq p(e_M)$.

Definition 1: (**Joint Packet Reception Probability**) Suppose u is transmitting a packet to a nonempty neighbor set $R(u) \subset N(u)$. We define u's Joint Packet Reception Probability (JPRP) as the probability that all the nodes in $R(u)$ successfully receive a packet, denoted as $p(R(u))$.

Definition 2: (**Joint Packet Loss Probability**) Given the receiver set $R(u)$, we define u's Joint Packet Loss Probability (JPLP) as the probability that all the nodes in $R(u)$ lose a packet, denoted as $p(\bar{R}(u))$.

JPRP and JPLP are introduced to describe the correlation among wireless links. Let $R_n(u)$ denote a forwarder set $R(u)$

1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications

with size n, where $n \leq M$. When there is only one node v_i in $R(u)$, i.e., $n = 1$, JPRP reduces to traditional link quality and $p(R_1(u)) = p(e_i)$.

We use these two definitions to build unified metrics, which adjust the estimation bias introduced by the widely used link independent metric. In the following, we derive the unified metric for correlated diversity in unicast and broadcast.

B. The Unicast Case

In unicast scenarios, a sender in opportunistic routing (OR) needs to keep transmitting a packet until at least one of the nodes in its forwarder set receives it. Network coding (NC) can help opportunistic routing eliminate layer abstraction and further exploit spatial reuse. In hybrid routing (HR), opportunistic routing and network coding are combined together, and the sender needs to reliably deliver k independent coded packets to its forwarder set.

We now theoretically analyze the expected number of transmissions to deliver k packets to an n-node forwarder set and demonstrate that diversity based routing schemes obtain *less* benefit when the links are more *highly* correlated. Let $p(e_1)$, $p(e_2)$, ..., $p(e_n)$ denote the link qualities of the n receivers respectively. The corresponding packet loss probabilities are $p(\bar{e_i}) = 1 - p(e_i)$. We define μ – the expected number of transmissions to deliver the k packets to the n-node forwarder set as the unicast cost.

Lemma 4.1: (**Unified Unicast Cost Metric** μ) The unicast cost of diversity based routing μ – the expected number of transmissions for the sender to successfully transmit k packets to a forwarder set $R(u)$ with size n, is given by

$$\mu = \frac{k}{1 - p(\bar{R}_n(u))}. \tag{1}$$

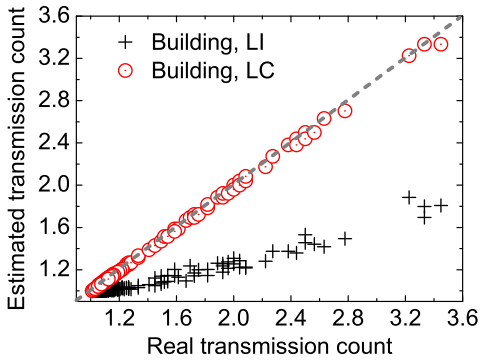
Lemma 4.1 is straightforward. The denominator in Eq.(1), i.e., $1 - p(\bar{R}_n(u))$, represents the probability that at least one node in the forwarder set receives one packet. The formula $\mu = \frac{k}{1-p(\bar{R}_n(u))}$ thus represents the expected number of transmissions for the sender to successfully deliver k packets to at least one node in the forwarder set.

The computation complexity of μ is $O(n)$, where n is the size of the forwarder set. When $k = 1$ and $n = 1$, i.e., the sender u transmits one packet to the forwarder set which contains a single node, μ equals one over the quality of the link from u to the forwarder node, which reduces to ETX [25]. When $k = 1$ and $n > 1$, μ indicates the cost for opportunistic routing (OR), i.e., the expected number of transmissions for sender u reliably deliver a packet to a set of nodes. Network coding (NC) needs to guarantee that the receiver gets k independent packets for the decoding purpose. The expected number of transmission for NC is thus $\frac{k}{p(e)}$, which is the case of μ when $k > 1$ and $n = 1$. When $k > 1$ and $n > 1$, it describes hybrid routing (HR) which is a combination of opportunistic routing and network coding. To summarize, our unified unicast cost metric μ embraces the following cases:

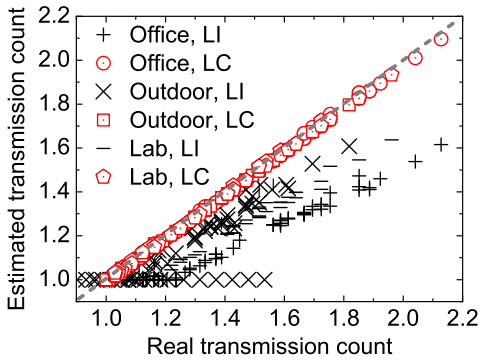
$$\mu = \begin{cases} 1/p(e), & k = 1, n = 1 \ (ETX) \\ \frac{1}{1-p(\bar{R}_n(u))}, & k = 1, n > 1 \ (OR) \\ \frac{k}{p(e)}, & k > 1, n = 1 \ (NC) \\ \frac{k}{1-p(\bar{R}_n(u))}, & k > 1, n > 1 \ (HR) \end{cases} \tag{2}$$

• **Special Case:** Furthermore, the legacy link independent metric, which is widely used in existing studies, e.g., MORE [6], is a special case of our unified metric. When we assume wireless links are independent, we have $p(\bar{R}_n(u)) = p(\bar{e_1})p(\bar{e_2}) \dots p(\bar{e_n})$, and the unicast cost μ reduces to

$$\mu = \frac{k}{1 - \prod_i^n p(\bar{e_i})}. \tag{3}$$



(a) 802.11 testbed



(b) 802.15.4 testbed

Figure 4. μ vs. Real transmission count

To examine the performance of μ, we conduct experiments on one 802.11 testbed and three 802.15.4 testbeds. The implementation details and experiment settings are described in Section V and VI. Figure 4 shows the comparison of the unicast cost metric μ, the Link Independent unicast metric "LI", and the real unicast transmission count on the 802.11 and 802.15.4 testbeds.

In both figure 4(a) and 4(b), we find that the transmission counts estimated by the link independence metric for unicast are lower than the real ones. This is because in a static environment, most links in 802.11 and 802.15.4 networks are positive correlated [8], [10] because of the interference from high power emission sources. The positive link correlation means that the nodes in a forwarder set lose similar packets. Under the assumption of link independence, when one forwarder

1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications

fails to receive a packet, others may have a better chance of receiving it. This optimistic view thus leads to an overestimate of diversity benefit and transmission efficiency.

In summary, on the 802.11 testbed (Figure 4(a)), the average error rate of the μ metric is 1%, while it is 12.9% for the link independent metric. Similarly, in Figure 4(b), we find that the error rate of μ on the 802.15.4 testbeds is 0.9%, while it is 9.3% for the link independent metric. The link correlation metric μ reduces the link independent metric 92% of the estimation error of the transmission cost in unicast.

C. The Broadcast Case

In the broadcast scenario, the problem changes to how to reliably send packets to a set of nodes, and make sure that each of these nodes receives k packets. Let ε be the number of transmissions needed by sender u to reliably broadcast k packets to a set of nodes – $\{v_1, v_2, \ldots, v_n\}$. In the following, we classify the existing broadcast routing schemes into three categories: (i) traditional broadcast schemes without considering reception diversity, (ii) broadcast schemes with inter-flow opportunistic coding, which finds coding opportunity in advance and delivers specified coded packets to all the potential receivers, and (iii) broadcast schemes with intra-flow random linear network coding (RLNC), which requires all the receivers receive k independent packets.

For the first two categories, the source node needs to guarantee that all the nodes in $R(u)$ receive the *same* k packets. In the third category, the sender only needs to make sure that all the nodes receive k *independent* packets. In the following, we first establish metrics to calculate the expected number of transmissions – $E(\varepsilon)$ for the first two categories, i.e., traditional broadcast and broadcast schemes with inter-flow opportunistic coding.

Lemma 4.2: The broadcast cost $E(\varepsilon)$ for successfully delivering k packets to the n receivers can be calculated as

$$E(\varepsilon) = \sum_{\substack{R_1(u)=\{v_i\}\\1\le i\le n}} \frac{k}{1-p(\bar{R}_1(u))} - \sum_{\substack{R_2(u)=\{v_i,v_j\}\\1\le i<j\le n}} \frac{k}{1-p(\bar{R}_2(u))} + \sum_{\substack{R_3(u)=\{v_i,v_j,v_l\}\\1\le i<j<l\le n}} \frac{k}{1-p(\bar{R}_3(u))} + \ldots + \frac{(-1)^{n-1}k}{1-p(\bar{R}_n(u))}. \quad (4)$$

The metric $E(\varepsilon)$ represents the expected number of transmissions for the sender to successfully deliver k packets to all the receivers. The derivation of ε is not straightforward and the proof detail is in Appendix A. To compute Eq.(4), we need to compute $\binom{n}{1}+\binom{n}{2}+\ldots+\binom{n}{n} = 2^n - 1$ polynomial terms. We thus propose an approximate algorithm to estimate $E(\varepsilon)$, denoted as ε.

Lemma 4.3: (**Unified Broadcast Cost Metric** ε) Assume nodes with higher link quality receive a broadcast packet earlier than those with lower link quality, then u's expected transmission count of k packets with n potential receivers is approximated by

$$\varepsilon = \sum_{i=1}^{n} \frac{k}{p(e_i)} - \sum_{i=2}^{n} \frac{k}{p(e_i)} \cdot \frac{p(R_i(u))}{p(R_{i-1}(u))}. \quad (5)$$

The proof of Lemma 4.3 is in Appendix B. The computational complexity of ε is $O(n^2)$.

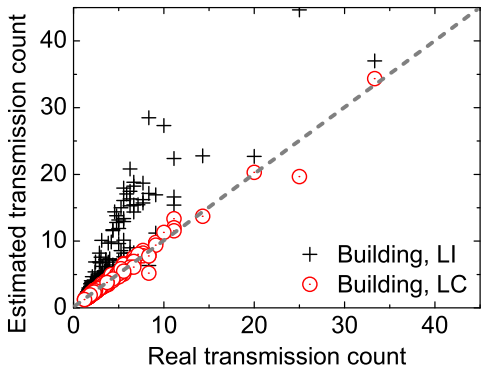
• **Special Case:** Eq.(5) includes the special case of link independence. When links are independent, we have $p(R_i(u)) = p(e_i)p(R_{i-1}(u))$, and ε reduces to

$$\varepsilon = \sum_{i=1}^{n} \frac{k}{p(e_i)} - kn + k. \quad (6)$$

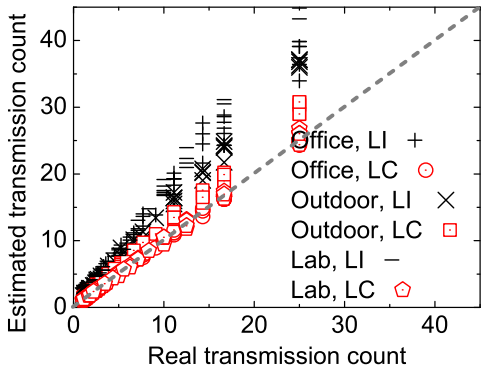
For broadcast schemes with intra-flow random linear network coding, when the sender broadcasts coded packets, it needs to make sure all that the receivers $\{v_1, v_2, \ldots, v_n\}$ get k independent packets, which finally goes to ensure that the node with the worst-quality link in the receiver set receives k independent packets. Without loss of generality, we assume $p(e_1) \ge p(e_2) \ge \ldots \ge p(e_n)$.

Lemma 4.4: Let the broadcast cost with random linear network coding be ε_R, then $\varepsilon_R = \frac{k}{p(e_n)}$, and ε_R is the infimum of ε. Here $p(e_n)$ is the packet reception ratio of the worst link.

The proof of Lemma 4.4 is in Appendix C. From this lemma, the expected number of transmissions with the random linear network coding is always less than (or equals to) that without using the coding technology. On the other hand, the value of ε_R can be quite close to ε, and the performance of RLNC may even worse than that of traditional routing protocols given the coding/decoding overhead.



(a) 802.11 testbed



(b) 802.15.4 testbed

Figure 5. ε vs. Real transmission count

Up to now, we have deduced the link correlation metric for the broadcast cost. We check the accuracy of ε on both 802.11 and 802.15.4 testbeds, as shown in Figure 5. From the figures, we find that the link independence metric overestimates the transmission count. In broadcast, we need to

1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications

make sure all the nodes receive the packets. When links are positive correlated, the source node needs fewer transmissions because the receivers miss packets in similar patterns. The link independence metric mistakenly assumes that the receivers lose different packets, which means more retransmissions.

Comparing the results from the three 802.15.4 testbeds, we find different gaps between the two metrics. The link independence metric (i.e., "LI") generates large errors on the lab and office testbeds because the cross network interference from high-power wireless networks (e.g., Wi-Fi) could introduce destructive noise to low-power sensor networks, causing correlated packet losses. The error is small for the outdoor testbed, where the links are less correlated.

To summary, the average error rate of ε is 4.3% in broadcast, while it is 63.9% for "LI" on the 802.11 testbed (Figure 5(a)). Similarly, in Figure 5(b), we find that the error rate of ε on the 802.15.4 testbed is 10.8% in broadcast, while it is 64.2% for "LI". ε reduces the link independent metric 94% of the estimation error of the transmission cost in broadcast.

V. Metric Implementation and Overhead

In the previous section, we introduce the unified metric for diversity based routing. In this section, we describe its implementation detail, overhead under link dynamics as well as the application in unicast and broadcast. The accuracy of the metric will be analyzed in the next section.

A. How to Obtain the Metric Online

To calculate the expected number of transmissions for either unicast or broadcast, we need to obtain each receiver's link quality $p(e)$ and joint packet loss (or reception) probability $p(\bar{R}_i(u))$ (or $p(R_i(u))$). Suppose each receiver maintains a packet reception report (e.g., [1100]) recording the reception status of a fixed number (e.g., 4) of most recent packets. With the reception report, the link quality is given simply by the number of 1s in the reception report divided by the length of reception report. The calculation of $p(R_i(u))$ or $p(\bar{R}_i(u))$ deserves a little more explanation. Assume a reception report of length W, we have

$$p(R_i(u)) = \frac{1}{W}\sum_{j=1}^{W} B_{v_1}(j)\&B_{v_2}(j)\&\ldots\&B_{v_i}(j), \qquad (7)$$

where $B_{v_i}(j)$ is a bit representing the receiver v_i's reception status of the jth packet. $B_{v_i}(j) = 1$ represents node v_i receives the packet, otherwise $B_{v_i}(j) = 0$.

• **Example of Calculating $p(R_i(u))$:** Here we use an example to show how to calculate $p(R_i(u))$. In Figure 6, node u has three receivers. We calculate u's $p(R_i(u))$ for $\{v_1, v_2\}$. Suppose the reception report of node v_1 is [1100], which indicates that v_1 receives the 1st and 2nd packets and misses the 3rd and 4th packets. When node u receives the reception bitmaps, it uses Eq.(7) to calculate $p(R_i(u))$, i.e., $p(R_2(u)) = (1\&0 + 1\&1 + 0\&0 + 0\&1)/4 = 25\%$.

B. Overhead under Link Dynamics

The calculation of the link correlation metrics – μ and ε introduce little additional cost. The main communication overhead of the link correlation metrics come from the exchange of the reception bitmaps which are needed by the calculation of link quality and $p(\bar{R}_i(u))$ (or $p(R_i(u))$). To calculate link quality, we need to use the bit information in rows in the reception bitmap, as shown in Figure 6. The calculation of $p(\bar{R}_i(u))$ needs to use the bit information in columns in Figure 6. The required information for link quality and $p(\bar{R}_i(u))$ (or $p(R_i(u))$) is exactly the same. In other words, the link correlation metric introduces no extra costs when the link quality measurement is required.

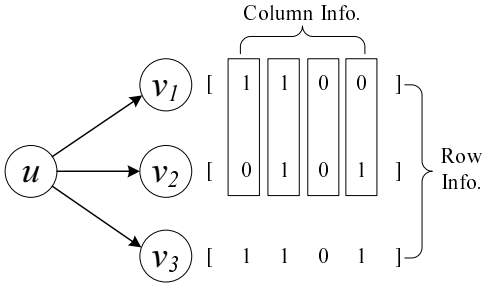


Figure 6. Example of Calculating u's $p(R_i(u))$ for $\{v_1, v_2\}$. With the same bitmap input, the link correlation metric not only utilizes the row information (which is used to calculate the link quality) but also utilizes the column information to exploit the correlation among links.

To collect reception bitmaps, we adopt the free piggyback mechanism with normal traffic data, which has been already applied by the existing protocols [26], [27] to measure link quality or to improve the robustness of the routing structure. In addition, the one-hop receivers can also send packet reception bitmaps directly to their upstream node to calculate the values of link quality and $p(\bar{R}_i(u))$ (or $p(R_i(u))$). Our measurement shows that the overhead of exchanging reception bitmaps for the link quality information only occupies less than 1% of the total energy cost since binary reception bitmaps are small and are much less frequently exchanged.

C. Metric Embedding

In the previous sections, we propose the unified link correlation metric which captures the transmission cost of diversity based routing protocols. However, a metric is of little use if it cannot be used in a range of protocols for performance improvements. This section demonstrates how our generic metric can be embedded into existing unicast and broadcast protocols. As shown in Figure 7, we can simply replace the link independent metric (e.g., the unicast metric $1/p(e_i)$) with our link correlation metric, a process that is transparent to upper layer designs (e.g., router selection) as well as derivative metrics (e.g., throughput). Such a transparent embedding process enables us to easily support a wide range of existing protocols that use link independent metrics and improve their performance broadly. In the following, we use two examples to show how to embed our metric with existing diversity based routing algorithms.

• **Unicast Application Example:** We first consider the unicast application using a hybrid routing scheme, e.g., MORE [6]. In the design of MORE, the source calculates and assigns a transmission cost to each node in the forwarder set, using

1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications

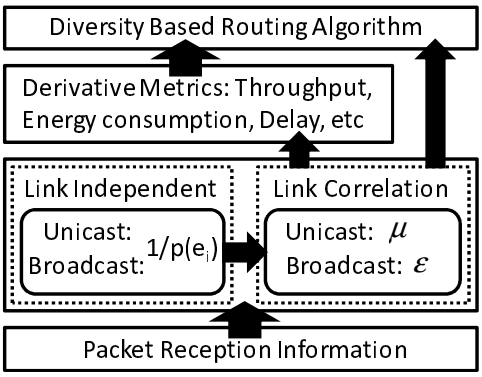


Figure 7. Link correlation metric embedding.

ETX [1], computed from each node's packet reception ratio under the link independence assumption. As discussed before, the forwarder set selected by the link independent metric may not only introduce the overhead of the forwarders scheduling, but also obtain little diversity gain if high correlated routing pathes are constructed.

We now show how to use our metric to maximize the reception diversity benefit by optimizing the forwarder set selection. Instead of using ETX, we use our metric μ in Eq.(2) to estimate the transmission cost. Then we adopt the path selection algorithms used in existing hybrid routing designs to choose the forwarder set to minimize the expected number of any-path transmissions needed for the reliable delivery of one packet from a source to a destination. We define the expected number of any-path transmissions needed for the reliable delivery of one packet from a source u to a destination d, with awareness of link correlation, as $E(u,d)$. We calculate $E(u,d)$ starting from destination d backward to the source u:

$$E(u,d) = \mu + \beta, \tag{8}$$

where μ is the expected number of transmissions for successfully transmitting k coded packets from the source node u to the next-hop forwarder set, and β captures the cost of forwarding the k packets from the forwarder set to the destination, which is given by

$$\beta = \sum_{i=1}^{n} \mu \cdot p(e_i) \cdot E(v_i,d), \tag{9}$$

where μ is used because of the implicit condition that the forwarder set has received the k packets. The term $\mu \cdot p(e_i)$ represents the number of packets delivered by the forwarder v_i and $E(v_i,d)$ is the expected transmission count for reliably delivering a packet from v_i to d. With $E(u,d)$, we prioritize the forwarder candidates, so that when a forwarder with lower $E(u,d)$ costs to the destination opportunistically receives the packet, it has a better chance of delivering it. By introducing μ, we are able to take advantage of reception diversity for the hybrid routing algorithms.

• **Broadcast Application Example:** In addition to transparent embedding (i.e., metric replacement), we can also utilize our metric to obtain maximum performance gain through explicit metric integration. In this section, we demonstrate a broadcast application using two dissemination protocols – Deluge [28] which does not use network coding, and Rateless Deluge [11] which uses intra-flow random linear network coding (RLNC) to reduce transmissions. Specifically, we describe how to integrate our metric with Rateless Deluge to encode packets efficiently, and thus improving the throughput. The metric will activate Rateless Deluge when it brings benefit and deactivate it when the coding cost exceeds the benefit. In the following, we use our metric to avoid unnecessary coding operations.

Let τ be the sum of the delivery delay of one packet and the delay between transmissions. With our link correlation metric, the time T_D for Deluge to successfully disseminate a batch of k packets is $T_D = \varepsilon \cdot \tau$,where ε is given in Eq.(5). In Rateless Deluge, we need to take into account the time spent by coding operations, and the delay T_{RD} is given by $T_{RD} = \varepsilon_R \cdot \tau + T_c$, where $\varepsilon_R = \frac{k}{p(e_n)}$. T_c is the time spent by coding operations for k packets in one page. It is a constant decided by the number of coded packets and the hardware settings. To avoid unnecessary coding/decoding overhead, in our design, Rateless Deluge only encodes packets when $T_{RD} < T_D$.

VI. Evaluation: Metric Accuracy and Application

In wireless networks, packet reception patterns vary significantly across network environments, as they are affected by environmental noise and external interference. In the following, we first evaluate the accuracy of the link correlation metric under different power levels and channels in subsection VI-A. The performance of the metric in the unicast and broadcast applications will be shown in subsection VI-B. We evaluate our link correlation metric on both 802.11 and 80.15.4 testbeds under different environments and settings. The basic information of the testbeds are summarized in Figure 8 and Table I.

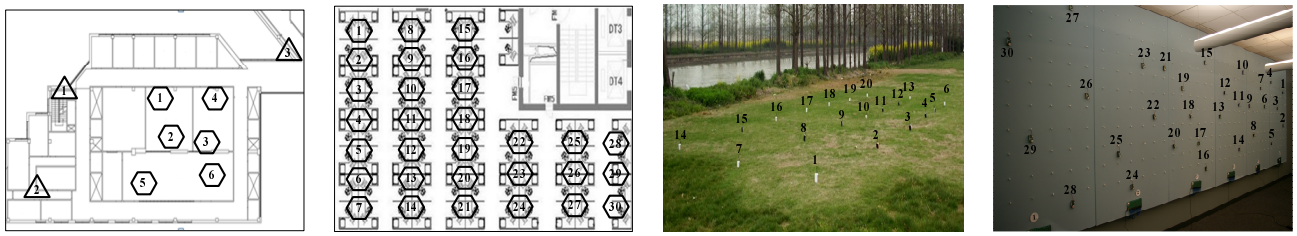
• **802.11 testbeds:** 802.11g experiments run on the fourth floor of the computer science department building in University of Minnesota. The map of the testbed is shown in Figure 8(a), where we use triangles to represent the APs. In the experiment, we use laptops with an Intel Core2 Duo T5470 processor, 2GB RAM, and an 802.11 wireless card with the Realtek to serve as APs. In the figure, laptops are labeled as hexagons. From the figure, we can see that three APs are placed at the corner of the floor while six laptops are distributed in labs and meeting rooms, separated by concrete walls.

The data traffic is generated using the `Lorcon2` packet injection library [29]. The data rate is 1 Mbps. During the experiment, three APs send 10^5 packets in turn and the receivers keep on listening and recording all the received packets in the environment. We then use Wireshark to extract our packets with the unique SSIDs. We thus get the packet reception information for all the receivers with a specified AP. With the reception bitmaps we are able to evaluate the two metrics. In the link correlation metric, we use Eq.(1) and Eq.(5) to estimate the transmission counts of unicast and broadcast. In unicast, the default forwarder set size is three, and in broadcast the default size of receiver set is six. In the link independence metric, Eq.(3) and Eq.(6) are used to compute the transmission count of unicast and broadcast.

• **802.15.4 testbeds:** The first 802.15.4 testbed consists of 30 TelosB nodes, labeled as hexagons in Figure 8(b). These nodes

1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications



(a) 802.11: Department building (b) 802.15.4: Open office (c) 802.15.4: Outdoor (d) 802.15.4: Research Lab

Figure 8. Testbeds experiment: (a) each triangle is an AP, and the hexagons represent the receivers (laptops); (b) we use hexagons to represent the sensors. The number is the Node ID; (c) the outdoor scenario; (d) the research lab scenario.

Platform	Location	Environment	Nodes	APs	Physical Size	Degree	Channel	Power
802.11g	UMN	Department Building	6	54+3	$73m \times 30m$	6	Ch3, Ch6	15dBm, 20dBm
TelosB	SIAT	Open Office	30	8	$18m \times 13m$	$6 \sim 21$	Ch16, Ch26	-25dBm
GreenOrbs	TRIMPS	Outdoor	20	0	$15m \times 5m$	$4 \sim 13$	Ch16	-25, -19.2dBm
MICAz	UMN	Research Lab	30	54	$8m \times 2.5m$	$7 \sim 23$	Ch16	-25dBm

Table I
TESTBED SETTINGS AND TOPOLOGY PROPERTIES.

are deployed in an $18m \times 13m$ open office environment following a grid pattern. Figure 8(c) plots the testbed in the outdoor environment, in which 20 GreenOrbs nodes are deployed on the grass-covered curb along a river. The third testbed is in a dedicated lab environment, as shown in Figure 8(d), in which a total of 30 MICAz nodes are randomly deployed on an $8m \times 2.5m$ wall. In all three 802.15.4 testbeds, the default transmission power is set to -25dBm and the default channel is 16. After deployment all nodes are synchronized and broadcast 8000 packets in turn. The delay between transmissions is 0.3s. All the received packets are recorded in the nodes' flash memory. When all the nodes finish broadcasting packets, they send their packet reception information to a sink node which is connected to a PC.

In both 802.11 and 802.15.4 networks, the maximum number of retransmission is set to be seven. In the following, we use "LI" to represent the Link Independent metric. Our Link Correlation metric (i.e., μ and ε) is labeled "LC". Note that we do not compare the accuracy of our link correlation metric with the κ factor since it is an indicator of correlation degree ranging from $[-1, 1]$, while our metric measures the expected number of transmissions. Instead, we compare the "LI", κ, and "LC" metrics in the application subsection since all of them infer the performance of network protocols.

A. Metric Accuracy

Link correlation is reflected by packet losses. On the testbeds, we can not control the link quality unless artificially introducing packet losses. In the following, we study the accuracy of the link correlation ("LC") metric under different power levels and channels, which have a great impact on packet losses and can be controlled freely.

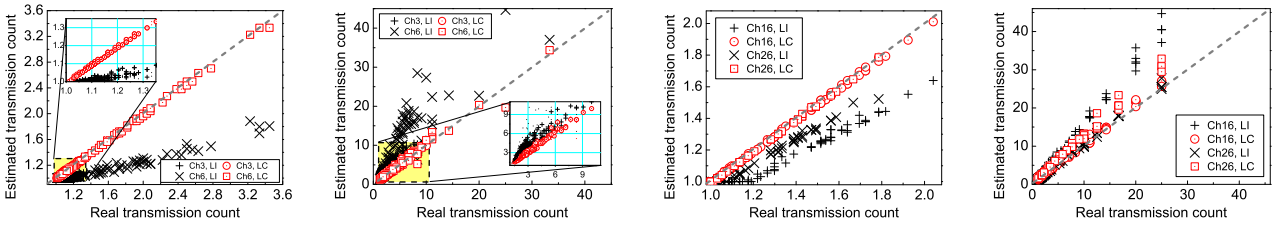
1) Varied Channels: **•802.11 testbeds:** In this experiment, we study the impact of channels on the two metrics. Figure 9(a) and 9(b) show the metric accuracy under channels 3 and 6. In the unicast case, the real transmission count range under channel 3 is between 1.01 and 1.35 (see the enlarge part in Figure 9(a)), which is lower than channel 6's $1.01 \sim 3.45$. Similar results are also observed in the broadcast cases. This is because the interference on channel 6 is much severer than channel 3 given that about one third of the 54 APs in our department use channel 6. From Figure 9(a) and 9(b), the estimation error of the link correlated metric in both unicast and broadcast scenarios is much lower than that of the link independent one. This is because the interference of Wi-Fi signals makes the 802.15.4 transmissions using channel 16 highly correlated, a phenomenon consistent with the observation in [8].

•802.15.4 testbeds: In this experiment, we use the data from the office testbed (Figure 8(b)), which includes a total of 8 access points. Note that channel 16 overlaps with a cohabiting access point's 802.11 channel and that channel 26 is free of Wi-Fi interference. We ran the experiments during normal office time. Figure 9(c) and 9(d) show the metric accuracy under two channels – channels 16 and 26. From Figure 9(c) and 9(d), we can see that the link correlation metric remains accurate for both channels. Compared with channel 26, the source needs more transmissions under channel 16. This is because the overlapped channel causes more packet losses. Besides, we see that the estimation error of the link independence metric significantly increases under channel 16. This is because the interference of Wi-Fi signals makes the 802.15.4 transmissions using channel 16 highly correlated.

2) Varied Powers: **•802.11 testbeds:** We change the power level from 15dBm to 20dBm to generate different packet losses. In Figure 10(a), the unicast transmission count under 15dBm ranges from 1.09 to 2.94, while that under 20dbm is from 1.01 to 3.45. The average transmission count estimation error rates of the LC metric under unicast are 1% for both power levels, while that of the LI metric are 13.8% (15dBm) and 12.9% (20dBm). Figure 10(b) plots the broadcast transmission count under the delivery ratio of 76.3% (15dBm) and 99.6% (20dBm). The average estimation error rates of the LC

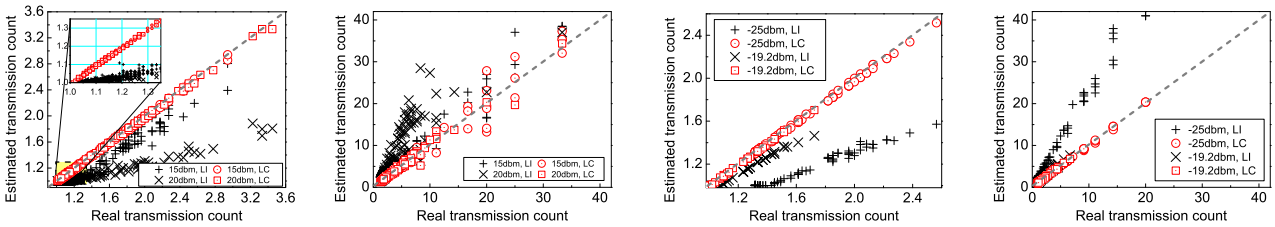
1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications



(a) 802.11: Unicast (b) 802.11: Broadcast (c) 802.15.4: Unicast (d) 802.15.4: Broadcast

Figure 9. Metric accuracy with varied channels.



(a) 802.11: Unicast (b) 802.11: Broadcast (c) 802.15.4: Unicast (d) 802.15.4: Broadcast

Figure 10. Metric accuracy with varied power levels.

metric for broadcast are 3.8% (15dBm) and 4.3% (20dBm), while they are 55.5% and 63.9% for the LI metric.

•**802.15.4 testbeds:** We conduct this experiment on the outdoor testbed (Figure 8(c)), in which we can control the transmission range among nodes freely. The power level for transmission is set from -25dBm to -19.2dBm. Figure 10(c) and 10(d) show the estimated transmission count along with the ground truth for unicast and broadcast separately. We find that most of the "LC" dots appear on the diagonal line under both power levels, indicating the accuracy of our metric. The accuracy of the link independence metric improves when the power level increases. It works well under a high power level but is still worse than the link correlation metric because of the existence of correlated losses.

B. Application Performance

In this section, we report the experiment results about the performance improvement from the link correlation metric, compared with the link independent metric and the κ factor [8], in both unicast and broadcast applications.

1) Unicast Application: In the unicast experiment, 20 nodes are deployed on a $8m \times 2.5m$ wall. The power is set to be -25dBm to form a two-hop network. We run MORE with the "LI", κ [8], and "LC" metrics separately. Figure 11 shows the packet transmission counts of MORE with the three metrics. On average, the "LC" metric saves MORE (with the "LI" metric) about 23% of transmissions while the κ factor saves MORE 15% of transmissions. Our "LC" metric has a better performance than the κ factor since it is a direct indicator of the transmission cost and fully exploit the diversity benefit.

Table II lists the estimated transmission counts, i.e., μ, and the ground truth from the source to the forwarder sets in MORE. The "LI" metric selects forwarder set 1, because it uses ETX and the average link quality of forwarder set 1 is better than that of forwarder set 2. However, we find

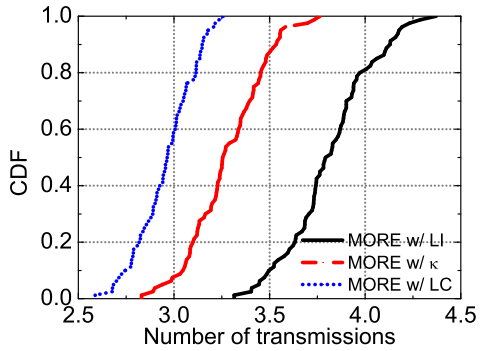


Figure 11. Unicast application.

	Forwarder Set 1 (PRR=0.51)	
	Estimated	Ground Truth
LI	1.68	2.37
LC	2.34	2.37
	Forwarder Set 2 (PRR=0.49)	
	Estimated	Ground Truth
LI	1.73	1.83
LC	1.81	1.83

Table II
LC VS. LI IN FORWARDER SET SELECTION

that the "LI" metric's estimates for forwarder set 1 are much lower than the real ones, since it ignores correlated losses. It makes the source node make wrong decisions on the selection of forwarder nodes. With our metric, it accurately estimates the transmission count and helps the source select the most appropriate forwarder set.

2) Broadcast Application: This section validates the usefulness of our link metric to the performance of Deluge and Rateless Deluge. In the dissemination experiments, the source sends an image to 12 one-hop receivers under channel 16 and

1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications

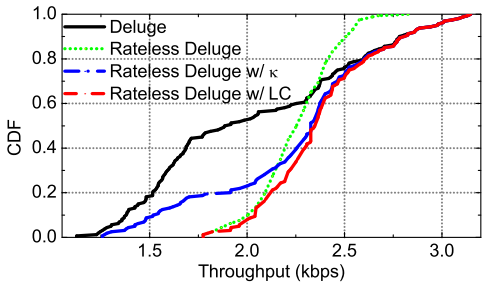


Figure 12. Broadcast application.

26 at -25dbm. The disseminated image comprises 9 pages with 20 packets each. The experiment ends once every node has collected the entire image in its memory. The delay between transmissions is 7ms. The average packet delivery delay is about 50ms. We thus have $\tau = 57ms$. The encoding/decoding delay T_c for one page with 20 packets is 1.56s on average.

Figure 12 plots the throughput of Deluge, Rateless Deluge, and Rateless Deluge with the κ [8] and "LC" metrics. From the figure, we find that Rateless Deluge does not always outperforms Deluge. This is because when link correlation is high, the transmission cost with and without network coding is almost the same, and Rateless Deluge may have a worse performance than Deluge due to the coding/decoding delay. Our design, i.e., Rateless Deluge with "LC" metric has better performance than both Deluge and Rateless Deluge since the "LC" metric accurately estimates the transmission benefit with and without network coding, and thus execute coding operation appropriately. The κ factor works not as good as the "LC" metric since the threshold setting with κ factor (i.e., specified correlation degrees) can not always provide the appropriate guidance in coding switching, especially in link dynamic scenarios.

VII. Conclusion

This paper shows that packet receptions over different links are not independent. We thus present a unified link correlation metric for diversity based routing, which covers opportunistic routing, network coding and hybrid routing. Compared with the traditional link independent metric, our metric provides a better estimate on transmission cost. It guides the diversity based routing to fully exploit the wireless diversity with negligible costs. We show the versatility of the metric on both unicast and broadcast applications. Experiment results from one 802.11 and three 802.15.4 testbeds confirm the accuracy and efficiency of the link correlation metric.

References

[1] S. Biswas and R. Morris. Exor: Opportunistic multi-hop routing for wireless networks. In *ACM SIGCOMM*, 2005.
[2] Olaf Landsiedel, Euhanna Ghadimi, Simon Duquennoy, and Mikael Johansson. Low power, low delay: Opportunistic routing meets duty cycling. In *IPSN*, 2012.
[3] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft. Xors in the air: Practical wireless network coding. In *SIGCOMM*, 2006.
[4] T. Ho, M. Medard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 52:4413–4430, 2006.
[5] R. Bruno and M. Nurchis. Survey on diversity-based routing in wireless mesh networks: Challenges and solutions. *Computer Communications*, 2010.
[6] S. Chachulski, M. Jennings, S. Katti, and D. Katabi. Trading structure for randomness in wireless opportunistic routing. In *SIGCOMM*, 2007.
[7] Y. Lin, B. Liang, and B. Li. Slideor: Online opportunistic network coding in wireless mesh networks. In *IEEE INFOCOM*, 2010.
[8] K. Srinivasan, M. Jain, J. I. Choi, T. Azim, E. S. Kim, P. Levis, and B. Krishnamachari. The κ-factor: Inferring protocol performance using inter-link reception correlation. In *ACM MOBICOM*, 2010.
[9] T. Zhu, Z. Zhong, T. He, and Z.-L. Zhang. Exploring link correlation for efficient flooding in wireless sensor networks. In *NSDI*, 2010.
[10] S. Wang, S. Kim, Y. Liu, G. Tan, and T. He. Corlayer: A transparent link correlation layer for energy efficient broadcast. In *MOBICOM*, 2013.
[11] A. Hagedorn, D. Starobinski, and A. Trachtenberg. Rateless deluge: Over-the-air programming of wireless sensor networks using random linear codes. In *ACM/IEEE IPSN*, 2008.
[12] Z. Zhong and S. Nelakuditi. On the efficacy of opportunistic routing. In *IEEE SECON*, 2007.
[13] Song Min Kim, Shuai Wang, and Tian He. Exploiting causes and effects of wireless link correlation for better performance. In *IEEE INFOCOM*, 2015.
[14] Z. Zhao, W. Dong, G. Guan, J. Bu, T. Gu, and C. Chen. Modeling link correlation in low-power wireless networks. In *IEEE INFOCOM*, 2015.
[15] A. Khreishah, I. Khalil, and J. Wu. Distributed network coding-based opportunistic routing for multicast. In *ACM MobiHoc*, 2012.
[16] Song Min Kim, Shuai Wang, and Tian He. Incorporating spatiotemporal correlation for better wireless networking. In *ACM SENSYS*, 2015.
[17] Manjunath Doddavenkatappa, Mun Choon Chan, and Ben Leong. Splash: Fast data dissemination with constructive interference in wireless sensor networks. In *NSDI*, 2013.
[18] W. Du, J. C. Liando, H. Zhang, and M. Li. When pipelines meet fountain: Fast data dissemination in wireless sensor networks. In *ACM SENSYS*, 2015.
[19] Z. Zhao, W. Dong, J. Bu, Y. Gu, and C. Chen. Link correlation aware data dissemination in wireless sensor networks. *IEEE Transactions on Industrial Electronics*, 2015.
[20] W. Du, Z. Li, J. C. Liando, and M. Li. From rateless to distanceless: Enabling sparse sensor network deployment in large areas. In *ACM SENSYS*, 2014.
[21] W. Du, Z. Li, J. C. Liando, and M. Li. From rateless to distanceless: Enabling sparse sensor network deployment in large areas. *IEEE/ACM Transactions on Networking*, 2015.
[22] Mobashir Mohammad, Xiangfa Guo, and Mun Choon Chan. Oppcast: Exploiting spatial and channel diversity for robust data collection in urban environments. In *ACM/IEEE IPSN*, 2016.
[23] Song Min Kim, Shuai Wang, and Tian He. Exploiting causes and effects of wireless link correlation for better performance. In *IEEE INFOCOM*, 2015.
[24] D. Koutsonikolas, C. Wang, and Y. Hu. Efficient network-coding-based opportunistic routing through cumulative coded acknowledgments. *IEEE/ACM Transactions on Networking*, 2011.
[25] D. Couto, D. Aguayo, J. Bicket, and R. Morris. A high-throughput path metric for multi-hop wireless routing. In *ACM MOBICOM*, 2003.
[26] Nouha Baccour, Anis Koubaa, Luca Mottola, Marco Zuniga, Habib Youssef, Carlo Boano, and Mario Alves. Radio link quality estimation in wireless sensor networks: a survey. *ACM TOSN*, 2012.
[27] A. Kamthe, M. A. Carreira-Perpinan, and A. E. Cerpa. M&m: Multi-level markov model for wireless link simulations. In *ACM SENSYS*, 2009.
[28] J. Hui and D. Culler. The dynamic behavior of a data dissemination protocol for network programming at scale. In *ACM SENSYS*, 2004.
[29] Lorcon packet injection library. https://code.google.com/p/lorcon/.

Appendix A
Proof of the Lemma

Proof of Lemma 4.2: We consider the scenario where one source node delivers k original or coded packets to n receivers. For an arbitrary positive integer threshold t, we denote $\frac{\varepsilon}{k}$ as the number of transmissions the source node u needs to deliver *one* packet to the

1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications

receivers. The probability that $\frac{\varepsilon}{k}$ exceeds the threshold t satisfies the following equation:

$$Pr(\frac{\varepsilon}{k} > t) = \sum_{\substack{R_1(u)=\{v_i\} \\ 1\le i\le n}} p(\bar{R}_1(u))^t - \sum_{\substack{R_2(u)=\{v_i,v_j\} \\ 1\le i<j\le n}} p(\bar{R}_2(u))^t + \sum_{\substack{R_3(u)=\{v_i,v_j,v_l\} \\ 1\le i<j<l\le n}} p(\bar{R}_3(u))^t + \ldots + (-1)^{n-1} p(\bar{R}_n(u))^t. \tag{10}$$

In addition,

$$Pr(\frac{\varepsilon}{k} = t) = Pr(\frac{\varepsilon}{k} > t-1) - Pr(\frac{\varepsilon}{k} > t).$$

Then, the expected number of broadcast transmission $E[\varepsilon]$ can be calculated as

$$\begin{aligned} E[\varepsilon] =& k \cdot E[\frac{\varepsilon}{k}] = k \cdot \sum_{t=1}^{+\infty} t \cdot Pr(\frac{\varepsilon}{k} = t) \\ =& k \cdot \sum_{t=1}^{+\infty} t \cdot (Pr(\frac{\varepsilon}{k} > t-1) - Pr(\frac{\varepsilon}{k} > t)) \\ =& k \cdot \sum_{t=0}^{+\infty} Pr(\frac{\varepsilon}{k} > t) \end{aligned}$$

Substitute $Pr(\frac{\varepsilon}{k} > t)$ with Eq.(10), we have

$$E[\varepsilon] = \sum_{\substack{R_1(u)=\{v_i\} \\ 1\le i\le n}} \frac{k}{1-p(\bar{R}_1(u))} - \sum_{\substack{R_2(u)=\{v_i,v_j\} \\ 1\le i<j\le n}} \frac{k}{1-p(\bar{R}_2(u))} + \sum_{\substack{R_3(u)=\{v_i,v_j,v_l\} \\ 1\le i<j<l\le n}} \frac{k}{1-p(\bar{R}_3(u))} + \ldots + \frac{(-1)^{n-1}k}{1-p(\bar{R}_n(u))}.$$

In the paper, we use ε as the abbreviation of $E[\varepsilon]$. Q.E.D.

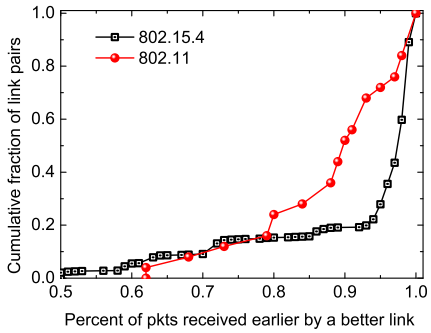


Figure 13. Statistics of receiving probability. The node with better link quality receives about 92% (802.15.4 testbed) and 89% (802.11 testbed) of the packets earlier or at the same time than the node with worse link quality.

APPENDIX B
PROOF OF THE LEMMA

Proof of Lemma 4.3: We provide the approximation of ε by assuming that the nodes with higher link quality receive a broadcast packet before (or at the same time) those with lower link quality. To confirm this hypothesis, we conduct experiments on both 802.15.4 and 802.11 testbeds. Figure 13 shows the statistic results. From the figure, we can find that the node with better link quality receives more than 90% of the packets earlier (or at the same time) than the node with worse link quality.

Based on the hypothesis, we propose the approximation algorithm to calculate ε. First we estimate the number of transmissions for the source node u to reliably send one packet to the node v_i with better link quality. Then we consider the transmissions of delivering one packet to the node v_j with worse link quality under the situation that v_j fails to receive the packet when u sends it to v_i. Let $p(R_i(u))$ be the probability that all the i nodes in $R_i(u)$ successfully receive one packet. We have

$$\begin{aligned} \hat{\varepsilon} =& \frac{k}{p(e_1)} + \frac{k\Pr(\bar{e_2}|e_1)}{p(e_2)} + \ldots + \frac{k\Pr(\bar{e_n}|\bigcap_{i=1}^{n-1} e_i)}{p(e_n)} \\ =& \frac{k}{p(e_1)} + \frac{k(p(R_1(u)) - p(R_2(u)))}{p(e_1)\cdot p(e_2)} + \ldots \\ & + \frac{k(p(R_{n-1}(u)) - p(R_n(u)))}{p(R_{n-1}(u))p(e_n)} \\ =& \sum_{i=1}^{n} \frac{k}{p(e_i)} - \sum_{i=2}^{n} \frac{k}{p(e_i)} \cdot \frac{p(R_i(u))}{p(R_{i-1}(u))} \end{aligned}$$

Q.E.D.

APPENDIX C
PROOF OF THE LEMMA

Proof of Lemma 4.4: The proof consists of two parts. First, we show that $\varepsilon_R = \frac{k}{p(e_n)}$, which is straightforward. When the worst-quality link receives k independent packets, the remaining links must have already received k packets. We now prove that the infimum of $\hat{\varepsilon}$ is ε_R. Since $p(e_i) \ge p(e_{i+1})$, $p(R_i(u)) \ge p(R_{i+1}(u))$ and $p(e_i) \ge p(R_i(u))$, we have

$$\frac{p(e_{i+1})}{p(e_i)} \ge \frac{p(R_{i+1}(u))}{p(R_i(u))}. \tag{11}$$

Therefore,

$$\begin{aligned} \hat{\varepsilon} =& \sum_{i=1}^{n} \frac{k}{p(e_i)} - \sum_{i=2}^{n} \frac{k}{p(e_i)} \cdot \frac{p(R_i(u))}{p(R_{i-1}(u))} \\ =& \frac{k}{p(e_n)} + \sum_{i=1}^{n-1} (\frac{k}{p(e_i)} - \frac{k}{p(e_{i+1})} \cdot \frac{p(R_{i+1}(u))}{p(R_i(u))}) \\ \ge& \frac{k}{p(e_n)}. \end{aligned} \tag{12}$$

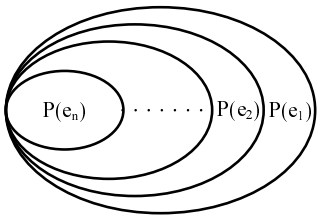


Figure 14. The case of $p(e_i) = p(R_i(u))$.

When $p(e_i) = p(R_i(u))$, as shown in Figure 14, we have

$$\hat{\varepsilon} = \frac{k}{p(e_n)} = \varepsilon_R. \tag{13}$$

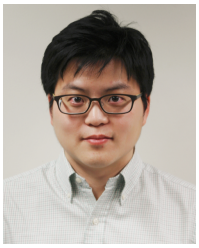
From the above proof, we find that the transmission cost without network coding will be quite close to that with network coding when the links are high correlated. When the links are perfect positive correlated (i.e., the case shown in Figure 14), the transmission cost with and without network coding is exactly the same. Given the encoding/decoding delay, network coding may even have a worse performance than the traditional routing protocols without network coding. Q.E.D.

1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

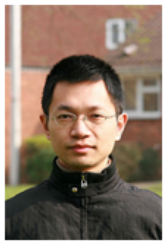
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2581821, IEEE Transactions on Wireless Communications

Shuai Wang received the B.S. and M.S. degrees from Huazhong University of Science and Technology, China. He is currently working toward the PhD degree with the Department of Computer Science and Engineering at the University of Minnesota, Twin Cities. His research interests include mobile cyber physical systems, bigdata analytics for intelligent transportation systems, wireless networks and sensors.

Anas Basalamah is an Assistant Professor at the Computer Engineering Department of Umm Al Qura University. He is currently leading the Data Collection Research Thrust of KACST GIS Technology Innovation Center at Umm Al-Qura University. He did his M.S. and Ph.D. Degrees at Waseda University, Tokyo in 2006, 2009 respectively. He worked as a Post Doctoral Researcher at the University of Tokyo and the University of Minnesota in 2010, 2011 respectively. His areas of interest include Embedded Networked Sensing, Real-world applications of Networked Sensing, Ubiquitous Computing, Participatory and Urban Sensing.



Song Min Kim is currently an assistant professor in the Department of Computer Science at George Mason University. He received his Ph.D. from the Department of Computer Science and Engineering at the University of Minnesota in 2016. He received his M.E. and B.E. degrees from the Department of Electrical and Computer Engineering at Korea University in 2009 and 2007, respectively. His research interests include wireless and low-power embedded networks, mobile computing, Internet of Things, and Cyber-physical Systems.



Guang Tan is currently a Professor at Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences, China, where he works on mobile computing systems and networks. He received his Ph.D. degree in computer science from the University of Warwick, U.K., in 2007. From 2007 to 2010 he was a postdoctoral researcher at INRIA Rennes, France. His research is sponsored by National Science Foundation of China and Chinese Academy of Sciences. He is a member of IEEE, ACM, and CCF.

Yunhuai Liu received his PhD degree in Computer Science and Engineering from Hong Kong University of Science and Technology in 2008. From 2008 to 2010 he worked in Hong Kong University of Science and Technology as a research assistant professor. In the year 2010, he joined Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences as an associate professor. And in the year 2011, he joined the Research Center of Internet of Things, Third Research Institute of Ministry of Public Security located in Shanghai, China. His research interests include wireless sensor networks, cognitive radio networks, and extreme-scale datacenter and data networks. His research papers have been published in many prestigious conferences and journals such as ACM Mobicom, IEEE INFOCOM, IEEE ICDCS, IEEE ICPADS, and IEEE TPDS, IEEE TMC. The paper titled "opportunity-based topology control in wireless sensor networks" obtained the only Best Paper Award in IEEE ICDCS 2008 (1 out of 638). He is an IEEE and ACM member.

Tian He is currently an associate professor in the Department of Computer Science and Engineering at the University of Minnesota-Twin City. Dr. He is the author and co-author of over 200 papers in premier network journals and conferences with over 17,000 citations (H-Index 54). Dr. He is the recipient of the NSF CAREER Award, George W. Taylor Distinguished Research Award and McKnight Land-Grant Professorship and five best paper awards in networking. Dr. He served a few program chair positions in international conferences and on many program committees, and also served as an editorial board member for six international journals. His research includes wireless sensor networks, cyber-physical systems, intelligent transportation systems, real-time embedded systems and distributed systems, supported by National Science Foundation, IBM, Microsoft and other agencies.

1536-1276 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.