Comparing Simulation Tools and Experimental Testbeds for Wireless Mesh Networks

Kefeng Tan, Daniel Wu, An (Jack) Chan, Prasant Mohapatra
Department of Computer Science
University of California, Davis, CA 95616
E-mail: {kftan,danwu,anch,pmohapatra}@ucdavis.edu

***Abstract*—Simulators provide full control for researchers to investigate wireless network behaviors, but do not always reflect real-world scenarios. Although previous works point out such shortage is due to the limitation of radio propagation models in the simulators, it is still unclear how imperfect modeling affect network behavior and to what degree. In this paper, we investigate wireless mesh network behavioral differences between simulations and real-world testbed experiments. We compare and analyze the experimental results with NS-2 and Qualnet simulations. We find that in the PHY layer, the distribution of received signal strength in experiments is usually different from simulation due to the antenna diversity. However, path loss, which is regarded as a dominating factor in simulator channel modeling, can be configured to fit to real-world behaviors. At the MAC layer, simulators and testbeds response differently to heavy traffics. While a significant performance degradation is observed in experiments, it is less obvious in simulations. The inadequacy in capturing interference further widen the discrepancy. Sensitivities differences exist in hardware receivers result in significant unfairness in flow-level goodputs, but never be a problem in simulators. On the IP layer, we focus on route prevalence and persistence, and find that a few dominant routes exist in experimental testbed while routes in simulators are less stable. These findings give the wireless research community an improved picture about the differences between simulations and testbed experiments. They will help researchers to choose between simulations and experiments according to their particular research needs.**

I. Introduction

In the past decade, due to the emphasis on practicality, testbed based approaches are more popular in wireless networking research than simulations. Many of the seminal works, such as Roofnet [1], MeshNet [11], ORBIT [15], and QuRiNet [20] are based on testbed experiments. On the other hand, due to the hardware limitations and high costs in testbed experiments, simulators still play an important role in wireless network protocol development [4], [5], wireless system performance evaluation [3], and on-line wireless traffic analysis [7].

Simulations have advantages that can hardly be replaced by testbed experiments. In simulations, network scenarios can be easily constructed and modified, and data can be easily collected. More importantly, simulations can model large scale network topologies which would be very expensive, if not impossible, in testbed experiments. Building a wireless network testbed would require hardware and labor resources. Moreover, testbed experiment results are heavily affected by the testing environment, which is often highly random and uncontrollable. For example, a little, tiny change surrounding a wireless communication parties such as temperature increases, or a door is closed, can affect the communication quality, and thus change their throughput.

On the other hand, however, wireless network simulators have their own limitations. Due to the inadequacy of modelings, especially at the PHY layer, simulators are often accused of not being able to provide as trustworthy results as real testbed does. In this paper, we focus on the discrepancies between simulators and testbed experiments. We aim to find out how the imperfectness of channel modeling in simulators affect network behaviors and to what degree. In particular, we are interested in multihop flow behavior differences.

Our contributions are three-fold:

- To the best of our knowledge, we give the most comprehensive and systematic comparisons between simulations and experiments in wireless mesh network.
- In each of the PHY, MAC and IP layers, we find some discrepancies between simulations and experiments, and analyze the corresponding root cause. In the PHY layer, we find that simulators fail to model the antenna diversity which is widely exploited in real world Wi-Fi devices. [1] However, path loss, a a dominating factor in simulator channel modeling, can be configured to fit some of real-world scenarios. In the the MAC layer, a serious flow level unfairness exist in real testbed, but not captured in wireless simulators. In the network layer, a few dominant routes exist in testbed experiments while simulation routes are less stable.
- We give a clear picture about the major differences between simulators and testbeds, and how these differences affect the network traffic flow behavior. It helps researchers weigh pros and cons in choosing between simulations or testbeds when carrying out wireless network studies.

The rest of paper is organized as follows. In Section II, we briefly introduce the background and related works. We then describe the experimental setup in Section III. In section IV, we start with point to point measurements. The PHY, MAC

[1]Most modern Wi-Fi devices are equipped with multiple antennas and automatically select the antenna with the strongest received signal strength to use. It is worth noting that antenna diversity here is different from the MIMO (multiple-input multiple-output) technique in IEEE 802.11n [14].

and IP layer differences between simulations and testbeds are detailed in Section V, Section VI, and Section VII, respectively. We conclude the paper in Section VIII.

II. Background and Related Work

NS-2 (network simulator 2) is one of the most popular wireless network simulators. Due to its open source nature, it's widely used in academia. Ever since the IEEE 802.11b PHY/MAC model was added into NS-2, many enhancements modules (i.e., IEEE 802.11a/g/e, multirate schemes, energy consumption models) have been incorporated. QualNet is a commercial alternative to NS-2. It was developed from the GloMoSim simulator [23]. Compared to the former, QualNet takes more factors into the considerations (i.e., antenna parameters, weather factor) . In addition, QualNet is suitable for parallel simulations and has a built-in support for some of the mostly useful statistics about simulations.

However, both of them are still plagued by inaccurate modeling of wireless networks. One of the more prominent problems is inadequate modeling of the wireless physical layer [10], [2]. Many assumptions that simulation modelings rely on have been shown to be invalid in realities. For example, received signal strength (RSS) has been modeled as a simple function of distance in simulators, but so far there is no perfect physical model that can capture all the factors due to the random nature of the wireless channel.

Nevertheless, simulators are still very useful due to its simplicity, flexibility, and result repeatability in studying network behaviors. Understanding discrepancies between simulations and real-world testbed experiments can provide important insights for the networking research community.

Previous works have studied wireless network behavior in the real world, but do not compare that with simulations. For example, Ratul et al. [12] analyzed the MAC behaviors and Krishna [13] compared the routing stability in Roofnet and MeshNet, but neither of them compared their results with simulation results. The work in [8] also compare traffic flow behavior in simulations with those from testbed experiments. However, the study is limited to an indoor network and evaluated NS-2 simulator only. In contrast, our comparisons are most comprehensive and systematic, which carried over NS-2, QualNet, and testbed experiments on different hardware platforms in a wide range of network scenarios.

III. Methodology

Our approach are experiment-driven. We design and set up testbeds for different network scenarios, and then replicate a similar scenario in simulators. We compare metrics which include goodput, delay, delay jitter and some other statistical results like probability distribution of dominant routes to measure the differences between experiments and simulations.

We study many parameters: path loss and antenna diversity (PHY layer); traffic load, bit rate, number of hops, flow fairness and interference (MAC layer); and route length, persistence, stability and diversity of routes (routing layer). To analyze how one particular parameter affects the system performance, we fix all other parameters to minimize their impacts. For example, to find out how bit rate affects a multihop flow, we use static routing to avoid potential influences from different routing protocols.

A. Testbed Platforms

We use three testbed platforms in this study for different purposes. In our settings, wireless mesh network is constructed with laptops. Meanwhile, realizing the hardware differences in computation power, buffer sizes, and number of antenna, we also conduct experiments on embedded system equipped with a Wi-Fi device. To better analyze IP layer characteristics, we leverage a larger size of outdoor network, QuRiNet that consisted of more than 30 nodes. As for simulators, we choose NS-2 and Qualnet. We now describe them in details in the following.

Soekris: We use Soekris boards as the mesh routers. They are 266 MHz x86 Soekris net4826 embedded devices running custom built Linux distribution with 2.6.23 Linux kernel [17]. Each of them has 128MB SDRAM main memory and 64MB compact flash, and equipped with one antenna. Associated clients with the mesh routers are HP nc6000 laptops running Linux kernel of 2.6.25. Each client also has a wireless card with an Atheros chipset for wireless connection to the router. MadWifi (version 0.9.4) is installed in all nodes, including routers and clients. We deliberately configure the wireless mesh network with IEEE 802.11a to reduce interference from existing 802.11b/g networks. All nodes are operating on channel 36 (5.18GHz). This platform represents a series of low-end 802.11 devices with one antenna, and have limited processing capability and buffer size.

The Soekris testbed is used for indoor experiments only because it needs AC power. As shown in Fig. 1, the testbed is composed of nine nodes. The squares represent the laptops, while the circles are the Soekris boards. There are two competing parallel flows. This experiment is done on the second floor of a three-story office building, where concrete walls create non-line-of-sight transmission environment. The solid line indicates the connectivity among nodes. By *connected*, it means a node can directly (without routing) *ping* another node in the network with more than 50% successful rate at 6 Mbps. Otherwise, the link between them is defined as unconnected. The interference range, from our measurement, is roughly two-hops away.

Laptops: In this testbed, only laptops are used to build up a mesh network platform. Laptops are HP model nc6000. Each of them is equipped with Intel Pentium M 1.6 GHz processor with 512 MB DDR SDRAM, and HP W500 802.11a/b/g wireless LAN cards with an Atheros chipset. The operating system is Linux with kernel version 2.6.25 and WLAN driver is MadWifi (version 0.9.4). As in Soekris, all experiments are conducted on Channel 36. This platform represents a set of modern 802.11 devices with dual antennas (one on each side of LCD screen), and have sufficient processing capability and buffer size.

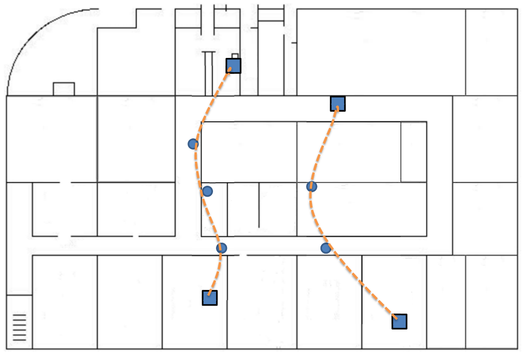


Fig. 1. Soekris indoor mesh network topology.

We carry out both indoor and outdoor experiments on this laptop-based mesh network testbed. The outdoor case is performed on an open grass field, so all transmissions are line-of-sight. The indoor case is conducted in a reading room of a public library, where bookshelves are neatly arranged in the center, surrounded by reading desks and chairs. The room has the width of 32 m and has the length of 51 m. In indoor scenario, transmissions among nodes are non-line-of-sight.

QuRiNet: To collect the network performance results, we also perform some routing experiments in QuRiNet, an outdoor mesh network [20]. QuRiNet is located in the Quail Ridge Natural Reserve, California. It consists of more than 30 nodes and provides wireless coverage over 2,000 acres of hilly terrain. All the access points in QuRiNet are dual radio and utilize multiple channels to achieve higher goodput. For the detailed topology of QuriNet, interested readers are referred to [20].

B. Simulations

We use NS-2 (version 2.34) and QualNet (version 4.0) as target simulators [6], [18]. They are the two most commonly used wireless network simulators. We run all the simulations on them. In each simulation, we configure the simulation parameters to represent the values observed in the real network testbed. Due to space limitations, these parametric values are not individually presented here.

C. Duration for Simulations and Experiments

For each simulation, each test runs for five minutes, and is repeated for five times. For experiments on testbeds, each test runs for two minutes, and is repeated for five times as well. Unless we state explicitly, all the results presented in this paper are means of multiple runs.

IV. Point to Point Measurement

Point to point performance comparisons lay down the foundation for multihop network comparisons. The meaning of comparing is towfold. On one hand, it will help us understand some of the fundamental differences between simulations and experiments. On the other, the observations and rules of thumb from it can better prepare us for multi-hop comparisons in later sections.

Method: In this set of tests, we send packets from one laptop to another at different bitrate schemes, distances, traffic loads and with TCP and UDP protocol respectively. Indoor test is conducted in an empty garage, and outdoor test is carried on a campus grass field. TCP and UDP are studied respectively. Goodputs across platforms are compared in TCP, and delay jitter is also included in UDP.

A. TCP Performance

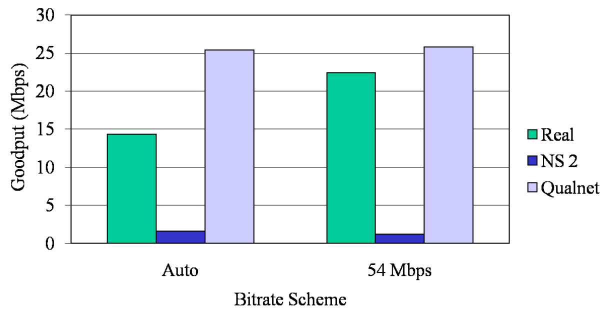
We first compare their TCP performances. Fig. 2- (a) and (b) show goodputs from a pair of laptops, NS-2 and QualNet. To quantify the impacts of bitrate scheme, we change the bitrate from auto to 54 Mbps. In experiments, laptops have better performance indoor than outdoor. Because inside the empty garage, the environment not only does not have any hurdles to block radio transmissions, the indoor reflections can also facilitate the propagations. On the grass field, the laptops are laid down on ground, and the height of antenna (on the edge of LCD screens) is merely 0.15 meters, which results in a lower goodput. In both cases, goodput of Qualnet is higher than that of the laptops, and the goodput of NS-2 is the lowest. In indoor case, 54 Mbps rate outperforms the auto bitrate adaptation, while in outdoor case the latter works better. But simulations are not affected much by the environment.

We then change the transmission distance and show the results in Fig. 2- (c) and (d). After the pair is set further apart, the laptop sender delivers less amount of packets. So does NS-2, reporting almost zero goodput. Qualnet, on the other hand, shows no changes of goodput in response to distance increase and the environment change.

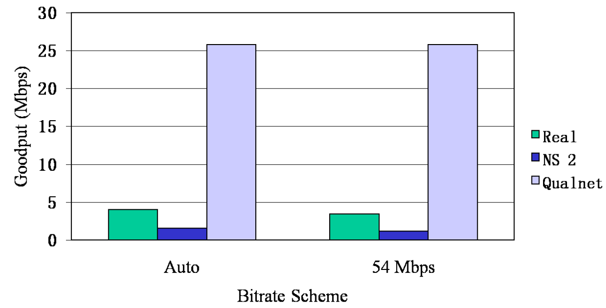
B. UDP Performance

As we have shown, TCP performance in simulations and experiments have major diverges. This is because TCP protocol has a close loop control on transmission rate based on round trip time (RTT), packet retransmission rate, packet loss rate, etc. To mitigate the impacts from the transportation layer, we also evaluate the UDP performance. On top of UDP, constant bit rate (CBR) application is running. As shown in Fig. 3 (a) - (d), their goodput performance is similar when the channel is relatively reliable.

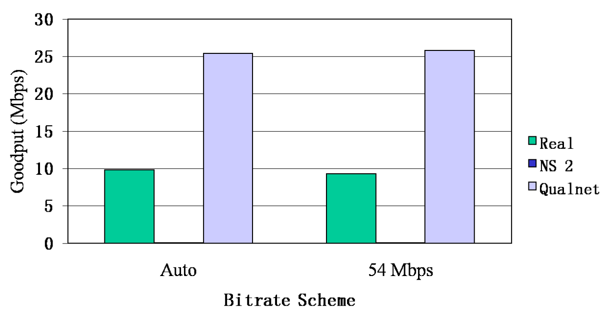
Either separating the laptop pair further apart to 80 meters, or just moving it outdoor, its goodput starts diverging from simulations, as shown in Fig. 3 (e) - (h). In indoor cases, packets are delivered at about 10 Mbps in experiments and Qualnet, but NS-2 reports merely 5 Mbps goodput. In outdoor cases, however, laptop goodput plunges dramatically while NS-2 and Qualnet do not response to the environment change. In another word, the parameters modifications in simulations are simply not significantly enough to reflect the change. Just as in the TCP cases, laptop UDP goodputs are consistently higher indoor than outdoor, and Qualnet always gets the highest result among the three.



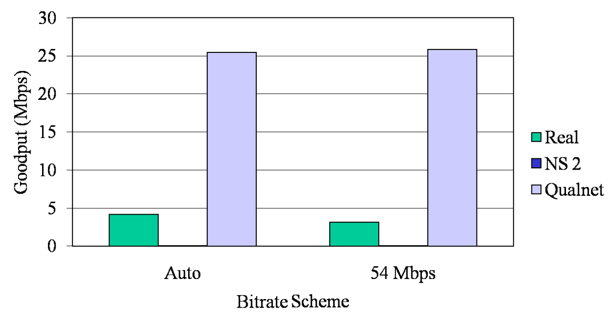
(a) TCP goodput: 40m apart indoor



(b) TCP goodput: 40m apart outdoor



(c) TCP goodput: 80m apart indoor



(d) TCP goodput: 80m apart outdoor

Fig. 2. Point to Point TCP Goodput Comparisons

For time related metric, we compare the delay jitter, and present results in Fig. 4. Delay jitter here is defined as the smoothed mean of differences between consecutive transit times. It roughly inverse proportional to the goodput, meaning the higher the goodput, the lower the delay jitter. But when all packets are successfully delivered and the jitter is really low, like in 4- (a) and (b), the difference is on the order of 0.1 ms.

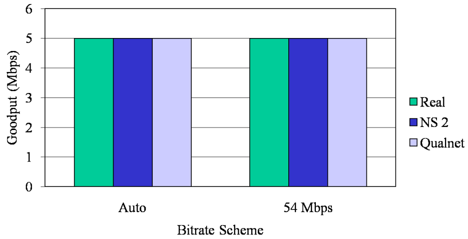
As their differences in goodput getting larger, the differences in jitter becomes remarkable. It is understandable that Qualnet almost always has the lowest delay jitter since it transmits the largest amount of packets. For NS-2 and laptops, it is noticeable that both of them report higher delay jitter at 54 Mbps than that at autorate across many scenarios, e.g. Fig. 4-(c) to (h) even if the goodputs at two rate schemes are similar. This can be explained as transmitting at 54 Mbps incurs more retransmissions than autorate schemes, and retransmissions alter the transit times. But in Fig. 4-(e) and (g), there are exceptional cases that a smaller jitter at 54 Mbps on laptops is observed. Rather than that, there exist a major divergence in delay jitter between simulations and realities.

C. Conclusions

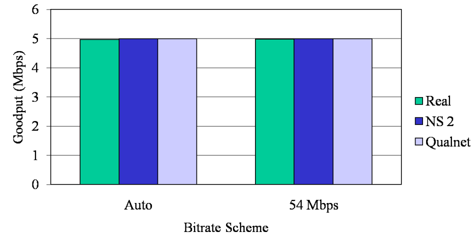
In the above experiments and simulations, we followed the instructions in simulator manuals, and tried to adjust a few key parameters to adapt to the environment in experiments. Nevertheless, we do not claim our configurations in simulators are most accurate, such that they can vividly replicate everything in reality. Detailed evaluations of parameters on how they affect results will be discussed in Section V. What we did are simply an average user will do in simulations. Above results suggest that certain differences exist between simulations and realities, and need to be used with cautions.

The results also indicate that Qulanet typically has an optimistic channel modeling, and NS-2 has a relatively conservative estimation of channels. The difference is more obvious as wireless channels are deteriorated, such as in outdoor, or when transmission distance is increased. The existing modelings in simulations are generally good enough for reliable channels, but are not sufficiently accurate to capture unstable channels.

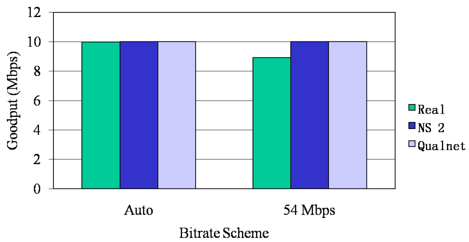
For a fair comparison of multihop network, we want to separate the problems on high layers from underneath, and the impacts from PHY layer should be mitigated. To this end, we summarize a few take-home lessons. Realizing that TCP protocol has its own congestion control and recovery algorithm that dramatically impacts goodputs and deepens the gap between simulations and experiments, we use UDP instead for further comparisons. Besides, to minimize the possibility of operating on unreliable channels, long distance transmissions, e.g. over 40 meters, should be avoided. Moreover, the simulators report reasonably similar goodputs as experiments does when loaded with medium traffics, e.g. 5 Mbps. The similarity does not hold any more when heavy traffic is loaded. Thus, we do not use more load more than 5 Mbps for multihop cases. Finally, due to the correlations between goodput and jitter, we primarily focus on goodput in the next.



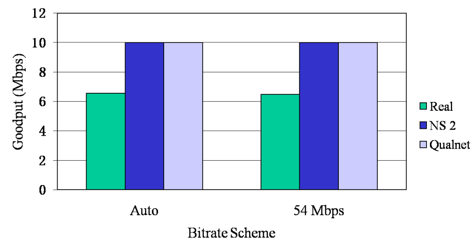
(a) UDP goodput: 40m apart indoor with 5 Mbps CBR load



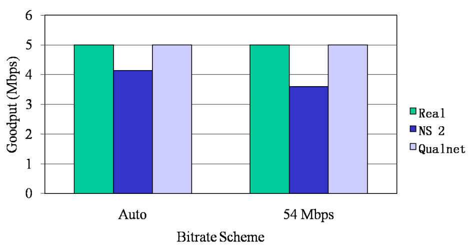
(b) UDP goodput: 40m apart outdoor with 5 Mbps CBR load



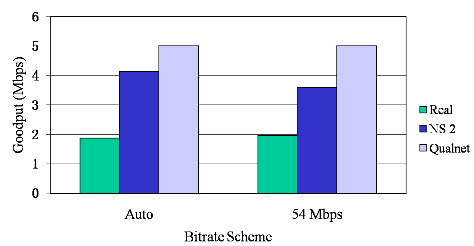
(c) UDP goodput: 40m apart indoor with 10 Mbps CBR load



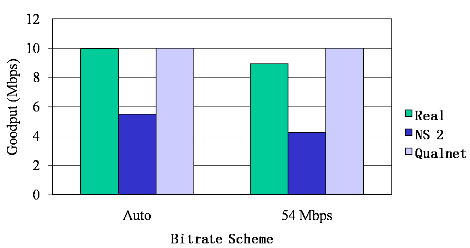
(d) UDP goodput: 40m apart outdoor with 10 Mbps CBR load



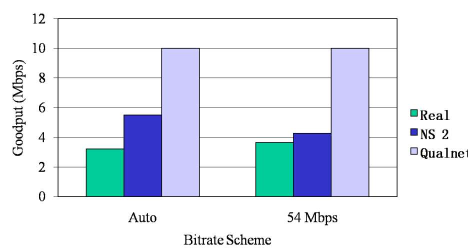
(e) UDP goodput: 80m apart indoor with 5 Mbps CBR load



(f) UDP goodput: 80m apart outdoor with 5 Mbps CBR load



(g) UDP goodput: 80m apart indoor with 10 Mbps CBR load

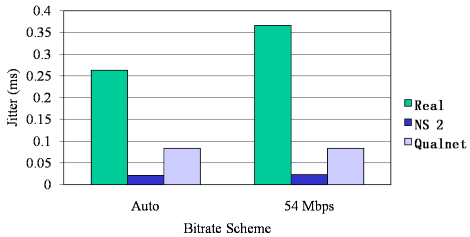


(h) UDP goodput: 80m apart outdoor with 10 Mbps CBR load

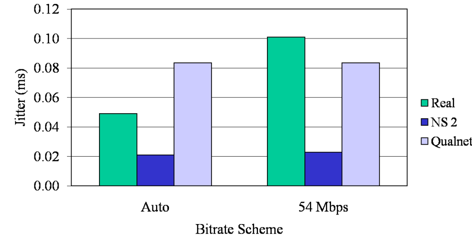
Fig. 3. Point to Point UDP Goodput Comparisons

V. PHY: Beyond Inaccurate Channel Modeling

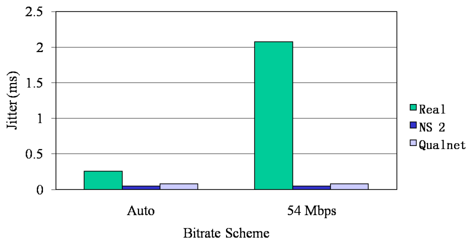
Wireless channel modeling plays a fundamental role in simulating wireless networks. Both NS-2 and QualNet have provided physical models like freespace, two-ray ground, shadowing model, Rayleigh fading or Ricean fading. Additionally, QualNet supports richer libraries than NS-2: Irregular Terrain Modeling (ITM), High speed fading, etc. Most of the channel models follow certain distributions or simply a math function. A few uses statistical models, but require empirical data from users. Radio propagation in reality, however, does not always follow the well-defined distributions. Previous research has pointed it out by showing differences in received signal strength (RSS), probability of symmetric beacon, reception ratio, etc [10]. Our results in Section IV also verify it by comparing TCP and UDP performance. In this section, we further investigate whether there are any other factors affecting PHY layer modeling accuracy, and to what degree the inaccuracy can be compensated through careful calibrations.



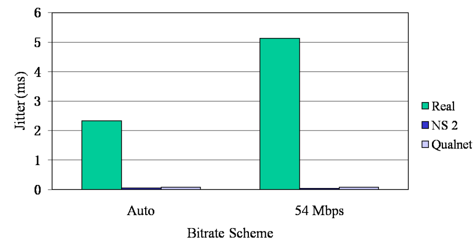
(a) UDP jitter: 40m apart indoor with 5 Mbps CBR load



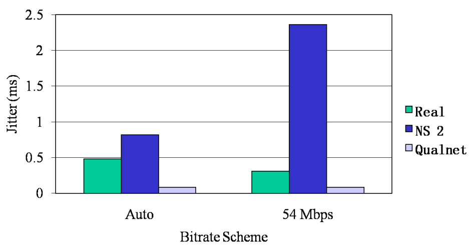
(b) UDP jitter: 40m apart outdoor with 5 Mbps CBR load



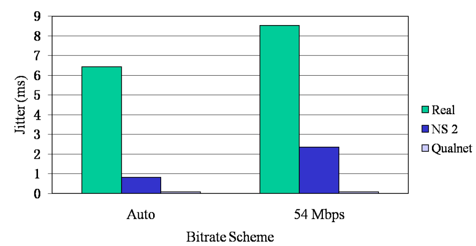
(c) UDP jitter: 40m apart indoor with 10 Mbps CBR load



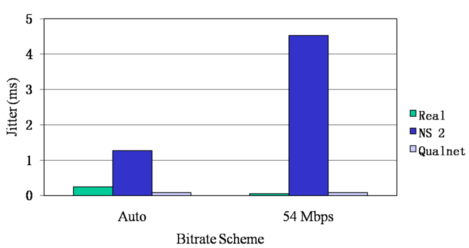
(d) UDP jitter: 40m apart outdoor with 10 Mbps CBR load



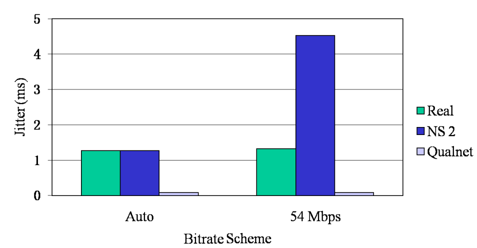
(e) UDP jitter: 80m apart indoor with 5 Mbps CBR load



(f) UDP jitter: 80m apart outdoor with 5 Mbps CBR load



(g) UDP jitter: 80m apart indoor with 10 Mbps CBR load



(h) UDP jitter: 80m apart outdoor with 10 Mbps CBR load

Fig. 4. Point to Point UDP Delay Jitter Comparisons

A. Antenna Diversity

Besides radio propagation problem, we also find another factor that has been missing from simulators is antenna diversity. Antenna diversity is a widely adopted technique in modern 802.11 devices. It takes advantage of the fact that quality of received signal at two antennas can differ dramatically if spaced at least one wavelength apart (12.5 cm at 2.4 GHz and 5.79 cm at 5.18 GHz). Based on this observation, the antenna with the best signal quality will be automatically selected for transmitting or receiving frames. It is worth noting that antenna diversity is different from another antenna technique MIMO (Multiple-Input Multiple-Output). On reception of a frame, MIMO would further utilize advanced signal processing technique to effectively combine received signals from multiple antennas, while antenna diversity simply picks up one of the best. Modern laptops are usually equipped more than one antennas.

Method: We use *wireshark* [19] to record frames from the **Laptops** testbed, in which each node has multiple antennas. We record all frames in the air, in which the received SNR (in dB) and the antenna ID from which it was received have been traced. We then categories the frames by the antenna ID, and plot the PDFs. While Yong et al. exploit two WiFi interfaces to track SNR [16], we are truly using one device with two

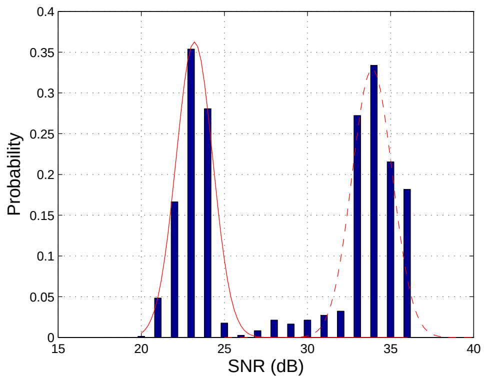


Fig. 5. Example of indoor SNR PDF: the plot shows the distribution of 123,744 frames received by a laptop in 52 minutes.

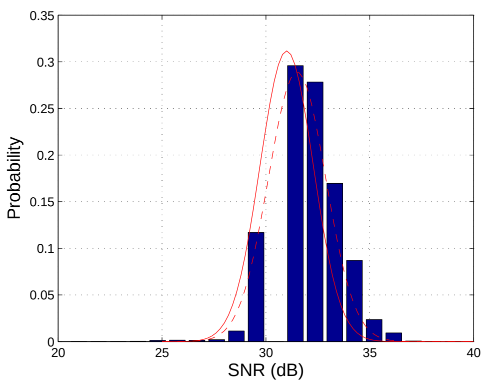


Fig. 6. Example of outdoor SNR PDF: the plot shows the distribution of 135,295 frames received by a laptop in 52 minutes.

antennas.

Results: In our experiments, we plot SNRs of all frames regardless of its antenna, and observe a bi-modal Gaussian distribution. From samples, it is also noticeable that more frames are received from the antenna with higher average SNR, because the one with better reception is more likely to be selected. Fig. 5 and Fig. 6 are two examples of received SNR patterns from the same laptop. Fig. 5 shows a bi-modal Gaussian distribution of received SNR in a multipath indoor environment. However, Fig. 6 shows that the bi-modal distribution becomes less obvious in outdoor, because two antennas then have more balanced reception capability with a strong line-of-sight path, in contrast to non line-of-sight multipath fading indoor. Moreover, SNR drops on the verge of bell-shapes in both cases, and we attribute that to the selection scheme. As the received SNR falls below a threshold, antenna switch is triggered consequently, and lower SNR samples will not be tracked. Received SNRs in simulations, on the other hand, are generated based on predefined functions or distributions. Unfortunately, the bi-modal Gaussian distribution is absent from current libraries in both simulators.

B. Configuring PHY Channel

While various distribution models add sort of randomness, it is well-known that signal is exponentially faded with the increase of propagation distance. In simulations, how fast a signal is faded is determined by path loss. In NS-2, path loss is set as a global parameter to control channel quality, while similar global setting is unseen in QualNet. Instead, an optional interface for user-defined channel is provided, and individual channel between each pair of nodes can be configured. Since our goal is to find out the differences between experiments and the most common scenario in simulators, we set the path loss in NS-2, but keep the default channel setting (path loss is 2.0 and variance is 4.0) in Qualnet. As we believe an average use will not undertake the trouble of configuring individual channel, since the difficulty increases factorially with the number of nodes.

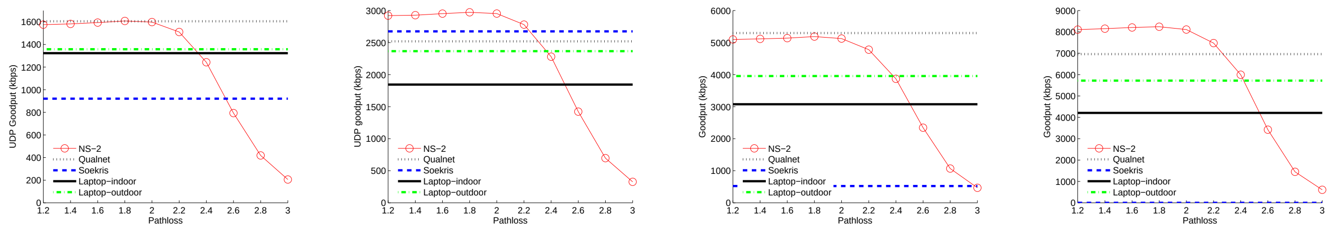
Method: We use both **Soekris** and **Laptops** (indoor and outdoor) form a multihop network topology. A single UDP traffic flow is loaded over the network. Their performances is evaluated at bit rates of 6 Mbps, 12 Mbps, 24 Mbps and 48 Mbps, respectively. Settings in simulators are the same, except we change the path loss in NS2 to see how goodput varies.

Results: Fig. 7 (a) (d) shows the goodput at different bit rates. In each graph, experimental and QualNet results are constant curves, but NS-2 goodput changes as the path loss increases. In experiments, the laptop-based outdoor scenario (laptop-outdoor) outperforms that of the laptop-based indoor scenario (laptop-indoor) due to line-of-sight transmissions. Soekris and the laptop-indoor are both tested in indoor, but they are in the different environment. The former is in an office building while the latter is in a library. Soekris' performance does not seem to be consistent, which can be attributed to its multipath fading environment. Especially, when links between Soekris nodes become extremely vulnerable at high rates, and the goodput goes down to almost zero. In simulations, the QualNet results are almost always better than those in the experiments due to its benign channel setting. The NS-2 results match laptop-indoor results most closely when the path loss equals to 2.4 for indoor and about 2.5 for outdoor. Even if we change the transmission bit rate, this setting still matches very well.

Even if we have pointed out lots of differences between simulators and experiments, the results suggest us that the differences might be mitigated through careful calibrations on path loss. The calibration in practice, however, is difficult. It takes empirical data, and can only apply to certain scenarios. In the following comparisons, we always use these path loss settings in NS-2 if not specified otherwise.

VI. MAC: Flow-level Unfairness Due to Interference

In this section, we focus on MAC layer problems: bit rate, multihop contentions and interference. We discover that the inter-flow interference can result in extreme unfairness at the flow level in reality, while cannot little has been captured in simulations.



(a) Bit rate: 6 Mbps (b) Bit rate: 12 Mbps (c) Bit rate: 24 Mbps (d) Bit rate: 48 Mbps

Fig. 7. Throughput varies with channel path loss at different bit rates

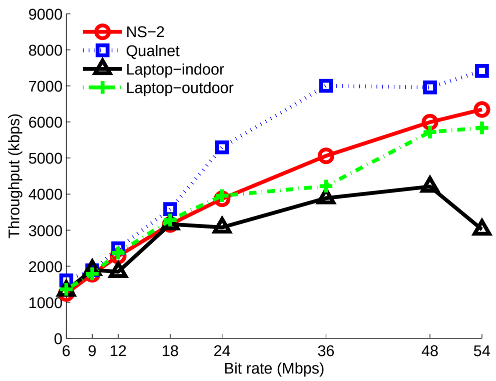


Fig. 8. Throughput varies with bit rates

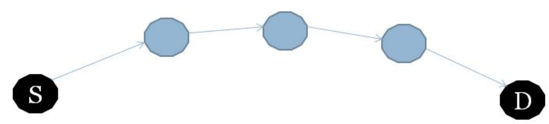
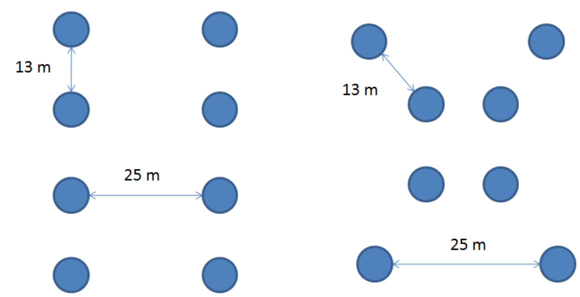


Fig. 15. Linear network topology for a multihop network.



(a) Network topology: H-shape (b) Network topology: X shape indoor

Fig. 16. Flow-level Contention Network Topologies

A. Impact of Transmission Data Rate

Fig. 8 presents goodputs of a multihop flow across all bit rates in IEEE 802.11a. At low rates, almost all results increase linearly. The increase slows down because the high rates make links more vulnerable. NS-2 results (path loss is set to 2.5 for outdoor environments) are in good match with the laptop-outdoor results more closely. The dramatic drop at 54 Mbps in laptop-indoor is captured neither in NS-2 or QualNet.

B. Impact of Multihop

We now investigate the performance difference induced by by multiple hops in mesh network.

Method: Five nodes are closely placed and reside in the same collision domain. We let the sender transmit constant bit rate (CBR) traffic to the nodes two hops away, then to the nodes further away sequentially. The interference comes from the flow itself, we call it "intra-flow interference". Bit rate is set to 6 Mbps to ensure the wireless links remain reliable, which also is the bit rate at which simulations match well with experiments as shown in Fig. 8. We also vary traffic load from 1.5 Kbps, 2 Kbps, to 2.5 Kbps. Again this is far below the stable traffic load threshold 5 Mbps in Section IV. Fig. 8 shows five Soekris nodes forming a linear topology in an office environment.

Results: The results are shown in Fig. 9 through Fig. 14. In each figure, we show how goodput or delay vary with the number of hops. When the number of hops is 2 or 3, the results in NS-2, QualNet and the testbed are close. However, they diverge as the number of hops or traffic load increases. To see why, we need to understand where packets get lost. Packet loss in a mesh network can be attributed to network-induced factors such as interference and wireless channel, as well as to node-induced factors like processor queue in the operating system [22]. If packets can not be proceeded at the wireless line speed, they will be dropped even if correctly received. In a resource constrained embedded system, overwhelming traffic is very likely result in queue overflow. Intra-flow interference intensifies the contentions and more packets are queuing up, thus a good portion of it is dropped away. This explains why Soekris board experience low performance, and none of these happens in simulators. Besides, optimistic channel settings in Qualnet can explain its consistent high goodput and low delay.

C. Flow Unfairness Under Interference

Only intra-flow interference is taken into considerations in previous experiments. In this subsection, we investigate how inter-flow interference affects the fidelity of simulators. Strictly speaking, interference modeling is a physical layer issue. But since our discussions involve flow level medium contentions, we discuss it at the MAC layer. Previous work

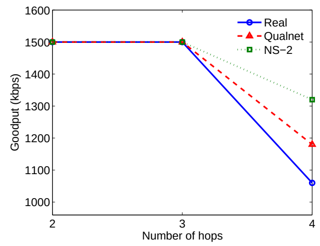


Fig. 9. Traffic load 1500k: goodput

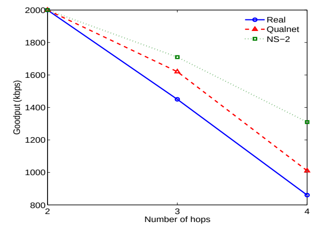


Fig. 10. Traffic load 2000k: goodput

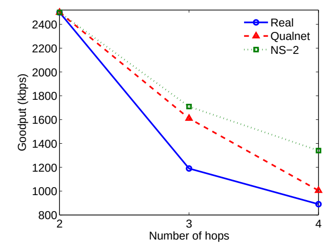


Fig. 11. Traffic load 2500k: goodput

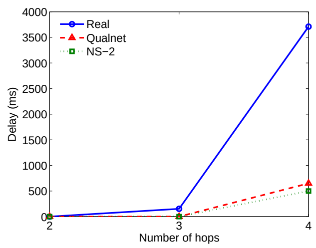


Fig. 12. Traffic load 1500k: delay

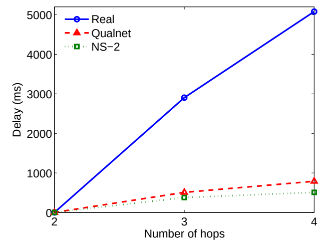


Fig. 13. Traffic load 2000k: delay

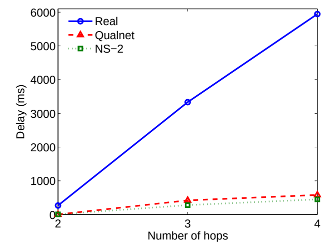


Fig. 14. Traffic load 2500k: delay

studied inefficiency of resolving collisions in simulators [9]. Neither NS-2 or QualNet follows the physical layer capture technique used in the real-world. The latest models in both simulators use signal to interference and noise ratio (SINR) to solve physical layer capture issues.

Method: We are interested in interference impacts on flow level performance. To this end, we construct a scenario where two flows consisted of multihop nodes transmitting simultaneously, and thus have to compete for wireless medium. Impacts of inter-flow interference are then quantified. We use the goodput of a single flow without any interruptions of interference as a benchmark, and measure results in interfered scenarios. The goodput degradation then can be calculated, which is the ratio of the latter over the former.

We consider the following factors in this comparison: environment, topology, and bitrate. The experiments are conducted on two three-hop flows in an open library hall, and a grass field, respectively. In each flow, laptops are set about 13 meters apart to ensure reliable transmissions, and two flows are not left far away to generate interference to each other, as shown in Fig. 16. In addition, X-Shape and H-Shape (flows in parallel) are constructed. The bitrats we used ranges from 6 Mbps to 36 Mbps, at which the network can maintain good conductivities. The PHY parameters in simulators are configured to best match environment.

Results: When no interference is presented, each flow reports similar goodputs in experiments. But an extreme unfairness between two flows are observed when they transmit simultaneously, as shown in Fig. 17-(a) (d). Particularly, Flow 1 constantly grabs more wireless medium than Flow 2. In sharp contrast, flows can compete fairly in simulations. Due to the space limit, we present only H-shape result from simulators, and X-shape result is highly similar.

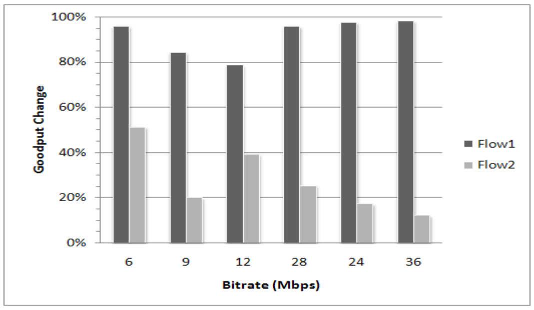
The aggregated goodput of two flows can be higher than that of any of single flows because of the PHY layer capture effects. Even under interference, one receiver can still decode a frame if the signal to interference ratio is not too low. The facilitates simultaneous transmissions, and thereafter improve aggregated goodput.

In the outdoor case, the unfairness is even more severe than in the indoor case, because the interference is even stronger for no obstacles block the way radios propagate. The unfairness in experiments exists regardless to network topologies or bitrates.

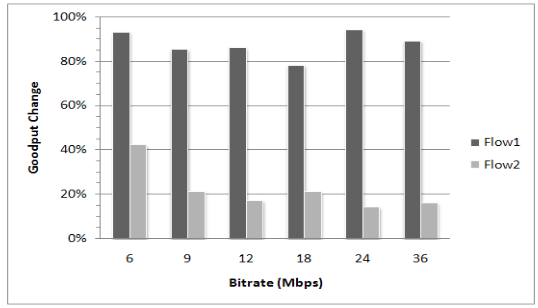
Analysis: The hardware processing speed discussed in subsection VI-B is not a concern here since we do not overload the flows. The performance difference can be largely attributes to the difference in transceivers' capability such as transmitting power, receiving sensitivity, and career sensing threshold. The factor that one flow consistently has lower goodput than the other indicates some of laptops in this flow have weaker transceiving capability. When the flow comes into competition with the other flow, the weakness is magnified. The weak flow delivers less packets, thus introduces less interference than the other. The strong flow is therefore less impacted by the interference. Similar results are also observed in [21], but only single hop result is presented. On the other hand, the simulators assume flows are identical if the transmit power, the distance, and the channel quality between nodes are the same. Thus in our simulations two flows contend equally.

VII. Routing: Route Stability

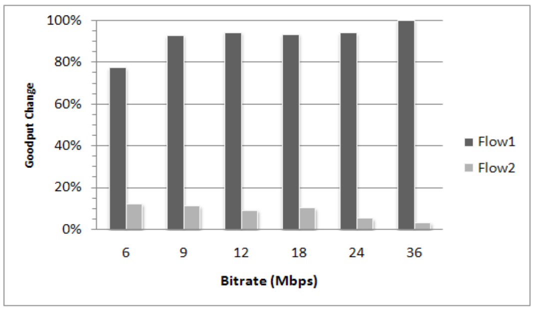
In this section, we quantify the differences in routing stability between simulators and real-world networks. The meaning is two-folded. First, it can provide a better understanding on simulating wireless routing protocols. Most of existing routing protocols do not consider the routing stability. Routing instability, however, can lead to routing pathologies like packet reordering. Second, routing stability affects simulation results of wireless mesh network management, planning, and radio placement.



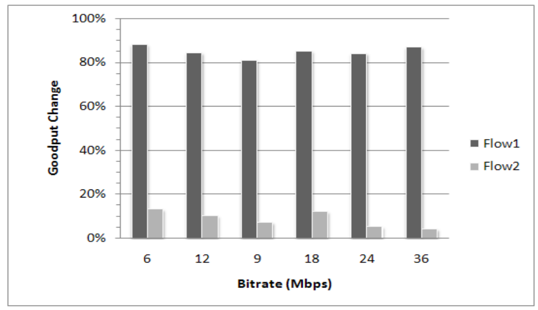
(a) Laptop: two flows compete in H shape indoor



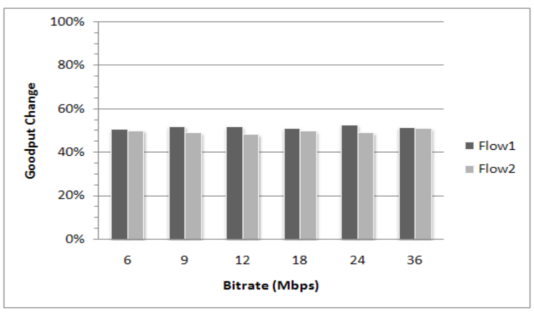
(b) Laptop: two flows compete in X shape indoor



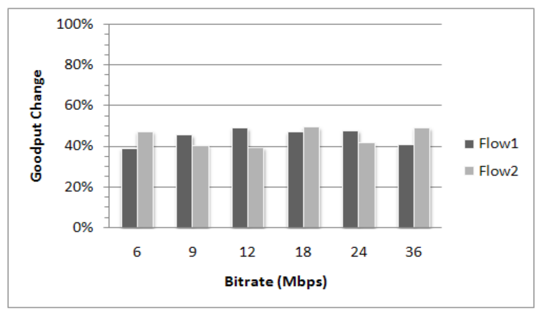
(c) Laptop: two flows compete in H shape outdoor



(d) Laptop: two flows compete in X shape outdoor



(e) NS-2: two flows compete in parallel



(f) Qualnet: two flows compete in parallel

Fig. 17. Flow-level Goodput Comparisons under Interference

A. Network Settings

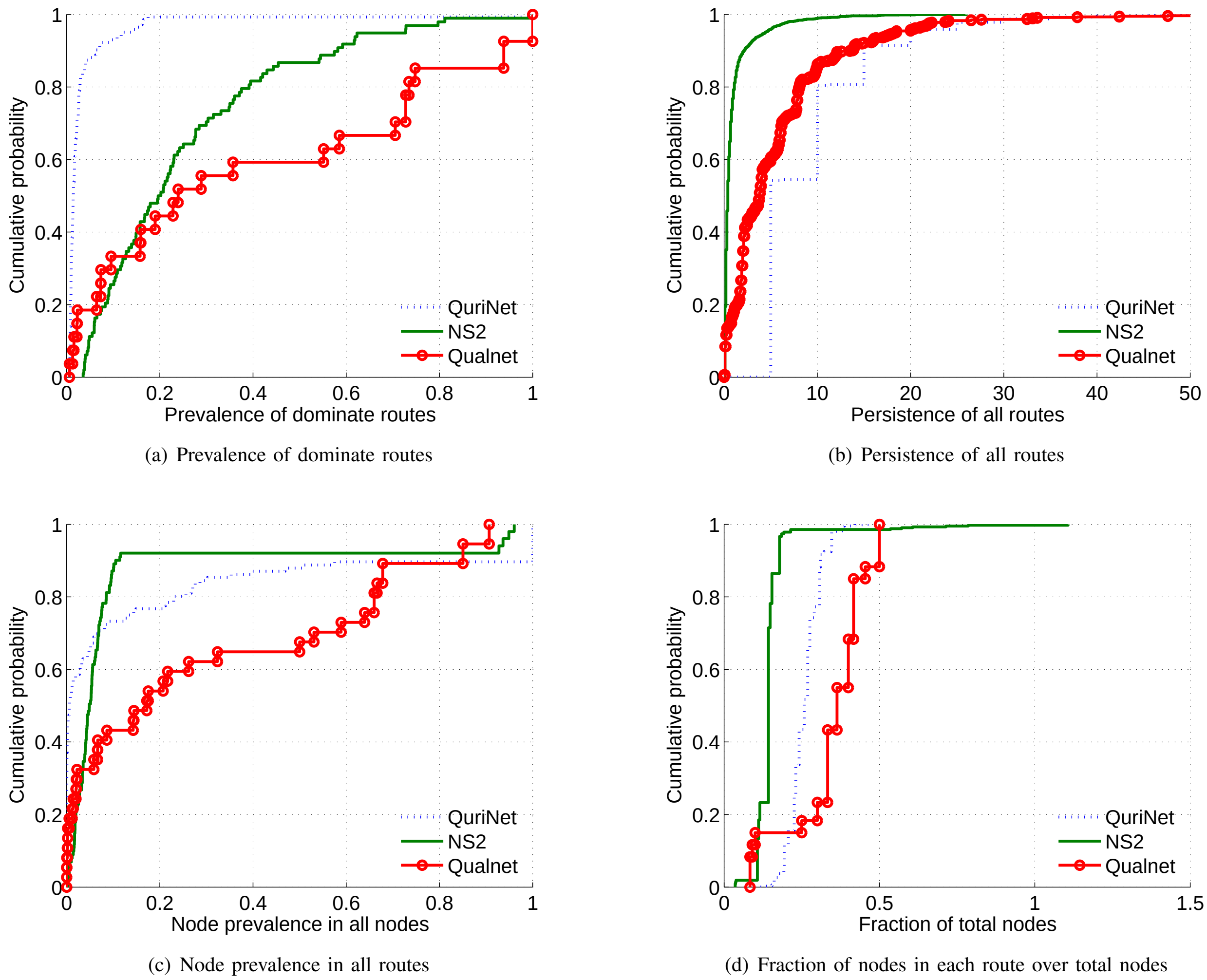
OLSR is used as the target evaluating routing protocol. We first pick up thirty source-destination pairs across the network. In simulation, we track routing table changes every 0.1 seconds for a given source-destination pair. In experiments on QuRiNet, we probe the route between a source-destination pair every 5 seconds to avoid overloading the network. The low probing rate may cause some misses of route changes, but does not affect the trend of discrepancy.

QuRiNet is a heterogeneous network, and nodes are equipped with omnidirectional and/or directional antennas. Nodes are configured to use different transmit powers. Instead of trying to reconstruct the exact same network in the simulators, we replicate the network's connectivity. From an arbitrary node in either network or simulators, it has similar reachability to all other nodes. We argue this should be enough for our purpose because OLSR use expected transmission times (ETX) as the routing metric and the background traffic in QuRiNet is low enough that it does not affect ETX. Therefore, the dominating factor for routing stability is the channel quality and channel stability. We collect four-days data from QuRiNet and more than 2 GB data from simulations.

B. Route Prevalence and Persistence

Method: Two metrics are used to analyze routing stabilities, *prevalence* and *persistence*. The *prevalence* is normalized occurrence frequency: occurrence of a route over that of the *dominant route*. The *persistence* means the duration in seconds for which a route lasts before it changes.

For a given source-destination pair, we analyze its routing prevalence. The *dominate route* is defined as the one being observed most often over all records. If the dominate route occurs n seconds in total, and another route lasts m seconds cumulatively, we define the *prevalence* as a ratio $p = \frac{m}{n}$, and the range is (0, 1]. For example, a route's prevalence is $\frac{1}{50}$ means that the dominant route appears 50 times more often than this route. Intuitively, the prevalence can also be defined



(a) Prevalence of dominate routes

(b) Persistence of all routes

(c) Node prevalence in all routes

(d) Fraction of nodes in each route over total nodes

Fig. 18. Impacts on Routing Stability and Prevalence

as a cumulative duration over total time. However, we've found in that case the scale will be very small even for the dominate route. Alternatively, this relative prevalence can capture the relationship among all routes.

Results: Fig. 18(a) shows the cumulative distribution of the route prevalence over all transmission pairs. For clarity, routes that occur three orders less than the dominate route have been omitted. We observe that the dominate route occurs much more often than the rest in QuRiNet. Most of the routes are merely occurring 20% as often or less as the most popular one. The dominate route often turns out to be the one with the minimum number of hops. This trend is less obvious in both simulators, where the routes' occurrence frequency are distributed more evenly.

The route persistence analysis is shown in Fig. 18(b). Most of the routes are short lived. Similar to QuRiNet, the median persistence of QualNet is 5 seconds, while NS-2 has 80% of its routes last less than 1 second. The frequent route changes is possible even if OLSR HELLO messages are exchanged every 2 seconds among the nodes because message exchange is not synchronized, and every message exchange may induce a route change. In QuRiNet, since each route is recorded every 5 second, some discreteness in time is observed.

C. Spatial Prevalence and Spatial Distribution

In this subsection, we will compare the spatial prevalence.

Method: Here the *spatial prevalence* is defined as follows. Every time a node appears in a route, including source and destination, we count its occurrence once. The *spatial prevalence* is the number of cumulative occurrences of one node over the number of all records. The source and the destination will appear in every route between them, and their spatial prevalence is 1. This metric can reveal how often a node appear in all routes. In another words, it shows the routes spatial diversity.

We are also interested in the spatial distribution of routes. We calculate the length of routes in all records, then divide the length by the total number of nodes in the network. We want to use this ratio to measure the distribution of the route length. Note for both simulators and QuRiNet, the total number of nodes in the network is the same, thus the metric is comparable.

Results: Fig. 18(c) shows the cumulative distribution of node spatial prevalence, where the x-axis indicates the fraction of all routes, and the y-axis is the cumulative probability of nodes involved. NS-2 exhibits the most spatial diversity, with approximately 17% of routes covering 90% of nodes; followed by QuRiNet with 20% of routes covering 78% of nodes. This distribution almost matches the power law. NS-2 and QuRiNet are similar in terms of route spatial diversity. The routes in QualNet are spatially concentrated with 60% of routes going through 70% of nodes. In other words, the similarity in QualNet is higher than NS-2 and QuRiNet. This is due to the better channel quality and relative stability of channel model in QualNet simulation.

Fig. 18(d) demonstrates the cumulative distribution results of route length. The medians for NS-2, QuRiNet, and Qualnet are 15%, 25%, and 38% respectively. This indicates that the number of hops in each route in QualNet is generally more than the other two. Together with the results of route spatial prevalence in Fig. 18(a), we have found that even if the length of routes in QualNet is longer, they involve less number of nodes. This strengthens our conclusion about high similarity among routes in QualNet.

VIII. Conclusion

In this work, we performed a study on the discrepancies between simulators and real-world testbeds. While the discrepancies are broadly expected, this paper provides a detailed, systematic, and comprehensive quantification. We perform the simulations on NS-2 and QualNet which are two of the most commonly used wireless network simulators. We then have compared the results with those from experiments on three different testbeds, single antenna embedded system, laptops with modern Wi-Fi devices, and QuRiNet (an outdoor Wi-Fi network deployed in a natural reserve). We examine PHY, MAC and IP layers, and study factors including antenna diversity, path loss, multihop, transmission rate, interference and routing stability.

We summarize the discrepancies between simulations and testbed as follows. In addition of inaccurate channel modeling, simulations do not model the antenna diversity either. This will make it more difficult to replicate complicated scenario like indoor. However, since the dominating factor in channel modeling is path loss, simulations can still have a good match with experiments in simple environments like outdoor with line-of-sight transmissions, even if multiple different bit rates are used. Also, transceiving capability in simulation is just a simple function of distance, while in reality, differences in hardware can result in extreme flow level unfairness in interfered scenarios. Last, routes in simulators are less stable and persistent than that in a real network. In this paper, we give a clear picture about the major differences between simulators and testbeds. It provides a reference for researchers to weigh pros and cons in choosing between simulations or testbeds in doing wireless network studies.

References

[1] John Bicket, Daniel Aguayo, Sanjit Biswas, and Robert Morris, "Architecture and evaluation of an unplanned 802.11b mesh network," in *Proceedings of the 11th ACM International Conference on Mobile Computing and Networking (MobiCom '05)*, Cologne, Germany, August 2005.

[2] Y. Sasson D. Cavin and A. Schiper, "On the accuracy of manet simulators. Technical report," Tech. Rep. Technical report, Distributed Systems Laboratory, Oct 2002.

[3] An Chan and Soung Chang Liew, "Performance of voip over multiple co-located ieee 802.11 wireless lans," *IEEE Transactions on Mobile Computing*, vol. 8, no. 8, pp. 1063–1076, 2009.

[4] Chi Pan Chan, Soung Chang Liew, and An Chan, "Many-to-one throughput capacity of ieee 802.11 multihop wireless networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 4, pp. 514–527, 2009.

[5] Alix L.H. Chow, Hao Yang, Cathy H. Xia, Minkyong Kim, Zhen Liu, and Hui Lei, "Ems: Encoded multipath streaming for real-time live streaming applications," in *Network Protocols, 2009. ICNP 2009. 17th IEEE International Conference on*, Oct. 2009, pp. 233–243.

[6] VINT group, "UCB/LBNL/VINT Network Simulator – ns (Version 2)," *http://mash.cs.berkeley.edu/ns*.

[7] Fang Hao, Murali Kodialam, and T. V. Lakshman, "On-line detection of real time multimedia traffic," in *Network Protocols, 2009. ICNP 2009. 17th IEEE International Conference on*, Oct. 2009, pp. 223–232.

[8] S. Ivanov, A. Herms, and G. Lukas, "Experimental validation of the ns-2 wireless model using simulation, emulation, and real network," in *Proceedings of the 4th Workshop on Mobile Ad-Hoc Networks*, 2007.

[9] Andrzej Kochut, Arunchandar Vasan, A. Udaya Shankar, and Ashok Agrawala, "Sniffing out the correct physical layer capture model in 802.11b," in *ICNP '04: Proceedings of the 12th IEEE International Conference on Network Protocols*, 2004.

[10] D. Kotz, C. Newport, R. Gray, J. Liu, Y. Yuan, and C. Elliott, "Experimental evaluation of wireless simulation assumptions.," in *In Proc. ACM/IEEE Intl Symp on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2004.

[11] Henrik Lundgren, Krishna Ramachandran, Elizabeth Belding-Royer, Kevin Almeroth, Michael Benny, Andrew Hewatt, Alexander Touma, and Amit Jardosh, "Experiences from the Design, Deployment, and Usage of the UCSB MeshNet Testbed.," *IEEE Wireless Communications Magazine*, 2006.

[12] Ratul Mahajan, Maya Rodrig, David Wetherall, and John Zahorjan, "Analyzing the mac-level behavior of wireless networks in the wild," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 4, pp. 75–86, 2006.

[13] Vivek Mhatre, Henrik Lundgren, Francois Baccelli, and Christophe Diot, "Joint mac-aware routing and load balancing in mesh networks," in *CoNEXT '07: Proceedings of the 2007 ACM CoNEXT conference*, New York, NY, USA, 2007, pp. 1–12, ACM.

[14] S. Nanda, R. Walton, J. Ketchum, M. Wallace, and S. Howard, "A high-performance MIMO OFDM wireless LAN," *IEEE Communications Magazine*, vol. 43, no. 2, pp. 101–109, Feb 2005.

[15] D. Raychaudhuri, I. Seskar, M. Ott, S. Ganu, K. Ramachandran, H. Kremo, R. Siracusa, H. Liu, and M. . Singh, "Overview of the orbit radio grid testbed for evaluation of next-generation wireless network protocols.," in *Proc. of IEEE Wireless Communications and Networking Conference*, 2005.

[16] Yong Sheng, Keren Tan, Guanling Chen, David Kotz, and Andrew Campbell, "Detecting 802.11 MAC Spoofing Using Received Signal Strength.," in *Proc. IEEE INFOCOM*, 2008.

[17] "Soekris engineering," http://www.soekris.com/.

[18] Scalable Network Technologies, "QualNet," *http://www.scalable-networks.com/products/qualnet/*.

[19] Wireshark, "Wireshark," *http://www.wireshark.org/*.

[20] Daniel Wu, Dhruv Gupta, Stephanie Liese, and Prasant Mohapatra, "Qurinet: quail ridge natural reserve wireless mesh network," in *WiNTECH '06: Proceedings of the 1st international workshop on Wireless network testbeds, experimental evaluation & characterization*, New York, NY, USA, 2006, pp. 109–110, ACM.

[21] Lei Yang, Ben Y. Zhao, and Haitao Zheng, "The spaces between us: setting and maintaining boundaries in wireless spectrum access," in *MobiCom '10: Proceedings of the sixteenth annual international conference on Mobile computing and networking*, New York, NY, USA, 2010, pp. 37–48, ACM.

[22] Hua Yu, Daniel Wu, and Prasant Mohapatra, “Experimental anatomy of packet losses in wireless mesh networks,” in *Proceedings of IEEE SECON*, 2009.

[23] Xiang Zeng, Rajive Bagrodia, and Mario Gerla, “Glomosim: a library for parallel simulation of large-scale wireless networks,” *SIGSIM Simul. Dig.*, vol. 28, no. 1, pp. 154–161, 1998.