The κ Factor: Inferring Protocol Performance Using Inter-link Reception Correlation

Kannan Srinivasan†, Mayank Jain†, Jung Il Choi†, Tahir Azim†, Edward S Kim*,
Philip Levis†, and Bhaskar Krishnamachari*
{srikank, mayjain, jungilchoi, tazim}@stanford.edu, e.skim@usc.edu,
pal@cs.stanford.edu, bkrishna@usc.edu
†Stanford University *University of Southern California
Stanford, CA-94305 Los Angeles, CA-90089

ABSTRACT

This paper explores metrics that capture to what degree packet reception on different links is correlated. It shows that a widely used metric is a bad measure and presents normalized cross correlation index as a new metric. We call this new metric κ(kappa). Measuring κ sheds light on when and why opportunistic routing and coding protocols perform well (or badly) in a network. κ shows a direct correlation to the performance of opportunistic routing protocols like ExOR. We also show that protocols using network coding are not always suitable for all environments. Comparing Deluge and Rateless Deluge, Deluge's network coding counterpart, we find that κ can predict which of the two is best suited for a given environment.

Measuring κ in several 802.15.4 and 802.11 testbeds, we find that it varies significantly across network topologies and link layers. This suggests that the results of experimental comparisons can differ depending on the network used, and that reporting κ is important for experimental repeatability.

Categories and Subject Descriptors: C.2.1 [Network Architecture and Design]:Wireless communications

General Terms: Measurement, Design, Experimentation, Performance.

Keywords: 802.15.4, Wireless measurement study, Low power wireless networks, Wireless protocol design.

1. INTRODUCTION

One of the principal opportunities that wireless provides is the ability to send a single packet to multiple receivers. Data dissemination protocols such as Deluge [9] and broadcast protocols such as RBP [20] can broadcast data to every neighbor at once. Routing protocols can snoop, or use opportunistic receptions [4], in order to forward packets even when delivery to the primary intended recipient fails.

In these schemes, correlations in packet reception on different links can greatly impact protocol performance. For example, if all links receive the same packets (i.e their reception are completely correlated,) then opportunistic routing protocols will not work better than the shortest path protocols, as there is no spatial diversity to exploit. At the other extreme, if the receivers are negatively correlated – a reception at one implies a failure at the other – then opportunistic routing can be of great benefit.

In prior work, researchers have often concluded that packet reception on different links are largely independent [19, 16]. Miu et al. use the cross conditional measure, $P(A = 0|B = 0) - P(A = 0)$, as a metric for inter-link correlation between links A and B [16]. We refer to this metric as χ. This metric was subsequently used by Reis et al. [19] and Laufer et al. [13] to conclude that most of the link pairs have independent packet reception. In Section 3, we show that χ is a bad measure of correlation: its value depends highly on the packet reception ratios (PRRs) of the two links. Section 5 further shows that χ is not indicative of how well opportunistic routing protocols work.

Protocol designs typically assume that reception on different links is independent [6, 8, 24, 13, 5, 11]. Network simulators [3, 14] and work on network analysis also typically ignore these correlations as it is commonly believed that these correlations are too hard to capture.

From a set of wireless measurements over the past 4 years, we have found that reception at different receivers are not always independent. This observation is not new: one of the earliest sensor network deployment studies, Great Duck Island, observed correlated reception [21]. However, the degree of correlation varies greatly across network setups and link layers.

This paper presents a new metric called κ that captures this degree of correlation. κ is a 3-tuple quantity that measures packet reception correlation on two links that have a common transmitter. A κ of 1 means that

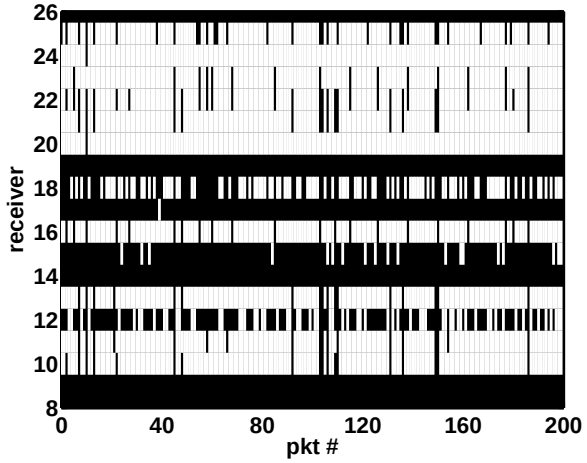


Figure 1: Packet reception at multiple receivers on Mirage channel 17. Packet loss is marked by black overlines. Several packets are lost at multiple nodes.

the reception at the two receivers are highly correlated, zero means they are independent, and -1 means that the losses on one are highly correlated with receptions on the other. Section 3 discusses the κ metric in detail.

Sections 4 and 5 systematically explore the usefulness of κ, and show that κ is a good indicator of how well opportunistic routing protocols like ExOR [4] perform. This paper also shows how κ can be used to understand when a network coding protocol such as Rateless Deluge [8] is beneficial over protocols that don't do network coding. This can allow researchers to make informed decisions while selecting protocols for testbeds and deployments. Section 6 describes this comparison in detail.

Measuring κ on IEEE 802.15.4 [23] and 802.11 (WiFi) [1] networks, we find that 802.15.4 networks see more correlated receiver pairs than WiFi; in an 802.15.4 network 70% of the receiver pairs have a κ higher than 0.8, while in all of the 802.11 networks that we measured, less than 20% of the WiFi pairs fall in that range. We find that external noise from 802.11 makes 802.15.4 receivers to be highly correlated.

Overall, this paper makes four research contributions. First, it shows that a widely used inter-link reception correlation metric is a bad measure. Second, it presents a new metric called κ. Third, it shows that κ explains how well an opportunistic protocol like ExOR performs in a network. Fourth, it shows how κ can be used to select a protocol that works best in a network. Specifically, κ helps to choose between Deluge and Rateless Deluge for a wireless network.

This paper shows that reception on different links can, unlike the general belief, be correlated and that measuring this correlation can help us understand when and why protocols perform the way they do. We believe that this insight is useful in designing efficient future protocols.

2. TESTBEDS

In an attempt to observe and understand various degrees of correlation present in wireless networks, this paper uses measurements and experiments from both 802.15.4 [23] and 802.11b [1] networks. 802.15.4 is an IEEE PHY-MAC low power, low data rate network standard with a 16 channel spectrum that overlaps the spectrum of 802.11b. It provides a data rate of 250 kbps and maximum transmit power of 0dBm, which are much lower than 802.11b's 11 Mbps capability and 23dBm maximum transmit power.

We run 802.15.4 experiments using TinyOS running on the Intel Mirage testbed [10], which consists of 100 Micaz [22] nodes placed along the ceiling. An Ethernet back-channel provides communication to all the nodes.

802.11b experiments run on the SWAN testbed at Stanford University [2] using data rates of 1, 2, 5.5 and 11 Mbps. This testbed consists of 40 nodes located along the hallways of 2 adjacent department buildings; Computer Science and Electrical Engineering. The Electrical Engineering building houses 15 nodes specific to this network spread out across 4 floors. 25 nodes are located in the Computer Science building and are evenly distributed across 6 floors. These nodes use the Madwifi driver and Click modular router [12] on a Linux kernel. Ethernet cables provide an IP based back-channel. In both the buildings several other 802.11b/g networks exist that are not under our control.

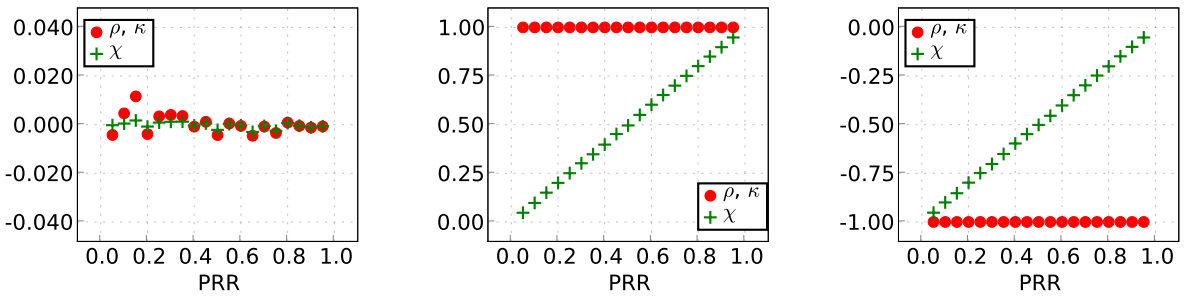
Apart from these two testbeds, we use several adhoc testbeds to explore answers to specific questions that arise. We also use publicly available Roofnet [15] datasets.

3. THE κ METRIC

This section shows that reception at multiple links is correlated. It investigates two existing metrics to measure this inter-link correlation; χ, a conditional probability metric that led earlier work to conclude that reception on links are independent of each other [19, 16] and ρ, the standard cross-correlation index used in statistics [7]. This section shows that χ is highly biased and so, defines a normalized form of ρ as the inter-link reception correlation metric, κ. It presents κ measurements for several 802.15.4 and 802.11 testbeds, showing that κ varies across networks and link layers.

3.1 Evidence of Existence

A small experiment shows the existence of inter-link reception correlations. A single node on the Mirage



(a) Independent (b) Positively Correlated (c) Negatively Correlated

Figure 2: ρ, κ and χ for synthetic data traces with varying PRRs. The PRRs for the two links are equal for the independent and positive correlation case and $PRR_2 = 1 - PRR_1$ for the negative correlation case. For these cases, $\kappa = \rho$. The cross-conditional metric, χ does not always correctly identify receiver pairs as correlated or negatively correlated, while ρ and κ do.

testbed transmits 200 broadcast packets on channel 16. All the nodes that hear the packets report to a centralized server.

Figure 1 shows packet reception at different receivers. Packet losses are marked by black overlines. Long vertical overlines indicate that packets are lost at multiple receivers. Visually, the packet losses on different links are correlated. While Figure 1 shows that reception correlations exist, quantifying such correlations is desirable.

3.2 Exploring Correlation Metric: χ

Previous work have used cross-conditional probability, χ, as the inter-link correlation metric and concluded that reception on links are generally independent [19, 16]. χ is a 3-tuple quantity, defined on a transmitter, t, and two receivers that can hear packets from t, as:

$$\chi_{t,x,y} = P^{(t)}_{x/y}(0/0) - P^{(t)}_{x}(0), \tag{1}$$

where, $P^{(t)}_{x/y}(0/0)$ is the probability, when t transmits, that a packet failed on link t→x given that it failed on link t→y and $P^{(t)}_{x}(0)$ is the probability that a packet failed on link t→x. If the failures on the two links, t→x and t→y are independent then χ is 0.

A simple example shows that χ has a PRR bias problem. We generate a synthetic trace of packet receptions on two links with varying PRRs for three cases. In the first case, the reception on the link pair are independent. In the second case, the reception on the link pair have perfect positive correlation; when a packet succeeds on one link it also succeeds on the other and when it fails on one it fails on the other as well. In the third case, the reception on the link pair have perfect negative correlation; when a packet succeeds on one link it fails on the other and vice versa.

Figure 2(a) shows calculated values of χ over a range of PRRs for two independent links. The metric properly reflects the independence of the uncorrelated links with near zero values. However, in Figure 2(b), for perfectly correlated links, χ fails to identify the correlation, but rather reflects the PRR of the receiver in question, causing confusion as to whether the links may be only partially correlated or even independent. As PRR approaches 0, it is clear why the two links can be misinterpreted as independent links when using χ metric. Figure 2(c) also demonstrates this shortcoming for negatively correlated links. This shows that the cross conditional probability metric fails to identify correlation.

3.3 Exploring Correlation Metric: ρ

As an alternative, we consider a popular quantity in statistics that measures correlation between two quantities: the cross-correlation index, ρ. ρ, in our definition, is a 3-tuple of one transmitter, t and two random variables, x and y, corresponding to reception at two receivers. This paper assumes that x and y are random variables representing 1 for a successful reception and 0 for a failure (they are therefore Bernoulli distributions). The rest of this paper uses "x" and "y" to refer to both the receivers and their corresponding random variables. ρ is defined as:

$$\rho_{t,x,y} = \begin{cases} \frac{E[x.y] - E[x].E[y]}{\sigma_x.\sigma_y}, & \sigma_x.\sigma_y \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $\sigma_x = \sqrt{P_x.(1 - P_x)}$ is the standard deviation of x, E[x.y] is the empirical mean of the product of x and y, E[x] is the mean of x, and E[y] is the mean of y. E[x.y] is the probability that both x and y receive the same packet, $P^{(t)}_{x,y}(1,1)$. The empirical means, E[x] and E[y], are the packet reception ratios of the links t→x

and t→y, respectively.

ρ compares the probability that both links actually receive a given packet to the probability that both would receive a given packet if their receptions were independent. If the difference between these two values is zero then the reception at x and y are independent. If the difference is positive: their reception are positively correlated. If the difference is negative: their reception are negatively correlated.

The graphs in Figure 2 show that, unlike χ, ρ properly identifies the correlation and decorrelation between links over all PRRs. While ρ provides a good indication of receiver correlation, ρ's scale changes based on the difference in PRR's of the links being used.

Lemma 1. *The range of ρ of [-1,1] is not tight. The true range of ρ depends on the packet reception ratio pairs, namely P_x and P_y. The maximum ρ is given by:*

$$\rho^{max} = \frac{min(P_x, P_y) - P_x.P_y}{\sigma_x.\sigma_y} \tag{3}$$

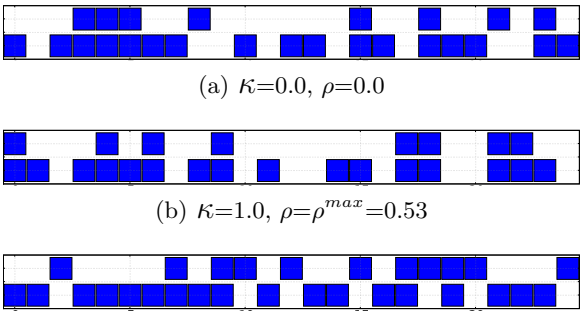
and the minimum ρ is given by:

$$\rho^{min} = \begin{cases} \frac{-P_x.P_y}{\sigma_x.\sigma_y}, & P_x + P_y \leq 1 \\ \frac{P_x+P_y-1-P_x.P_y}{\sigma_x.\sigma_y}, & otherwise \end{cases} \tag{4}$$

where $\sigma_x = \sqrt{P_x.(1-P_x)}$

Lemma 1 says that ρ cannot take arbitrary value in [-1,1] for arbitrary packet reception ratio pairs. For example, if two links have different packet reception ratios, it is impossible for the two receivers of the links to have a perfect correlation of $\rho = 1$. Similarly, if the packet reception ratios of the two links do not sum up to 1, then it is impossible for every packet to be received at exactly one of the two receivers. Only when both the packet reception ratios are 0.5, the true range of ρ is [-1,1]. The proof of Lemma 1 is in Appendix.

From a systems perspective, the limited range of ρ can limit its usefulness. For example, consider a pair of links with PRR 0.9 and 0.1. The maximum ρ for this case is $1/9 \approx 0.11$. Getting $\rho = 0.11$ means that for every packet lost at the higher PRR link, the corresponding packet is also lost at the lower PRR link, and for every packet received at the lower PRR link, the corresponding packet is received at the higher PRR link. For any protocol that exploits spatial diversity of links, after adding the higher PRR link, there is no gain by using the lower PRR link. Such a link pair should be classified as a highly correlated pair. For this reason, we further normalize ρ such that ρ^{max} is mapped to a 1 and ρ^{min} is mapped to -1. We define κ as this normalized ρ.



(a) κ=0.0, ρ=0.0

(b) κ=1.0, ρ=ρ^{max}=0.53

(c) κ=-1.0, ρ=ρ^{min}=-0.49

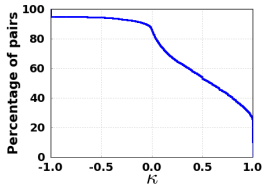
Figure 3: Packet reception sample for synthetically generated traces with different κ's. Blue squares correspond to successful packet reception. The PRR's for the two links are 0.4 and 0.7 in each case. (a) independent receiver pairs, (b) perfectly correlated reception and (c) negatively correlated reception.

$$\kappa_{t,x,y} = \begin{cases} \frac{\rho_{t,x,y}}{\rho^{max}_{t,x,y}}, & \text{if } \rho_{t,x,y} > 0 \\ \frac{-\rho_{t,x,y}}{\rho^{min}_{t,x,y}}, & \text{if } \rho_{t,x,y} < 0 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

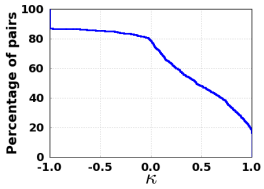
3.4 Understanding κ

A $\kappa_{t,x,y}$ of zero means that reception at x and y are independent for packets from t. κ can be a maximum of 1 corresponding to perfect reception correlation. If κ = 1 and if $PRR_x > PRR_y$, then if x receives a packet then necessarily y receives the same packet and if y loses a packet, then x also loses the same packet. κ can be a minimum of -1 corresponding to perfect negative correlation. If κ = -1 and if $PRR_x + PRR_y < 1$ then x and y never receive the same packet, and if $PRR_x + PRR_y > 1$, then x and y never lose the same packet.

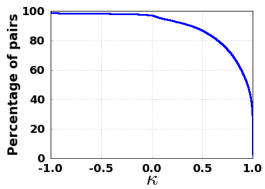
Figure 3 shows how reception from synthetically generated traces appear at receivers x and y, and the corresponding κ values. Figure 3(a) shows reception on links when the traces for the two links are generated independently. κ identifies these traces to be independent. Figure 3(b) shows traces for two links in which, when the lower PRR link receives a packet the other link also receives that packets and when the higher PRR link loses a packet the other link also loses the same packet. This link pair has a perfect correlation and κ is 1.0. Figure 3(c) shows traces for two links in which, when the higher PRR link loses a packet it is necessarily received on the other. This pair has a perfect negative correlation and κ is -1.0.



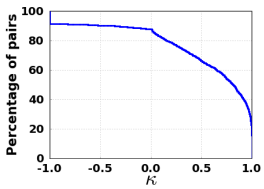
(a) Channel 26, Power 0dBm



(b) Channel 26, Power -25dBm



(c) Channel 16, Power 0dBm



(d) Channel 16, Power -25dBm

Figure 4: Complimentary CDF of κ for link pairs on Mirage. On channel 26, about 35% of the link pairs have a $\kappa > 0.8$. On channel 16, at the highest power level nearly 60% of the pairs have a $\kappa > 0.8$. At the lowest power level of -25dBm on channel 16, this percentage is close to 55%. Channel 16 shows more correlated link pairs than channel 26.

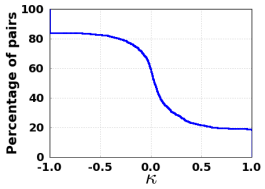
3.5 κ on Testbeds

Every node takes turn to send a burst of 50,000 broadcast packets. Every receiver that receives a packet, sends the successfully received sequence numbers over the wired back-channel to a server. The server runs this experiment for different parameters such as channel, data rate (for 802.11) and transmission power level.

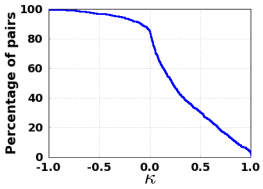
Figures 4 and 5 show the complimentary cumulative distribution function (CCDF) of κ for all communicating node pairs for all transmitters. These plots include, for two different testbeds, all possible pairs of links that heard at least one packet from a transmitter.

Figure 4 shows CCDFs for the Mirage testbed on channels 26 and 16 for two power levels: 0dBm and -25dBm. Channel 16 shows more link pairs to have correlated reception than channel 26. We found that channel 16 on Mirage overlaps with a cohabited 802.11 network. The 802.11 nodes being higher power systems than 802.15.4 cause losses on multiple 802.15.4 links just as Figure 1 showed.

Figure 5 shows the CCDFs for the SWAN testbed and the Roofnet datatraces at the maximum transmit power level and 11Mbps transmit rate. These plots are representative of CCDFs at other power levels and rates and are not shown for brevity. Less than 20% of all the link pairs have a $\kappa > 0.8$.



(a) Roofnet 11Mbps



(b) SWAN 11Mbps

Figure 5: Complimentary CDF of κ for receiver pairs on the SWAN and Roofnet 802.11 testbeds. Less than 20% of the link pairs have a $\kappa > 0.8$. Roofnet has many more negatively correlated links than SWAN.

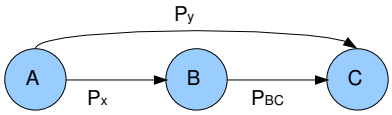


Figure 6: A 3-node 2-hop network. Node A is the source and Node C is the destination. The best shortest ETX path is A→B→C. However, Node C can sometimes hear A directly.

3.6 Summary

The κ metric is based on the cross-correlation index. It varies across networks. The same link pair can have different κ's depending on the channels, power levels and the data rate.

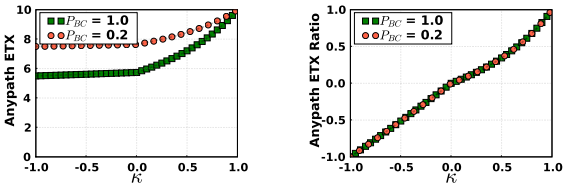
Overall, there are link pairs on all the testbeds that have highly correlated reception. The next three sections explore how inter-link reception correlation can impact the performance of three protocols: a simple opportunistic reception scheme, the ExOR protocol [4], and Rateless Deluge [8], and whether κ is a good metric for predicting this impact.

4. OPPORTUNISTIC ROUTING

This section examines how κ can help predict the performance and benefits of a simple opportunistic reception routing protocol. Because opportunistic routing protocols assume reception at different nodes to be independent, their estimates can be different from the actual average number of transmissions needed to get a packet to a destination. We find that the error of an opportunistic routing protocol's ETX estimate is correlated with κ.

4.1 A Simple Opportunistic Protocol

We consider an opportunistic routing protocol, similar in flavor to the ExOR [4] protocol in 802.11. Every



(a) Anypath ETX (b) Anypath ETX Ratio

Figure 7: Anypath ETX and Anypath ETX Ratio based on Equation 7 for P_{BC}=1.0 and 0.2 with P_x=0.1 and P_y=0.1. Anypath ETX Ratio is easier to visually compare than Anypath ETX.

node has a set of potential nexthop nodes. The nexthop list is prioritized such that if a node receives a packet, it will forward that packet only if none of the higher priority nodes receive the same packet. In reality, a receiver coordination scheme is needed to make sure that all the nexthops know which nodes received this packet. In this section, however, we assume perfect receiver coordination and analyze the performance. This assumption is favorable to opportunistic routing as it does not account for packets that get transmitted due to imperfect coordination. Throughout this paper, we refer to the average number of transmissions to get a packet from a source to destination using opportunistic routing as the anypath ETX [13].

We start our analysis with a simple network of 3 nodes, namely A, B and C. In this setup, A is the source and C is the destination. Figure 6 shows this setup along with the packet reception ratios of all the links. A→B→C is the shortest ETX path but we allow opportunistic routing i.e. if C hears the packets from A then B will not forward such packets. We use random variables x and y to indicate a successful reception on links A→B and A→C respectively. For this setup, the anypath ETX from A to B is:

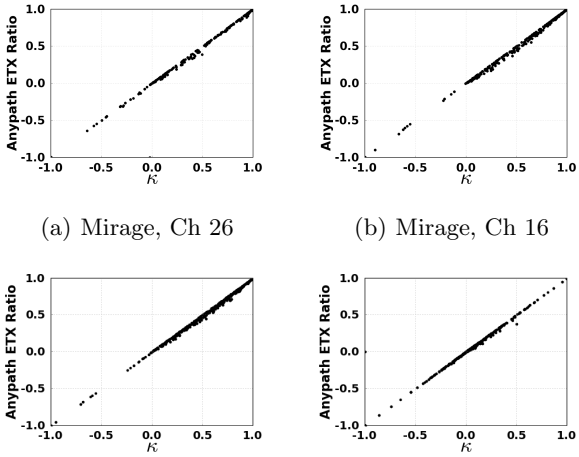
$$E[A] = P^{(A)}_{x,y}(1,0)(1+1/P_{BC}) + P^{(A)}_{x,y}(0,1) + P^{(A)}_{x,y}(1,1) + P^{(A)}_{x,y}(0,0)(1+E[A])$$

$$= \frac{1 + P^{(A)}_{x,y}(1,0)/P_{BC}}{\left(1 - P^{(A)}_{x,y}(0,0)\right)} \quad (6)$$

where, $P^{(t)}_{x,y}(b,c)$ is the probability that $x=b$ and $y=c$ when t transmits, $b,c \in \{0,1\}$.

4.2 Anypath ETX Ratio

From Equations 2, 5 and 6, we can write the anypath ETX as a function of κ as:



(a) Mirage, Ch 26 (b) Mirage, Ch 16

(c) SWAN, 11Mbps (d) Roofnet, 11Mbps

Figure 8: Anypath ETX ratio vs κ for the 2-hop, 3-node network on testbeds. On all the testbeds, the average anypath ETX ratio increases as κ increases.

$$E[A] = \begin{cases} \frac{1+[P_x.(1-P_y)-\kappa.\rho^{max}.\sigma_x.\sigma_y]/P_{BC}}{1-(1-P_x).(1-P_y)-\kappa.\rho^{max}.\sigma_x.\sigma_y}, & \kappa \geq 0 \\ \frac{1+[P_x.(1-P_y)-\kappa.\rho^{min}.\sigma_x.\sigma_y]/P_{BC}}{1-(1-P_x).(1-P_y)-\kappa.\rho^{min}.\sigma_x.\sigma_y}, & \kappa < 0 \end{cases} \quad (7)$$

where $\sigma_x = \sqrt{P_x.(1-P_x)}$, and ρ^{max}, ρ^{min} given by Equations 3 and 4.

Figure 7(a) plots the anypath ETX using Equation 7 by varying κ for P_x=0.1, P_y=0.1 and two values of P_{BC}: 1.0 and 0.2. It shows that the anypath ETX can be quite different depending on P_{BC} and so it is not convenient to observe how it relates to κ. To plot the anypath ETX for any value of P_{BC} and still observe how it relates to κ, we normalize the anypath ETX as:

Anypath ETX Ratio =

$$\begin{cases} \frac{E[A]-E[A]_{inde}}{E[A]_{max}-E[A]_{inde}}, & E[A] \geq E[A]_{inde} \\ \frac{E[A]-E[A]_{inde}}{E[A]_{inde}-E[A]_{min}}, & \text{otherwise.} \end{cases} \quad (8)$$

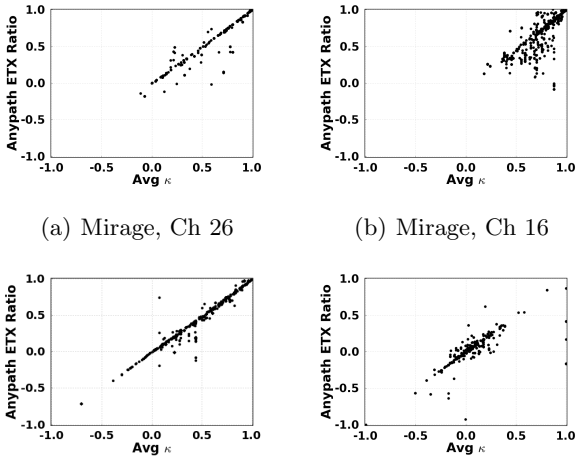
where, the max, independent and min anypath ETX's for the 3-node setup are given by:

$$E[A]_{max} = \frac{1 + (P_x - min(P_x, P_y))/P_{BC}}{max(P_x, P_y)} \quad (9)$$

$$E[A]_{inde} = \frac{1 + P_x.(1-P_y)/P_{BC}}{P_x + P_y.(1-P_x)} \quad (10)$$

$$E[A]_{min} = \begin{cases} \frac{1+P_x/P_{BC}}{P_x+P_y}, & P_x + P_y \leq 1 \\ 1 + (1-P_y)/P_{BC}, & \text{otherwise.} \end{cases} \quad (11)$$

The anypath ETX ratio is zero when the anypath ETX estimate from the opportunistic routing protocol



(a) Mirage, Ch 26 (b) Mirage, Ch 16

(c) SWAN, 11Mbps (d) Roofnet, 11Mbps

Figure 9: Anypath ETX ratio of a node vs average of all the κ's in the anypath for that node. On all the testbeds, the average anypath ETX ratio increases as the average κ of the anypath increases.

matches the actual anypath ETX a source-destination pair observes. It is positive when opportunistic routing protocol under-estimates the anypath ETX, with a maximum of 1 when the under-estimation is the maximum. The anypath ETX ratio is negative when the estimate is over-estimated, with a minimum of -1 when the over-estimation is the maximum.

Lemma 2. *The anypath ETX for the 3-node network is monotonically non-decreasing in κ; the anypath ETX is maximum when κ=1 and is minimum when κ=-1.*

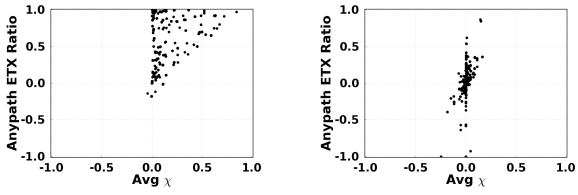
Lemma 2 says that the anypath ETX is maximum when receptions at B and C are correlated, and is minimum when receptions are negatively correlated. The anypath ETX when the receptions are independent is between the max and min anypath ETXs. The proof of this lemma is in the Appendix.

Figure 7(b) plots the anypath ETX ratio against κ for the same setup as in Figure 7(a). The anypath ETX ratio allows easy comparison for paths with different P_{BC}. The anypath ETX increases as κ increases. This is in agreement with Lemma 2

4.3 2-Hop Anypath ETX on Testbeds

This section compares the anypath ETX ratios and κ's for all the possible 2-hops in the testbeds.

We use the same Mirage, SWAN and Roofnet datasets as in Section 3.5. We look at all the cases where a node, A can send packets to node C, either directly or through B. For all such cases, we compute the actual anypath



(a) Mirage, Ch 26 (b) Roofnet, 11Mbps

Figure 10: Anypath ETX ratio of a node vs average of all the χ's in the anypath for that node for the Mirage and Roofnet testbed traces. Many anypath ETX ratios map to a χ of $\approx$ 0.0: χ is not a good indicator of protocol performance.

ETX from A to C, using the data trace. From this anypath ETX, we compute the anypath ETX ratio and compare it with κ of the link pair A→B and A→C.

Figure 8 plots the anypath ETX ratio against κ for both the 802.15.4 and the 802.11 testbeds. In all the testbeds, the average anypath ETX ratio increases as κ increases, as Lemma 2 pointed out.

5. EXTREMELY OPPORTUNISTIC ROUTING (EXOR)

The results presented so far are for the 3-node setup. This section explores if κ is useful in understanding how an opportunistic routing protocol will perform in a general multi-hop, multiple-receiver setting.

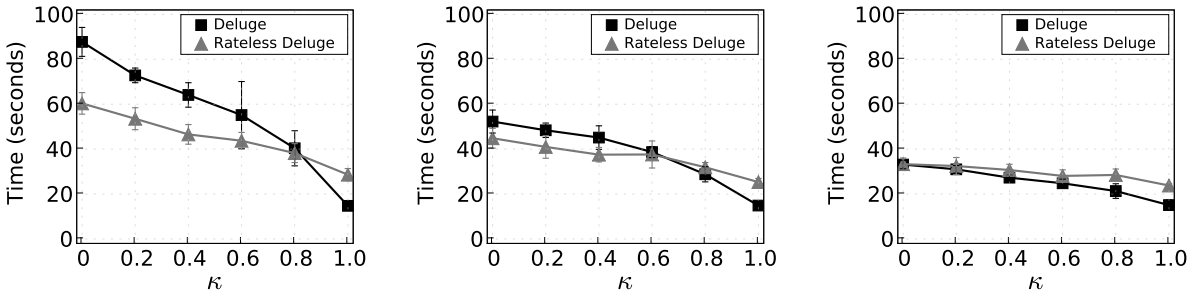
We extend the anypath ETX computation for the general case in which node t is the source, n is the destination and the anypath traverses through nodes 1,.., n-1.

$$E[t] = \frac{1}{1 - P^{(t)}_{1,2..n}(0,..,0)} \Big[1 + P^{(t)}_{1,2..n}(1,0,..,0).E[1] + P^{(t)}_{2,3..n}(1,0,..,0).E[2] + .. + P^{(t)}_{n-1,n}(1,0).E[n-1] + P^{(t)}_{n}(1).E[n]\Big] \quad (12)$$

5.1 Experimental Methodology

We use the same Mirage, SWAN and Roofnet data traces as in Section 3.5. For every source destination pair, the source computes the nexthop list from all the nodes that have a lower shortest path cost to the destination than the source. It includes a node in the nexthop list if that node can send at least 10% of the packets in a batch of size 100 packets. This methodology is same as the one for ExOR proposed by Biswas et al. [4].

For every node in the nexthop list, we compute the anypath ETX ratio using Equation 12. For every such node, n, we compute the average of all the κ's of all the



(a) PRR = 0.4 (b) PRR = 0.6 (c) PRR = 0.8

Figure 11: Time to disseminate for Rateless Deluge and Deluge for varying values of κ and PRR. Rateless deluge is better for lower κ, but Deluge is better for higher κ. The performance switch point varies based on the PRR of the links.

link pairs in which n is the transmitter and the receiver is from the n's nexthop list. For example, for a node 1 with nodes 2, 3 and 4 in its nexthop list, we compute the average of κ's of link pairs {1→2, 1→3}, {1→2, 1→4} and {1→3, 1→4}.

5.2 Results and Observations

Figure 9 plots the anypath ETX ratio against the average κ for every node for every source-destination pair for different testbeds. In general, the anypath ETX ratio increases as the average κ of the anypath increases; as more receiver pairs in the anypath are correlated, the error in the anypath ETX estimate from opportunistic routing increases and the estimate is usually an underestimate. Figure 10 plots the anypath ETX ratio against average anypath χ for the Mirage and Roofnet testbeds. χ is often close to 0 for different values of ETX ratio: using χ can lead to the incorrect conclusion that link pair receptions are uncorrelated. χ is a bad indicator of ExOR's performance.

Opportunistic routing protocols assume inter-link reception independence. When the reception on links are not independent of each other, these estimates are different from the actual anypath ETX. κ is capable of showing when these estimates are likely to be correct and when they are very different from the actual anypath ETX. Although κ is a 2-receiver metric, the average of all the κ's in the anypath still give us information on how good the anypath ETX estimates are.

6. DELUGE AND RATELESS DELUGE

Section 4 showed how κ can help us understand opportunistic routing protocol performance. This section explores if κ can give insights on network coding protocols, and shows that network coding is not always beneficial. Specifically, this section looks at the performance of two dissemination protocols: Deluge (which does not use network coding) [9] and Rateless Deluge [8] which uses network coding to reduce the number of packet transmissions needed for dissemination. We find that κ can guide us as to *when* network coding protocols are beneficial. The network coding technique of Rateless Deluge is more efficient only when the network reception correlation is low.

6.1 Controlled Experiment

To understand the implications of inter-link correlation to Rateless Deluge, it's imperative to vary κ in a controlled way and look at Rateless Deluge's performance for each κ. A single transmitter disseminates on channel 26 at maximum transmit power to 7 single-hop, near-by receivers. The maximum transmit power gives perfect links to all the receivers. The dissemination image has 9 pages with 20 packets per page.

We introduce two kinds of random losses: correlated and independent. To introduce correlated losses, the transmitter randomly drops packets from its transmit queue, causing same packets to be lost at all the receivers. For independent losses, every receiver randomly drops packets independently of other receivers. If P_t is the probability of packet loss at the transmitter and P_r is the probability of loss at every receiver, then varying P_t and P_r varies κ and the PRR of all the links. For example, if P_t=0.0 then the losses at the receivers are independent with a κ of 0.0 for any link pair. On the other hand, P_r=0 make all the losses to be correlated with a κ of 1.0 for any link pair. The PRR is given by $1 - (P_t + (1 - P_t) * P_r)$.

This experiment runs for PRR values between 0.4 and 0.9 and for each PRR value κ is varied between 0 and 1. For each combination of PRR and κ values, we run 10 experiments with Rateless Deluge and Deluge.

For all the PRRs, Rateless Deluge sends same or fewer number of packets as Deluge. This is an expected be-

	Simple (Avg κ=0.55, Avg PRR=0.91)		WiFi (Avg κ=0.85, Avg PRR=0.85)		Movement (Avg κ=0.04, Avg PRR=0.57)	
	Packets Sent	Latency (sec)	Packets Sent	Latency (sec)	Packets Sent	Latency (sec)
Rateless Deluge	339.2	29.8	424.6	36	473	42.8
Deluge	329.8	22.4	377.4	25	638.6	50.8
Improvement	-3%	-33%	-13%	-44%	26%	16%

Table 1: Comparison of the performance improvement obtained by using Rateless Deluge instead of Deluge for dissemination in different network environments. Each reading in the table is calculated as the average of readings from five different experiments. The table shows that when κ and PRR are high, Deluge performs more efficiently than Rateless Deluge. But in the presence of uncorrelated links and lower PRR, Rateless Deluge clearly has better performance.

havior and is not shown for the same purpose. However, for the time to disseminate there is no clear winner. Figure 11 shows the total dissemination time for both Rateless Deluge and Deluge for different κ values for different PRRs (in different plots).

When κ is close to 1.0, Deluge finishes dissemination sooner than Rateless Deluge, for all the PRRs. When the links lose the same packets, both Deluge and Rateless Deluge send the same number of packets. Therefore, the additional time taken by Rateless Deluge is purely the coding overhead. The low-power cpu in Telosb motes causes the coding overhead time to be significant.

On the other hand, when κ is close to 0, Rateless Deluge finishes sooner than Deluge. For the κ values in between, the κ value where Deluge starts to out-perform Rateless Deluge depends on the PRR of the links. As links get poorer, this transition point shifts to the right.

This result suggests that the κ value of a network can be used to help decide which protocols should be used for the network.

6.2 Performance Comparison

The previous experiments had controlled losses introduced at the motes to understand protocol behavior across a range of κ values. Further, due to the nature of the controlled experiment, the link PRRs at all the receivers were the same. We now explore, without any controlled losses introduced by us, and with different link PRR's across links, how Rateless Deluge and Deluge perform.

We compare Deluge versus Rateless Deluge in a small testbed of 9 telosb motes. We evaluate the performance of the two protocols under three different scenarios:

- **Simple Scenario:** We placed nodes randomly in a small part of a room measuring roughly 5'x3'x5'. The experiments were run during normal office time. This experiment uses channel 26, so that it does not experience any WiFi interference. The power level for transmissions is set to 1.
- **WiFi Scenario:** 802.15.4 and WiFi networks share the spectrum and are usually co-located. For this scenario, we use the same topology as in the simple scenario above, but use channel 25 on the motes instead. This makes the motes experience correlated losses due to WiFi interference.
- **Movement Scenario:** In this scenario, nodes are spaced further apart with transmission power level set to 2. We constantly keep moving the transmitter and have people walking around, both of which reduce packet loss correlation between links.

Prior to running dissemination experiments for the three scenarios, for each scenario, the transmitter sends 50,000 broadcast packets. The receivers note down which packets they receive. We use this information to compute κ and PRR for every link pair and report the average of all such κ's and PRR's.

Table 1 shows the results for the three scenarios along with the average κ values. Unlike, the controlled experiment, Table 1 also shows the number of packets sent by both Deluge and Rateless Deluge, because in this case, Rateless Deluge is not consistently better than Deluge in terms of packets sent. This is because in the controlled experiment all the links had same PRR. This is not the case for the three uncontrolled scenarios.

For the simple scenario, Deluge outperforms Rateless Deluge in terms of dissemination time – Rateless Deluge takes 33% longer to finish. The average κ for this scenario is 0.55 and average PRR is 0.91. As our controlled experiment predicted, for very high PRR links, Deluge is better than Rateless Deluge, even for medium κ values.

For the WiFi scenario also, Deluge outperforms Rateless Deluge. The average κ for this scenario further increased to 0.85 and the average link PRR dropped to 0.85. The higher κ is because the higher power WiFi systems cause same packets to fail at multiple 802.15.4 nodes, making links more correlated. The number of packets sent by both Deluge and Rateless Deluge

increases by 15-25% compared to the simple scenario. This is due to losses caused by WiFi. The higher κ value increases the gap in Deluge and Rateless Deluge performance compared to the simple case, with Rateless Deluge taking 44% more time and 13% more packets than Deluge.

The results for the movement scenario match our predictions about the two protocols for the case of independent receptions. Rateless Deluge is significantly more efficient since it effectively uses network coding to reduce the number of packets that need to be rebroadcast. Rateless Deluge sends 26% fewer packets and finishes 16% earlier than Deluge. The average κ for this scenario is 0.04.

6.3 Summary

These results show that a disseminating node can use the distribution of all its receiver pairs' κ values to decide whether to use Deluge or Rateless Deluge for data dissemination. If the distribution tends towards higher values of κ, or if the network has very high PRR links, Deluge is likely to be more efficient. Otherwise, Rateless Deluge generally does better than Deluge. This also validates our assertion that network protocols can benefit from a knowledge of the inter-receiver correlation to improve their performance and efficiency.

7. RELATED WORK

In the Wireless Communications field, there is much work in understanding received signal strength correlation over distance [18]. This has implications to PHY level receiver design such as multiple antenna systems. Our work looks at correlations that manifest at the link-layer i.e packet-level correlations. Such correlations have implications to higher layer protocols using existing hardware.

Xu et al. explored the correlation of average PRRs of links over the geographic space [25]. Our work, in contrast, is agnostic to geographic location and looks at how, a packet's reception is correlated on two links. We have yet to relate κ to the geographic location of the nodes.

Pattem et al. explore the spatial correlation of physical quantities such as temperature and humidity to optimizing data aggregation by eliminating redundant geographic-specific information gathered by densely populated nodes [17]. This work is complementary to ours.

To our knowledge, our work is the first to investigate a metric that captures the inter-link reception correlations and its implications to protocol performance.

8. CONCLUSION AND DISCUSSION

This paper shows that packet receptions over different links need not be independent of each other. It presents a metric called κ that captures this inter-link correlation. κ can indicate how the performance of opportunistic routing protocols like ExOR varies due to link correlations. It also shows that knowing how correlated inter-link receptions are in a network can guide us pick the protocols that work well in that network. Specifically, it shows that when the receivers have correlated reception, using Deluge is preferable over its network-coding counterpart, Rateless Deluge.

These results suggest that protocols should consider the dynamics of the network and adapt accordingly to be more efficient. For example, a modified version of Rateless Deluge that falls back to regular Deluge when the receivers have correlated reception will benefit the best of both the worlds and achieve greater efficiency.

κ proves to be a better metric for measuring inter-link correlations than the cross-conditional metric commonly used in current literature. Quantifying reception correlations using κ shows that some networks show high levels of correlation. For the testbeds measured in this paper, correlations tend to be higher for 802.15.4 networks compared to 802.11 networks. This could in part be attributed to the lower physical distance between nodes in 802.15.4 networks.

In our experience, same network can have different inter-link correlations on different channels. Therefore, when researchers publish protocol comparison results, it will be useful to report κ distribution of their network for the corresponding channel. This will allow for reproducibility and deeper understanding of the comparison.

Variations in signal and noise cause links to be positively or negatively correlated. An obstacle, such as a person, may block line of sight from the transmitter to two receivers, preventing them from receiving packets. If the obstacle moves away, the signal strength at both receivers will increase causing positive signal strength correlation leading to positive packet reception correlation at the two receivers. External interference can also cause positive correlations; positive correlations on channel 16 are due to external 802.11 sources. A spike of external interference causes the signal-to-noise ratio at both receivers to drop, causing correlated reception events.

We have also noticed that a moving obstacle can create negative correlations. The obstacle can move such that it blocks one receiver at a time, while allowing line-of-sight to the other receiver. In this case, the receptions are negatively correlated and the packet reception is also negatively correlated. High power wireless interferers can cause negative interference as well. The two interferers may communicate using CSMA/CA, such that one interferer is transmitting at a time. If each interferer causes packet drops at a different node, they cause negative reception correlation.

Current simulators like TOSSIM do not take link reception correlations into account. Incorporating κ as an input to simulators can lead to more accurate, testbed-specific simulations. This will allow for fair protocol performance comparisons and more realistic results. Addressing this issue is an open problem.

Significant changes to the environment can alter κ. For example, adding a new WiFi access point will introduce more correlated losses and increase κ. Computing κ online will keep track of such changes, which is a work in progress.

9. REFERENCES

[1] ANSI/IEEE Std 802.11 1999 Edition.

[2] Stanford Wireless Access Network (SWAN). `sing.stanford.edu/swan`.

[3] The Network Simulator ns-2 (v2.1b8a). http://www.isi.edu/nsnam/ns/.

[4] S. Biswas and R. Morris. Exor: opportunistic multi-hop routing for wireless networks. In *SIGCOMM '05: Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, 2005.

[5] S. Chachulski, M. Jennings, S. Katti, and D. Katabi. Trading structure for randomness in wireless opportunistic routing. In *SIGCOMM '07: Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, 2007.

[6] R. Fonseca, O. Gnawali, K. Jamieson, and P. Levis. Four bit wireless link estimation. In *The Sixth Workshop on Hot Topics in Networks (HotNets-VI)*, Nov. 2007.

[7] R. Gray and L. Davisson. *Introduction to Statistical Signal Processing*. Cambridge University Press, New York, NY, USA, 2005.

[8] A. Hagedorn, D. Starobinski, and A. Trachtenberg. Rateless deluge: Over-the-air programming of wireless sensor networks using random linear codes. In *IPSN '08: Proceedings of the 7th international conference on Information processing in sensor networks*, pages 457–466, Washington, DC, USA, 2008. IEEE Computer Society.

[9] J. W. Hui and D. Culler. The dynamic behavior of a data dissemination protocol for network programming at scale. In *Proceedings of the Second International Conferences on Embedded Network Sensor Systems (SenSys)*, 2004.

[10] Intel Research Berkeley. Mirage testbed. `https://mirage.berkeley.intel-research.net/`.

[11] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft. Xors in the air: practical wireless network coding. In *SIGCOMM '06: Proceedings of the Conference on Applications, technologies, architectures, and protocols for computer communications*, 2006.

[12] E. Kohler, R. Morris, B. Chen, J. Jannotti, and M. F. Kaashoek. The Click modular router. *ACM Transactions on Computer Systems*, 18(3):263–297, August 2000.

[13] R. Laufer, H. Dubois-FerriÃĺre, and L. Kleinrock. Multirate anypath routing in wireless mesh networks. In *Proceedings of IEEE Infocom 2009*, Rio de Janeiro, Brazil, April 2009.

[14] P. Levis, N. Lee, M. Welsh, and D. Culler. TOSSIM: Simulating large wireless sensor networks of tinyos motes. In *Proceedings of the First ACM Conference on Embedded Networked Sensor Systems (SenSys 2003)*, 2003.

[15] MIT Roofnet. 802.11 testbed. `http://pdos.csail.mit.edu/roofnet/`.

[16] A. Miu, G. Tan, H. Balakrishnan, and J. Apostolopoulos. Divert: Fine-grained path selection for wireless lans. In *Proceedings of the Second International Conference on Mobile Systems, Applications, and Services (MobiSys 2004)*, 2004.

[17] S. Pattem, B. Krishnamachari, and R. Govindan. The impact of spatial correlation on routing with compression in wireless sensor networks. In *IPSN '04: Proceedings of the 3rd international symposium on Information processing in sensor networks*, pages 28–35, New York, NY, USA, 2004. ACM.

[18] T. Rappaport. *Wireless Communications: Principles and Practice*. Prentice-Hall, 1996.

[19] C. Reis, R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan. Measurement-based models of delivery and interference in static wireless networks. In *SIGCOMM*, pages 51–62, 2006.

[20] F. Stann, J. Heidemann, R. Shroff, and M. Z. Murtaza. RBP: Reliable broadcast propagation in wireless networks. Technical Report ISI-TR-2005-608, USC/Information Sciences Institute, November 2005.

[21] R. Szewczyk, J. Polastre, A. Mainwaring, and D. Culler. Lessons from a sensor network expedition. In *Proceedings of the First European Workshop on Sensor Networks (EWSN)*, Berlin, Germany, Jan. 2004.

[22] C. Technology. Micaz datasheet. `http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/MICAz_Kit_Datasheet.pdf`, 2006.

[23] The Institute of Electrical and Electronics

Engineers, Inc. Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs), Oct. 2003.

[24] TinyOS. MultiHopLQI. `http://www.tinyos.net/tinyos-1.x/tos/lib/MultiHopLQI`, 2004.

[25] Y. Xu and W.-C. Lee. Exploring spatial correlation for link quality estimation in wireless sensor networks. In *PERCOM '06: Proceedings of the Fourth Annual IEEE International Conference on Pervasive Computing and Communications*, pages 200–211, Washington, DC, USA, 2006. IEEE Computer Society.

APPENDIX

Proofs of the Lemmas

Proof of Lemma 1: We wish to determine the maximum and minimum values that ρ can take on. If P_x and P_y are kept constant, then $E[x] = P_x, E[y] = P_y, \sigma_x = \sqrt{P_x \cdot (1 - P_x)}, \sigma_y = \sqrt{P_y \cdot (1 - P_y)}$ are all constant as well. From the expression of ρ in Equation 2, we can see that the only variable term is $E[x \cdot y]$. Thus to maximize/minimize ρ we must maximize/minimize $E[x \cdot y]$.

We have that $E[x \cdot y] \leq E[x] = P_x$ and $E[x \cdot y] \leq E[y] = P_y$. These imply that $E[x \cdot y] \leq \min(P_x, P_y)$. Assuming without loss of generality that $P_x \leq P_y$. This inequality can be achieved as an equality by the distribution where:

$\{Px, y(1,1) = P_x, P_{x,y}(1,0) = 0, P_{x,y}(0,1) = P_y - Px, P_{x,y}(0,0) = 1 - P_y\}$.

Thus the maximum value of $E[x \cdot y] = \min(P_x, P_y)$, which yields the maximum value of ρ indicated in Equation 3.

Since $(x-1), (y-1)$ are both negative, $E[(x-1) \cdot (y-1)] = E[x \cdot y - y - x + 1] \geq 0$. Rearranging terms, and noting again that $E[x] = P_x, E[y] = P_y$, we get that $E[x \cdot y] \geq P_x + P_y - 1$. Further, since x and y are both non-negative random variables, we have that $E[x \cdot y] \geq \max(0, P_x + P_y - 1)$. This inequality is achieved with equality under the following two distributions. When $P_x + P_y \leq 1$, the following distribution minimizes $E[x \cdot y]$ to 0:

$\{P_{x,y}(1,1) = 0, P_{x,y}(1,0) = P_x, P_{x,y}(0,1) = P_y, P_{x,y}(0,0) = 1 - P_x - P_y\}$.

When $P_x + P_y \geq 1$, the following distribution minimizes $E[x \cdot y]$ to $P_x + P_y - 1$:

$\{P_{x,y}(1,1) = P_x + P_y - 1, P_{x,y}(1,0) = 1 - P_y, P_{x,y}(0,1) = 1 - P_x, P_{x,y}(0,0) = 0\}$.

These minimum values of $E[x \cdot y]$ correspond to the minimum values of ρ indicated in Equation 4. □

Proof of Lemma 2: This lemma applies only if $P_{BC} > P_y$. This basically means that B is included as a possible next hop only if it has a better path to destination than A, which is a common assumption for opportunistic routing protocols. Under this constraint, the derivative $\frac{dE[A]}{d\rho}$ of the expression in Equation 7 can be shown to be:

$$\frac{dE[A]}{d\rho} = \frac{\sigma_x . \sigma_y . (P_{BC} - P_y)/P_{BC}}{\left(1 - (1 - P_x) \cdot (1 - P_y) - \rho . \sigma_x . \sigma_y\right)^2} \quad (13)$$

where $\sigma_x = \sqrt{P_x \cdot (1 - P_x)}$. Thus, for $P_{BC} > P_y$, $E[A]$ is monotonically non-decreasing in ρ. Further, for a given set of link PRRs, $\kappa \propto \rho$ for both $\rho > 0$ and $\rho < 0$. Also, $\kappa > 0$ for $\rho > 0$, $\kappa < 0$ for $\rho < 0$ and $\kappa = 0$ for $\rho = 0$. Then, ρ is monotonically non-decreasing with κ. By transitivity of monotonicity, $E[A]$ is monotonically non-decreasing with κ. □