RAIN: A Reliable Wireless Network Architecture

Chaegwon Lim[†‡], Haiyun Luo[†], and Chong-Ho Choi[‡]
[†]Department of Computer Science, University of Illinois at Urbana-Champaign, USA
[‡]School of Electrical Engineering and Computer Science, Seoul National University, Korea
chaegwon@csl.snu.ac.kr, haiyun@cs.uiuc.edu, chchoi@csl.snu.ac.kr

***Abstract*— Despite years of research and development, pioneering deployments of multihop wireless networks have not proven successful. The performance of routing and transport is often unstable due to contention-induced packet losses, especially when the network is large and the offered load is high. In this paper we propose RAIN, a *reliable wireless network* architecture for large-scale multihop wireless networks. A RAIN network enforces *contention control* by limiting the queue length at intermediate wireless routers to the minimum. To keep the queue short a RAIN network enforces *congestion control* through in-network implicit back-pressure. RAIN congestion control is built on wireless datalink layer mechanisms, e.g., mandatory per-frame acknowledgement and inter-frame backoff in popular CSMA/CA wireless transceivers, therefore very efficient and effective compared with those defined at the network or transport layer for the wired Internet. As a result of the built-in contention and congestion control, RAIN presents the end hosts a highly reliable network service model, even more reliable than that of the wired Internet. The end hosts only need to deal with packet losses due to router or routing failures. Therefore, the transport protocol can be significantly *simplified*. This is in stark contrast to the existing approach of adding more and more complexity to adapt TCP for multihop wireless networks. We propose the details of RAIN datalink layer protocol, and a simple transport protocol at the end hosts. Performance evaluation through intensive simulations shows that RAIN improves the throughput by up to 92% and fairness by up to 48%, with packet losses due to contention and congestion significantly reduced.**

I. Introduction

Multihop wireless networks first emerged in response to the demand for instant networking in the forms of mobile ad hoc networks (MANETs). It recently developed into wireless mesh networks to offer high-speed broadband Internet connectivity. However, despite years' research and development in MANETs and wireless mesh, pioneering deployments of multihop wireless testbeds [1], [2], [3], [12], [38] have encountered serious stability problem [4]. Routing and transport protocols, designed for the wired, first-/last-hop wireless Internet, are often extremely unstable and unpredictable in a multihop wireless network as the network grows, especially when the offered load is high. Reports of excessive packet drops [23], [46], unfair channel bandwidth sharing and starvation [25], [26], and extremely volatile path properties [9], [40], [46] have been frequently cited to question the feasibility of multihop wireless networks.

One major cause of these problems is that wireless transmissions are broadcast[1] in nature. They *contend* with each other even between packet transmissions of the same flow. Different from the Internet that are usually connected by point-to-point well-insulated wires, in wireless networks the transmission between a sender and a receiver relies on the medium access control (MAC) to resolve the contention for the shared wireless channel in a neighborhood with variable traffic demands. However, wireless transmissions interfere with each other in a range that is usually longer than the transmission/receiving range (or "one hop" direct communication range) and unknown *a priori*. A wireless MAC does not have the explicit information regarding those contending nodes that are interfering from beyond the direct communication range. It is therefore very challenging, if not impossible, for a wireless MAC to coordinate interfering nodes with which it cannot directly communicate [14], [13]. Furthermore, wireless interference comes and goes as interfering nodes start and finish transmitting a packet. This fine time granularity renders end-to-end mechanisms, e.g., TCP wireless variants [22], [27], [33], [41], [43], ineffective in contention control since end-to-end mechanisms operate at a significantly coarser time granularity of a round trip time. As a result, the contention becomes the primary cause to packet losses in a multihop wireless network [9], [23], [44], [46], as compared with the wired Internet where packet losses are mostly caused by network congestion and router buffer overflow.

The above analysis suggests that effective contention control must be in-network for timeliness, and go beyond the direct one-hop neighborhood of individual wireless routers. To this end, we start our design with a novel approach that bounds the contention level in a multihop wireless network through buffer management. Since all wireless nodes with non-empty packet queues will contend for the channel, the contention level is a function of the number of nodes with backlogged queues. Therefore, if we can reduce the number of backlogged queues in the network we can keep the contention level low. In specific, we limit the maximum length of the transit traffic queue[2] to the minimum possible, i.e., close to one packet. Because each intermediate wireless router only holds a very small number of packets, typically one packet only, the contention between upstream and downstream wireless

[1]Directional antenna may reduce the broadcast area, but cannot completely eliminate interference due to side lobes. It also introduces new sources of channel contentions such as carrier sense deafness [15].

[2]Note that the source or destination node will maintain extra queue for packets generated by or destined for the node itself.

1-4244-0593-9/06/$20.00 ©2006 IEEE

routers along the path of a packet stream can be minimized. To further prevent multiple flows from overwhelming the short queues at intermediate routers, we enforce congestion control through in-network implicit back-pressure. We achieve efficient implementation of the contention and congestion control by exploiting the unique features of the wireless datalink, i.e., per-frame acknowledgement and inter-frame backoff that are mandatory in popular CSMA/CA wireless transceivers [29]. Such back-pressure signaling implicitly extends the scope of the contention control at an individual wireless router's MAC layer all the way to the packet sources.

The potential impact of our design is profound. Today's Internet routers are equipped with more and more buffer to mitigate buffer overflow problem caused by traffic bursts. On the contrary to this common wisdom, we show that in multihop wireless networks the length of the transit traffic queue should be controlled to the minimum for effective contention and congestion control. Furthermore, the two buffers that are usually maintained separately at the network layer and datalink layer in the wired Internet architecture should be coalesced into one, shared between routing and medium access control. The end result is a *r*eli*a*ble w*i*reless *n*etwork (RAIN) that is free of both contention induced packet losses due to wireless interference and congestion induced packet losses due to buffer overflow, in which sense even more reliable than a wired network. Packet losses in a RAIN network will only happen when an intermediate wireless router or the routing fails. Therefore, the transport layer that guarantees end-to-end reliability can be significantly *simplified*, as opposed to the trend of patching TCP with more and more complex wireless extensions [22], [23], [35], [36], [41].

Our contributions in this paper are three-fold. First, we propose a novel reliable network architecture for large-scale multihop wireless networks. RAIN re-defines the service models for the datalink layer and transport layer, so as to better fit into the interference-prone multihop wireless networks. Second, we propose new datalink and transport protocols to demonstrate the feasibility of the RAIN architecture. Our protocol designs leverage the unique characteristics of popular CSMA/CA wireless transceivers, with a counter-intuitive queue management strategy. As a result, our protocols realize fine time granularity contention and congestion control. Finally, we have built the RAIN architecture and companying protocols in *ns-2*. Extensive simulation results show that RAIN's contention and congestion controlled datalink layer improves the throughput of individual unregulated UDP traffic by up to 92%, and improves the fairness among competing UDP traffic by up to 25%. Together with our simplified transport protocol, RAIN achieves an average 61% gain in aggregate throughput and an average 48% gain in fairness in highly mobile ad hoc networks.

The rest of this paper is organized as follows. We first compare with related work in Section II. We then present the RAIN network architecture and service models in Section III. The following two sections describe the details of RAIN datalink (Section IV) and RAIN transport (Section V) protocols. We evaluate our design in Section VI and finally conclude with future work in Section VII.

II. Related Work

IEEE 802.11 [29] defines the physical and datalink layers of the most popular wireless transceivers. The 802.11 MAC was designed for infrastructure mode, where nodes in a Basic Service Set (cell) can be *at most two hops* away from each other and *communicate only with the centralized access point*. 802.11 handles hidden/exposed senders well but does not address the hidden/exposed receiver problem [13], the primary cause to packet losses in a multihop wireless network as shown in [9], [23], [46]. MACAW [11] and FAMA [24] are early proposals on CSMA/CA wireless MAC. They handle hidden/exposed sender problem even better than 802.11, but leave the hidden/exposed receiver problem open.

BAPU [10] addresses channel contention due to hidden/exposed receiver problem, but it requires two channels and uses a dedicated control channel for signaling. Recently several multi-channel variations of 802.11 medium access control are also proposed, e.g., SSCH [7] and MMAC [42], but their goal is to increase network capacity, not to handle channel contention. The design of single-transceiver multiple channel involves extra latency for channel synchronization and requires time-synchronization hardware [7]. Furthermore, multiple unlicensed 802.11 channels are not always available as in Japan.

With high-quality directional antennas [8], [16] wireless radiation energy can be focused narrowly in a beam that behaves similar to a wired peer-to-peer link. However, high-quality directional antennas are not completely immune from interference due to the side lobes of the main beam. Furthermore, high-quality directional antenna are very expensive and require careful calibration. It therefore compromises the cost-effectiveness and instant deployability that distinguish multihop wireless networks from the wired alternatives.

Holland et al. [27] investigate the effect of mobility-induced link breakage of wireless ad hoc networks upon TCP performance. When packets are lost due to node mobility, congestion control mechanisms should not be applied. Studies in [22], [33], [41], [43] propose intelligent congestion detection in improving TCP in MANETs. In particular, [22], [41] use end-to-end measurements to differentiate packet losses due to wireless channel errors from those due to network congestion or buffer overflow. In [33] the network conditions are detected by ICMP (destination unreachable) and ECN messages based on the feedback of the intermediate nodes. In [43], Sundaresan et al. use the intermediate node's feedback to decide the sending and retransmission rate. Since congestion control is no longer a required function of the transport layer in our RAIN architecture, our transport protocol is much simpler than the above proposals.

Xu et al. [45] apply RED queue management in a local network neighborhood based on ideal and busy time slot measurements. We target not only fairness, but also high throughput that scales multihop wireless network design.

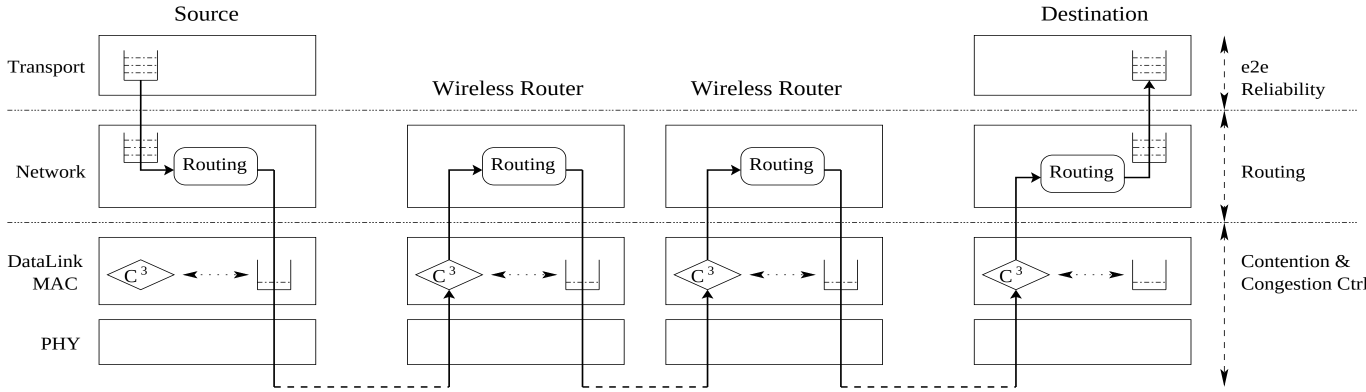

Fig. 1. RAIN architecture. "→" represents the flow of packets from the source to the destination. "C^3" represents "Contention and Congestion Controller". It sends out contention and congestion control signal based on the length of the DataLink layer queue. The network layer does *not* buffer transit traffic.

Congestion control based on hop-by-hop back-pressure has been proposed in ATM [30], switched Ethernet [28], and wireless networks [47]. We apply the technique to address the more severe problem of wireless network contention control, which is usually the primary cause of packet losses in a large multihop wireless network but often missed in the existing designs.

Finally, Ramanathan et al. [37] propose a new MANET architecture. Its main goal is to streamline packet forwarding at intermediate wireless nodes using the future full-duplex, relay-oriented wireless *physical layer* without having the upper MAC or network layer involved, similar to circuit switching in synchronous networks. RAIN is instead motivated by contention control for packet-switched asynchronous wireless networks with off-the-shelf half-duplex wireless physical layers. Important architectural functions, i.e., contention and congestion control, are implemented based on wireless datalink layer mechanisms that are usually mandatory in off-the-shelf CSMA/CA wireless transceivers.

III. RAIN ARCHITECTURE

We first present the overview of the RAIN architecture. The layers and packet flows are presented in Figure 1. We omit the application layer that sits on top of the transport layer. RAIN preserves the layering architecture of the Internet, but redefines the service models for the datalink layer, the network layer, and the transport layer. The most notable difference is that in RAIN the contention control is added and the congestion control is pushed down the stack to the datalink layer. The network layer interacts with the upper and lower layers the same way as that in the wired Internet. It populates the routing table and provides the default route lookup. The transport layer may detect and recover packets that are lost due to router or routing failures, in case perfect reliability is desired.

RAIN re-organizes the packet buffers in the protocol stack. For consistent contention and congestion control at the datalink layer, RAIN coalesces the network layer transit traffic buffering into the datalink layer. Note that the network layer still maintains a buffer, but only for packets generated by or destined for the node. That is, the network layer buffer can be non-empty only at the source or the destination nodes.

Whenever a wireless router receives a frame, the contention and congestion controller (C^3) at the datalink layer first checks the length of the transit traffic queue. If the queue length is above a small threshold, e.g., one packet, a contention and congestion control signal is forwarded, e.g., piggybacked to the datalink layer acknowledgement, to the upstream node before the frame is forwarded to the network layer for routing. RAIN intentionally stores a minimum number of transit packets at a wireless router for two reasons. One is that with a small number of packets in the queue an upstream wireless router will quickly yield from the channel contention after it forwards the packet downstream, therefore keeping the packet forwarding at upstream/downtream wireless routers smooth. The other reason is that with light buffering the contention and congestion control signal will quickly back-pressure to the sources, where the applications' next `send` system call will either be blocked or return with an error. Note that the minimum transit traffic queue length can be as small as one packet with popular IEEE 802.11 CSMA/CA wireless transceivers. It does not compromise the utilization of the IEEE 802.11 wireless channel capacity, because the pipe size of the wireless link is exactly one packet regardless of the wireless channel bandwidth due to the mandatory MAC-layer per-frame acknowledgement for unicast traffic.

To enable contention and congestion control summarized above, RAIN defines new horizontal interfaces across the datalink layers at intermediate wireless routers. These new interfaces are used for the neighboring wireless routers to advertise the queue status among one another. Note that the design of the RAIN architecture does not violate the elite end-to-end argument for system design [39] in general or contradict the Internet design philosophy [17] in specific. RAIN pushes the congestion control down to the datalink layer, where the congestion control is synergistically combined with contention control. The goal is to elevate the network reliability to the level on par with the wired Internet, e.g., less than one percent packet loss ratio [17]. The RAIN datalink

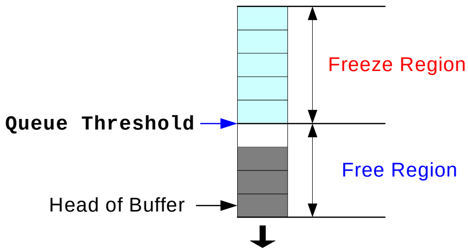

Fig. 2. SAFE transit traffi c buffer.

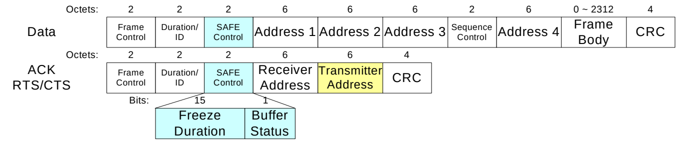

Fig. 3. SAFE frame format.

layer does not guarantee reliability or differentiate packets coming from different network layer flows, therefore does not raise the scalability concern given that the number of end-to-end flows can be large. RAIN also fully respects the layering of the network architecture and does not take the liberty of cross-layer design for granted [32].

In the following two sections we describe exemplary instantiations of those RAIN interfaces described above, in the specific context of multihop wireless networks based on the CSMA/CA wireless transceivers, e.g., IEEE 802.11 devices on which the majority of existing multihop wireless network testbeds are built [1], [2], [3], [12], [38]. We describe packet header specifications and the protocol details of a new datalink layer in Section IV and a new transport layer in Section V.

IV. RAIN DataLink

To demonstrate the feasibility of the RAIN architecture, we describe in this section how RAIN's contention and congestion control can be implemented using *simple* datalink layer mechanisms. Our proposal, called small buffer and adaptive freeze (SAFE), is based on the IEEE 802.11 standard. We assume that a wireless transceiver performs physical and virtual carrier sense before transmitting a frame. We also assume that per-frame acknowledgement be enforced. Due to the lack of space we omit further introduction to CSMA/CA and IEEE 802.11. Interested readers are referred to [29] for complete details.

RAIN contention and congestion control is all centered around the transit traffic buffer management. We therefore start from the structure of transit traffic buffer. We then describe the SAFE control fields in a SAFE frame header and finally the SAFE MAC sublayer.

A. Contention and congestion control

SAFE datalink maintains a single buffer, called transit traffic buffer, for all outgoing frames including those generated by the node itself. Transit traffic buffer is first-in-first-out (FIFO), subject to the freezing rule as described below. The buffer is divided into two regions according to the `Queue Threshold` parameter, as shown in Figure 2. The region that is close to the head of the queue is called Free Region, and the rest of the buffer belongs to the Freeze Region. The size of the Free Region is exactly `Queue Threshold` number of datalink layer frames, regardless of the frame size.

With the transit traffic buffer maintained at the datalink layer the contention and congestion control algorithm is simple. If the total number of frames in the queue is less than `Queue Threshold`, i.e., with the Freeze Region empty, the node behaves as a normal 802.11 device. Once frames start to be buffered in the Freeze Region, i.e., the number of buffered frames exceeds `Queue Threshold`, the contention and congestion controller returns a *contention and congestion control signal*, piggybacked to the per-frame acknowledgement, to the upstream sender. `Queue Threshold`, or the size of the free region, is set to a small number, typically one frame for effective contention control. The rationale is that with only one frame in the buffer, a wireless router yields from channel contention right after it forwards the frame downstream. Therefore, a stream of packets will be evenly paced along the path from the source to the destination. This even distribution not only eases the contention at any local area, but also maximizes the spatial reuse of the shared wireless channel.

Minimum buffering at intermediate wireless routers for transit traffic also ensures that the contention and congestion control signal be quickly back-pressured to the sources. The timeliness of the signaling is critical for effective congestion control because eventually the sources have to cut their sending rates for an overloaded network to recover from congestion. Note that SAFE's back-pressure is *implicit* in that the signaling is always piggybacked to the data or existing CSMA/CA handshake frames, and is applied on a *per-neighborhood* basis. It is slightly different from the back-pressure congestion control in traditional telecommunication networks, where explicit, per-flow signaling is employed.

The freeze region is used to accommodate the overflow of the small free region due to signaling errors. Since a SAFE wireless router starts to slow down its upstream routers at very early stage, far before the transit traffic buffer is close to full, packet losses due to buffer overflow are unlikely with a reasonable amount of buffer space in freeze region. In the rare cases the buffer is indeed full, due to either signaling error or small buffers at low-end transceivers, SAFE contention and congestion controller sends back a *negative acknowledgement* to the upstream wireless router. The upstream router that receives such a negative acknowledgement will then retransmit the frame after appropriate freezing.

In summary, through SAFE datalink layer contention and congestion control a RAIN network can effectively minimize the contention and completely eliminate packet losses due to buffer overflow. It demonstrates that on the contrary to the common belief, it is possible to build a multihop wireless network that is even more reliable than a wired network, using off-the-shelf CSMA/CA wireless transceivers.

MAC Address	$T_{freeze}(\mu s)$
00-13-20-56-6c-60	1,200
00-23-10-53-ab-12	300
00-15-30-4b-3a-34	12,200

TABLE I
A TABLE OF NEIGHBOR STATUS (TONS) EXAMPLE

B. SAFE Frame formats

To carry contention and congestion control signals, SAFE adds the `SAFE control` into the frame header, as shown in Figure 3. `SAFE Control` is composed of two fields, counting for 2 bytes in total. One field counts for 1 bit, and announces the status of the transit traffic buffer, i.e., whether the Free Region is full or not. The other field counts for 15 bits. It represents the requested freezing duration.

If transit traffic buffer status is set, i.e., the Free Region is full, the requested freeze duration is set to

$$\left\lceil \frac{\overline{T_{pkt}} \times N}{100\mu s} \right\rceil$$

where $\overline{T_{pkt}}$ and N are the average transmission time for a data frame and the number of data frames in the transit traffic buffer respectively. We choose $100\mu s$ as the unit to scale the requested freezing duration from 100μ to 3.2767 seconds, representing the minimum time it takes to transmit a frame (given the constant physical layer per-frame overhead) and the interval that is large enough to accommodate the longest queue length.

If the transmit traffic buffer status field is not set, then the requested freezing duration field must be zero. Any non-zero freezing duration together with the transit traffic buffer status unset represents a negative acknowledgement, as described in the above section.

$\overline{T_{pkt}}$ is maintained as a moving average over recent samples of the frame transmission time as follows:

$$\overline{T_{pkt}} = \alpha T_{pkt} + (1-\alpha)\overline{T_{pkt}}$$

where α is set to 0.3 in the performance evaluation. Note that T_{pkt} denotes the total time transmitting a head of line data frame in the transit traffic buffer. It includes the inter-frame random backoff, the requested freezing for the specific next-hop neighbor, and the retransmissions. It varies over time according to the channel contention level and routing stability, and therefore cannot be deterministically derived with the frame size and the channel rate.

C. Medium Access Control

SAFE maintains a Table of Neighbor Status (ToNS) with one neighbor per entry. ToNS has two fields: the identifier (MAC address) of neighbors and the remaining requested freezing duration time (T_{freeze}), as shown in Table I. The table entry is updated whenever a wireless router receives or overhears the two SAFE control fields as defined above. A ToNS entry is removed if a frame with buffer status bit unset is received from the corresponding neighbor.

Module	ReTP	TCP NewReno
Connection management	√	√
Sliding window	√	√
Slow start, congestion avoidance	×	√
Flow control	×	√
Timeout and Rx	√	√
Fast Rx and fast recovery	√	√
Delayed ACK	√	√
Adaptive delayed ACK	√	×

TABLE II
COMPARISON BETWEEN TCP NEWRENO AND RETP

With the ToNS maintained, SAFE simply goes through the transit traffic buffer, starting from the head of the queue, and transmits the first frame of which the destination address does not appear in ToNS. That is, SAFE transmits only to the downstream node whose transit traffic buffer is likely empty. This way, SAFE enforces the requested freezing duration advertised by neighboring nodes and realizes the contention and congestion control. Note that the way SAFE enforces the requested contention and congestion control is similar to the way virtual carrier sense works in the IEEE 802.11 standard. Therefore building SAFE into the existing CSMA/CA wireless transceivers will be straightforward.

V. RAIN TRANSPORT

A packet can be lost in a multihop wireless network due to the following five reasons: wireless channel errors, excessive contention for the shared wireless channel, congestion at intermediate wireless routers, wireless router failures, and routing failures due to high node mobility. Packet losses due to channel errors have been addressed in standard CSMA/CA wireless MAC through channel rate adaptation, per-frame acknowledgement and retransmission [29]. Link-quality aware routing [18], [20], [21] also helps to establish routes along high-quality wireless connections. Furthermore, with SAFE datalink layer implemented in a RAIN network, packet losses due to contention or congestion are highly unlikely. Therefore, for applications that demand 100% reliability, a reliable transport layer only needs to deal with packet losses due to failed wireless routers or routing. Although a full-fledged TCP suffices for the goal, there are a number of TCP functions unnecessary in a highly reliable RAIN network. Moreover, when node mobility is modest, e.g., in the wireless mesh network, router and routing failures are both rare. The transport protocol can exploit this fact, and acknowledges data segments less frequently to further improve the end-to-end throughput.

In this section we present ReTP, a reliable transport protocol designed specifically for a RAIN network. Table II shows the comparison between ReTP and TCP NewReno. ReTP removes congestion control and flow control from the TCP, and adds delayed ACK adaptation. We describe the way ReTP configures the sliding window size and delayed ACK adaptation in the rest of this section.

A. Sliding window size

ReTP applies a sliding window `swnd` to control the number of segments that ReTP will forward to the network layer before an ACK is received. Unlike TCP's congestion window `cwnd`, which has to be carefully adapted to ensure both fair and efficient utilization of the bottleneck bandwidth, ReTP's `swnd` adaptation can be much more coarse. As long as `swnd` is configured larger than the expected volume of the "pipe" along the path between the sender and the receiver, the in-network contention and congestion control will automatically keep excessive segments blocked at the source node, i.e., in the network layer buffer. Note that excessive queueing at the source nodes increases the delay and destabilizes round-trip-time estimate, leading to either conservative or premature retransmissions. Our goal is therefore to configure `swnd` that works the best with the adaptive retransmission that ReTP inherits from TCP.

Given the path between the sender and the receiver, the maximum number of active transmissions defines the path pipe volume. Note that because of the per-frame acknowledgement, the volume of the pipe between two neighboring wireless routers is exactly one segment, regardless of the segment size or the channel rate. Also note that because an IEEE 802.11 transceiver usually interferes with other transceivers within around twice the communication range, there are at most $\lceil \frac{h}{4} \rceil$ concurrent transmissions possible along a path of h hops. Therefore, `swnd` should be at least $\lceil \frac{h}{4} \rceil$ segments to allow full spatial reuse of the wireless channel. We find out that `swnd` $= 2\lceil \frac{h}{4} \rceil$ gives the best throughput in the majority of our simulated scenarios. We therefore fix ReTP's sliding window size to $2\lceil \frac{h}{4} \rceil$. Note that h is only available at the network layer routing module. RAIN opens a new interface between the network layer and the transport layer so that ReTP can be made aware of the path length.

A final tuning to `swnd` is that `swnd` should always be more than 4 segments for the following two reasons. One is that with less than 4 segments on the fly fast retransmissions can never be triggered for prompt retransmissions. The other reason is that ReTP delays the ACK aggressively to reduce the overhead, as presented in the next Section V-B. A small `swnd` limits the number of ACKs a ReTP can delay, and therefore unnecessarily bounds the achievable throughput gain especially when the network is stable.

B. Delayed ACK adaptation

ReTP is designed primarily to recover packet losses due to router or routing failures at high node mobility. Since those failures will be rare in a multihop wireless network with modest mobility, it is unnecessary for ReTP to acknowledge every data segment. Furthermore, end-to-end acknowledgement represents a large amount of overhead in a wireless network. Although the ACK segment itself is small compared with the size of a typical data segment, there is constant per-frame physical layer and datalink layer overhead, which is usually much larger than that in a wired network. For example, in the context of IEEE 802.11b with 2Mbps basic rate and 11Mbps data rate, transmitting a 40-byte ACK segment takes at least 35.6% of the time transmitting a 1500-byte TCP data segment. With RTS/CTS turned on or with higher channel rates (e.g., 6Mbps basic rate and 54Mbps data rate in IEEE 802.11a/g), the overhead increases up to 78%. Therefore, since both data and ACK usually traverse the same route between the sender and receiver, there is significant room for throughput optimization by decreasing the volume of ACK traffic.

Algorithm 1 Function - NmaxUpdate()

1: // *This function is called after sending an ACK*
2: // *Update* N_{ave} *with* $N_{current}$; $\alpha = 0.5$
$N_{ave} \leftarrow \alpha N_{ave} + (1-\alpha) N_{current}$
3: // *Reset* $N_{current}$
$N_{current} \leftarrow 1$
4: // *Update* N_{max} *with* N_{avg}
5: **if** $N_{ave} \geq 0.9 N_{max}$ **then**
6: $\quad N_{max} \leftarrow MIN(N_{max}+1,\ swnd)$
7: **else if** $N_{ave} \leq 0.7 N_{max}$ **then**
8: $\quad N_{max} \leftarrow MAX(1, 0.7 N_{max})$
9: **end if**

Ideally a ReTP receiver can simply acknowledge every `swnd` number of data segments, where `swnd` is the sender's sliding window size and can be communicated to the receiver during connection setup, i.e., three-way handshake. However, this strategy remains oblivious to the routing dynamics. When node mobility is high and routing failures are frequent, it is beneficial to acknowledge more frequently to avoid excessive waiting for the timeout of a lost packet. ReTP therefore builds adaptation into its delayed ACK. The idea is to maintain a moving average (N_{avg}) over the number of ACK's that a ReTP receiver recently delays ($N_{current}$), within the 500ms period and before receiving an out-of-order data segment. This moving average reflects the recent routing stability. We use the moving average to bound the maximum number of ACKs (N_{max}) that a ReTP receive can delay. The detailed algorithm is shown in Algorithm 1. The initial values of N_{avg} and N_{max} are set to 1.

Note that delayed ACK is a standard mechanism included in TCP to avoid large number of small ACK segments. Delaying more than two acknowledgements has been recently proposed to boost the TCP throughput in wireless networks [5], [6], [19]. ReTP is different from [5], [6] in that it adapts the number of delayed ACK to fit into the network dynamics. ReTP also differs from [19] in that the maximum window size at the sender is not bounded to a small fixed number 4.

VI. Performance Evaluation

In this section we evaluate the performance of the RAIN network through simulations. We present the RAIN's gains in throughput, fairness, and packet losses over both UDP and unmodified TCP. We start from simple linear network topologies, and then move on to the grid topology. We finally show the performance of RAIN in random, mobile networks.

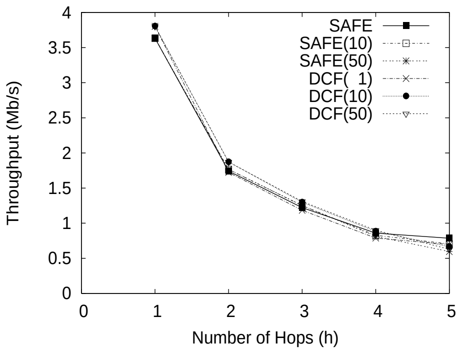

Fig. 4. UDP traffic in linear topologies.

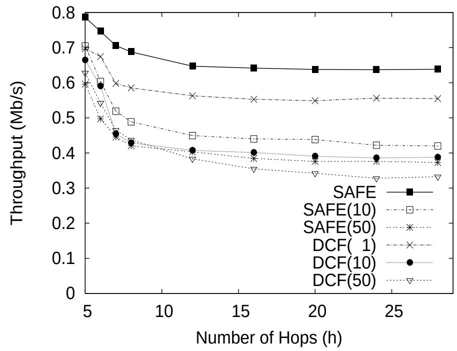

Fig. 5. UDP traffic in linear topologies.

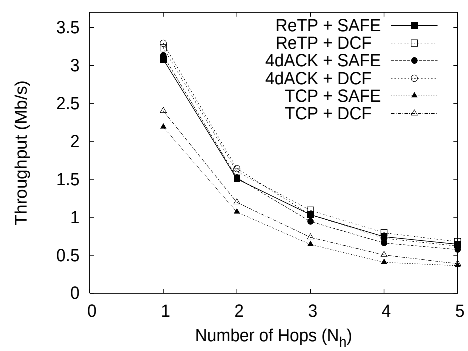

Fig. 6. Reliable transport in linear topologies.

Hop	Injected Packets From Sender			Routing Failure (pkts)			Buffer Overflow (pkts)			Drop Ratio (%)		
	DCF (1)	DCF (50)	SAFE	DCF (1)	DCF (50)	SAFE	DCF (1)	DCF (50)	SAFE	DCF (1)	DCF (50)	SAFE
1	42298	42347.4	41334.6	0	0	0	0	0	0	0	0	0
4	9880.2	10033.4	9760.4	0	0	0	1177.6	0	0	11.9	0	0
8	9232.2	5888.8	7815.4	2.6	471.4	0	2880.4	684.0	0	31.2	19.6	0
16	9295.2	5122.0	7296.6	3.8	486.0	2.4	3299.8	569.0	0	35.5	20.6	0.03
24	9229.8	4033.4	7242.0	4.6	261.2	1.8	3158.4	465.6	0	34.3	18.0	0.02

TABLE III

UDP PACKET LOSSES IN THE LINEAR TOPOLOGIES

We have implemented the RAIN architecture, as well as the SAFE datalink and the ReTP transport protocols in *ns*-2 simulator version 2.28. The transmission range of each node is configured to 100 meters and the carrier sensing range is set to 220 meters. The default buffer size for transmit packets at each node is 50 packets. We use AODV as the routing algorithm. For all experiments with UDP traffic, we set the packet rate sufficiently high so that the source sends out UDP data as fast as the network allows. All transport layer data segments are 1000 bytes in length. Every simulation runs for 100 seconds, and we present the data for the last 50 seconds when the simulations stabilize. We configure the data rates according to IEEE 802.11b standard, i.e., 11Mbps for data payload and 2Mbps for control signals.

We evaluate the performance using end-to-end throughput, inter-flow fairness, and packet drop ratios. Note that fairness is an important parameter because of the fundamental tradeoff between aggregate throughput and fairness in a multihop wireless network [34]. We use the Jain's fairness index [31] given as $\frac{(\sum_{i=1}^{n} x_i)^2}{n \sum_{i=1}^{n} x_i^2}$, where n is the number of flows and x_i is the throughput of flow i. Every data presented in this section is the average over 10 simulations configured with different random seeds.

A. Linear Topology

We first investigate the performance of UDP and reliable transport traffic, including TCP NewReno, TCP with delayed ACK, and ReTP, over SAFE and IEEE 802.11 DCF in simple linear topologies. All nodes are 90 meters apart from their immediate neighbors. The sender and the receiver are placed at both ends of the chain. We compare SAFE with the standard IEEE 802.11 distributed coordination function (DCF), the most popular CSMA/CA wireless MAC. We use DCF(N) to denote an IEEE 802.11 DCF node with packet buffer of N packets. We also compare ReTP with TCP NewReno and TCP with delayed ACK. For the latter we set the number of delayed ACK to 4, because it shows the best performance in our simulations and the previous evaluations [6], [19]. We denote this version of TCP with delayed ACK as "4dACK" and TCP NewReno as "TCP" in the legends of the figures and the rest of this section.

Figures 4–7 show that throughput comparisons, and Table III breaks down the packet losses due to routing failures and buffer overflow. Note that since nodes do not move in our simulations, all routing failures are caused by channel contentions. When the length of the path is small ($\leq$5 hops), the performance of SAFE and 802.11 DCF is similar as shown in Figure 4 and 6. In these cases almost all nodes are within the interference range, and 802.11 DCF handles the contention well through carrier sense and random backoff. Figure 5 shows the impact of buffer space on UDP throughput. As we can see from the figure, by simply decreasing the buffer space from the default 50 packets (i.e., DCF(50)) to 1 packet (i.e., DCF(1)), the throughput can be improved by up to 66.9%. Another 25.1% throughput improvement can be achieved with SAFE's contention and congestion control. Therefore, SAFE improves UDP throughput by up to 92% compared with DCF(50), the default multihop wireless network settings used in the majority of simulations in the literature.

The throughput gain can be clearly explained by the packet losses shown in Table III. As the buffer size is decreased from 50 to 1, packet losses due to routing failures (or channel contention) drop by almost two orders of magnitude. However, packet losses due to buffer overflow increase dramatically, close to one order of magnitude when the path is long. Overall,

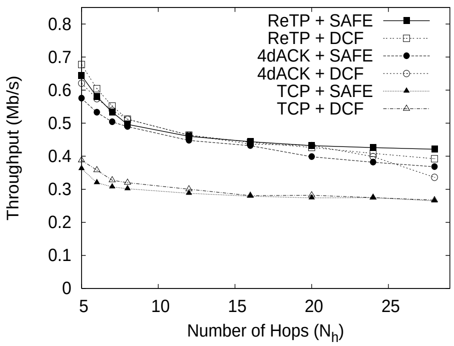

Fig. 7. Reliable transport in linear topologies.

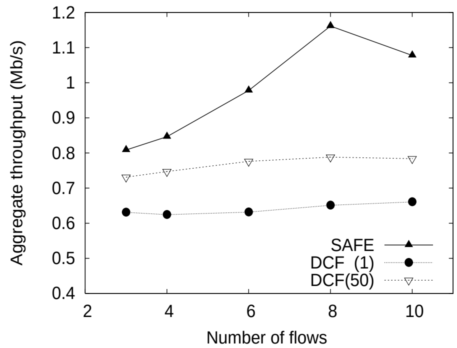

Fig. 8. UDP traffic in grid topologies.

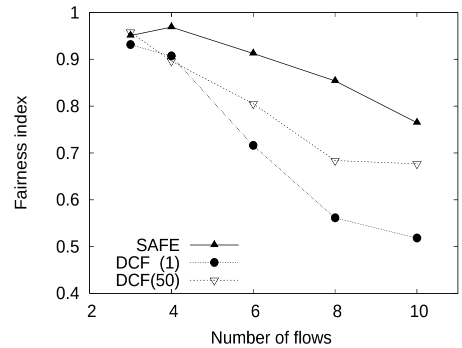

Fig. 9. UDP traffic in grid topologies.

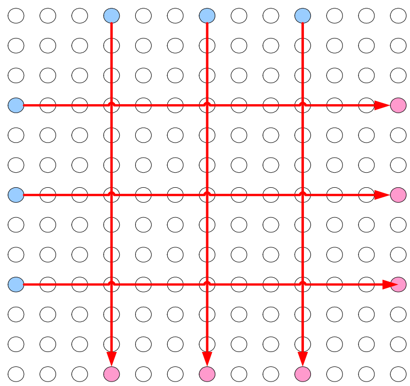

Fig. 13. Grid topology.

decreasing the packet buffer size from 50 to 1 increases the packet loss ratio by about 10%. SAFE benefits from small buffer in that the number of packet losses due to routing failure, or channel contention, is minimal (similar to DCF(1)). Meanwhile, thanks to its contention and congestion control SAFE does not suffer from buffer overflow. In fact, SAFE does not lose a single packet due to buffer overflow in our simulated scenarios. As a result, SAFE achieves zero packet losses when the path length is less than 8, and decreases the packet loss ratio from 18.0∼35.5% to as small as 0.03%.

Figures 6 and 7 show that both ReTP and 4dACK significantly improve the throughput over standard TCP by up to 57.6%, regardless of the link layer. This result demonstrates the importance of reducing the overhead of transport layer ACK messages, as we analyze in Section V-B. Furthermore, although SAFE improves the performance of both ReTP and 4dACK when the path is long, SAFE actually slightly decreases TCP throughput compared with 802.11 DCF. The reason is that SAFE seeks to deliver every frame reliably, including TCP ACKs, and therefore involves more overhead on the reverse path than an unreliable 802.11 DCF network. Since TCP usually opens its `cwnd` fairly large in an 802.11 DCF network, as observed in [23], a few lost ACKs do not cut the `cwnd` to the level that results in under-utilized wireless channel.

B. Grid Topology

We next study the performance of RAIN in a 13x13 grid topology. Similar to the linear topology every node is 90 meters apart from its four immediate neighbors. We simulate a number of flows in the grid, with half the flows horizontal and the other half vertical. All horizontal (or vertical) flows are evenly spaced. Figure 13 shows the network topology and the spatial distribution of 6 flows. Figures 8–12 show the performance of UDP and reliable transport traffic, with SAFE and 802.11 DCF. As we can see from Figures 8–10, DCF(50) outperforms DCF(1) by more than 10% in throughput, because of the significant buffer overflow problem at DCF(1). SAFE achieves a throughput gain between 10∼47%, depending on the number of flows, or the offered load in the network. It also significantly improves the throughput fairness by up to 25%, compared with the the fairest one between DCF(50) and DCF(1), as shown in Figure 9. It demonstrates that SAFE does not achieve higher throughput at the cost of lower fairness, e.g., flow starvation. Both throughput and fairness gain result from the ∼78% reduction in packet loss ratio, as shown in Figure 10.

Figures 11 and 12 show the comparison between various combination of datalink and transport protocols. Note that the 4dACK with SAFE or 802.11 DCF achieves the highest throughput, but significantly lags behind RAIN in fairness. In fact, 4dACK's fairness is even lower than TCP, with SAFE or 802.11 DCF. RAIN network (i.e., ReTP+SAFE) improves the throughput by 38∼54% over TCP with 802.11 DCF. Meanwhile, RAIN achieves a 23∼43% gain in fairness.

C. Random Mobile Topology

We finally evaluate RAIN in highly mobile multihop wireless networks. We randomly place 200 nodes within a 1000 m x 1000 m area. Nodes move according to the random walk model, i.e., each node stays one point for a random period between 0 and 1 seconds, and then starts to move towards a random point with a speed randomly chosen between 1m/s to 10m/s. Each simulation runs for 100 seconds. We start 10 flows randomly within the first 10 seconds. The senders and the receivers of all 10 flows are randomly chosen.

Table IV and V show the throughput, fairness, and packet

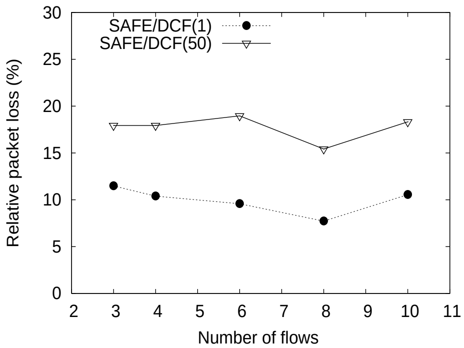

Fig. 10. UDP traffic in grid topologies.

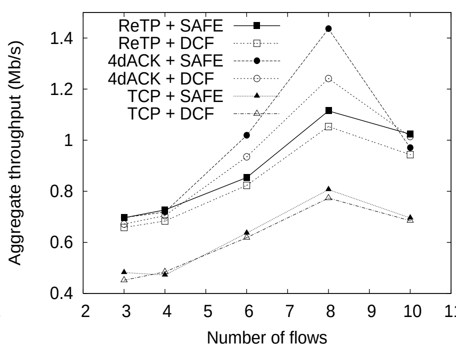

Fig. 11. Reliable transport in grid topologies.

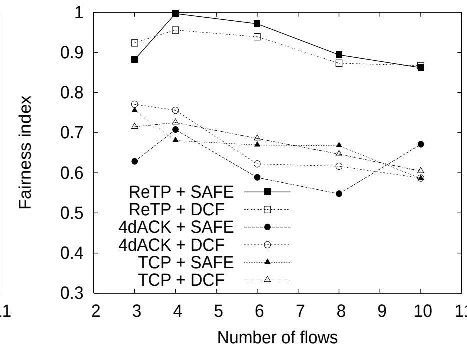

Fig. 12. Reliable transport in grid topologies.

Index	Aggregate throughput (Mbps)		Fairness index		Packet loss (pkts)	
	SAFE	802.11 DCF	SAFE	802.11 DCF	SAFE	802.11 DCF
1	3.64	3.30	0.50	0.44	4522	30721
2	2.26	1.73	0.32	0.30	4381	27562
3	0.85	0.68	0.70	0.38	6306	27228
4	0.97	0.81	0.50	0.28	4606	26186
5	3.12	2.98	0.36	0.30	3654	25797
6	2.19	1.98	0.29	0.25	4240	20355
7	0.66	0.49	0.64	0.44	5386	29000
8	1.43	1.30	0.52	0.34	4386	24235
9	1.95	1.56	0.38	0.30	4293	26185
10	1.83	1.19	0.32	0.23	4828	29793
Average	1.89	1.60	0.45	0.33	4660.2	26706.2
Percentage	118	100	136	100	17	100

TABLE IV

UDP TRAFFIC IN MOBILE RANDOM TOPOLOGIES

Index	Aggregate throughput (Mbps)						Fairness index					
	SAFE			802.11 DCF			SAFE			802.11 DCF		
	ReTP	4dACK	TCP	ReTP	4dACK	TCP	ReTP	4dACK	TCP	ReTP	4dACK	TCP
1	4.50	2.49	3.07	3.88	1.89	3.20	0.45	0.26	0.41	0.38	0.25	0.34
2	2.56	2.45	1.59	2.29	2.55	1.77	0.31	0.32	0.25	0.25	0.36	0.26
3	3.23	2.56	0.96	1.18	2.47	1.06	0.36	0.29	0.23	0.24	0.40	0.21
4	1.66	1.34	1.03	1.50	1.27	1.39	0.26	0.13	0.24	0.23	0.23	0.21
5	3.75	2.76	2.34	3.47	2.88	2.54	0.38	0.30	0.25	0.29	0.27	0.23
6	3.03	2.49	1.85	2.74	2.39	2.02	0.30	0.27	0.21	0.22	0.33	0.18
7	1.63	1.08	0.61	0.87	1.18	0.74	0.40	0.31	0.35	0.28	0.18	0.22
8	2.69	2.15	1.16	1.45	2.26	1.21	0.45	0.31	0.36	0.30	0.25	0.31
9	2.19	1.44	1.36	1.81	1.48	1.50	0.36	0.22	0.32	0.33	0.24	0.28
10	2.06	2.01	1.26	1.83	2.82	1.50	0.41	0.31	0.34	0.29	0.26	0.29
Average	2.73	2.08	1.52	2.10	2.12	1.70	0.37	0.27	0.30	0.28	0.28	0.25
Percentage	161	122	89	124	128	100	148	108	120	112	112	100

TABLE V

RELIABLE TRANSPORT IN MOBILE RANDOM TOPOLOGIES

losses for UDP and reliable transport traffic respectively. From Table IV we can see that SAFE consistently outperforms 802.11 DCF in all simulated scenarios in both throughput and fairness. On average, SAFE improves throughput by 18% and fairness by 36%, while cut packet losses by 83%. For reliable transport traffic, as shown in Table V, RAIN (i.e., ReTP+SAFE) consistently outperforms all other combinations in terms of both throughput and fairness. On average, RAIN improves throughput by 61% and fairness by 48% when compared with TCP+802.11 DCF. Even compared with 4dACK+802.11 DCF RAIN achieves 25.2% higher in throughput and 32.1% higher in fairness. Note that similar to the results with grid topology, as presented in Section VI-B, SAFE does not improve the performance of TCP because of the overhead involved in the highly reliable transport of every ACK message.

VII. CONCLUSION

Packet losses due to wireless interference or the contention for the shared wireless channel complicate the design and implementation of multihop wireless networks. In this paper we propose a new reliable architecture with built-in contention and congestion control for multihop wireless networks. RAIN re-defines the vertical interface between the datalink layer and the network layer to enable joint buffer management for contention control. It also re-defines the horizontal interface between neighboring wireless routers, in the form of new semantics for the standard per-frame acknowledgement and inter-frame backoff. Detailed protocol designs and evaluation through intensive simulations demonstrate the feasibility and potential performance gain that RAIN can bring into a multihop wireless network. We are currently implementing the RAIN architecture and proposed protocols in our wireless testbed, built on Linux wireless routers and the licensed version of the MADWiFi driver that allows programmability deep into the firmware.

References

[1] Champaign-Urbana community wireless network. http://www.cuwireless.net/.

[2] MIT roofnet. http://www.pdos.lcs.mit.edu/roofnet/.

[3] MSR self-organizing neighborhood wireless mesh networks. http://www.research.microsoft.com/mesh/.

[4] I. F. Akyildiz, X. Wang, and W. Wang. Wireless mesh networks: a survey. *Elsevier Computer Networks Journal*, 47(4):445–487, March 2005.

[5] M. Allman. On the generation and use of TCP acknowledgments. *ACM SIGCOMM Computer Communication Review*, 28(5):4 – 21, 1998.

[6] E. Altman and T. Jimenez. Novel delayed ack techniques for improving TCP performance in multihop wireless networks. In *Proceedings of PWC*, 2003.

[7] P. Bahl, R. Chandra, and J. Dunagan. SSCH: Slotted seeded channel hopping for capacity improvement in IEEE 802.11 ad-hoc wireless networks. In *Proceedings of ACM MobiCom*, 2004.

[8] L. Bao and J. Garcia-Luna-Aceves. Transmission scheduling in ad hoc networks with directional antennas. In *Proceedings of ACM MobiCom*, 2002.

[9] D. Berger, Z. Ye, P. Sinha, S. Krishnamurthy, M. Faloutsos, and S. K. Tripathi. TCP-friendly medium access control for ad-hoc wireless networks: Alleviating self-contention. In *Proceedings of IEEE MASS*, 2004.

[10] V. Bharghavan. Performance evaluation of algorithms for wireless medium access. In *Proceedings of IEEE Performance and Dependability Symposium*, 1998.

[11] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang. MACAW: A medium access protocol for wireless LANs. In *Proceedings of ACM SIGCOMM*, 1994.

[12] J. D. Camp, E. W. Knightly, and W. S. Reed. Developing and deploying multihop wireless networks for low-income communities. In *Poster of IEEE INFOCOM*, 2005.

[13] C. cheng Chen and H. Luo. The case for heterogeneous wireless MACs. In *Proceedings of HotNets*, 2005.

[14] C. cheng Chen, E. Seo, H. Kim, and H. Luo. Self-learning collision avoidance for wireless networks. In *Proceedings of IEEE INFOCOM*, 2006.

[15] R. R. Choudhury and N. Vaidya. Deafness: A MAC problem in ad hoc networks when using directional antennas. In *Proceedings of IEEE ICNP*, 2004.

[16] R. R. Choudhury, X. Yang, R. Ramanathan, and N. Vaidya. Medium access control in ad hoc networks using directional antennas. In *Proceedings of ACM MobiCom*, 2002.

[17] D. D. Clark. The design philosophy of the darpa internet protocols. In *Proceedings of ACM SIGCOMM*, 1988.

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

[19] R. de Oliveira and T. Braun. A dynamic adaptive acknowledgment strategy for TCP over multihop wireless networks. In *Proceedings of IEEE INFOCOM*, 2005.

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

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

[22] Z. Fu, B. Greenstein, X. Meng, and S. Lu. Design and implementation of a tcp-friendly transport protocol for ad hoc wireless networks. In *Proceedings of IEEE ICNP*, 2002.

[23] Z. Fu, H. Luo, P. Zerfos, S. Lu, L. Zhang, and M. Gerla. The impact of multihop wireless channel on TCP performance. *IEEE Transactions on Mobile Computing*, 4(2):209–221, March/April 2005.

[24] C. L. Fullmer and J. Garcia-Luna-Aceves. Solutions to hidden terminal problems in wireless networks. In *Proceedings of ACM SIGCOMM*, 1997.

[25] V. Gambiroza, B. Sadeghi, and E. W. Knightly. End-to-end performance and fairness in multihop wireless backhaul networks. In *Proceedings of ACM MobiCom*, 2004.

[26] M. Garetto, J. Shi, and E. W. Knightly. Modeling media access in embedded two-flow topologies of multihop wireless networks. In *Proceedings of ACM MobiCom*, 2005.

[27] G. Holland and N. Vaidya. Analysis of TCP performance over mobile ad hoc networks. In *Proceedings of ACM MobiCom*, 1999.

[28] IEEE Computer Society. Specification for 802.3 full duplex operation. IEEE standard 802.3x-1997, 1997.

[29] IEEE Computer Society. Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. IEEE standard 802.11, 1999.

[30] C. Ikeda and H. Suzuki. Adaptive congestion control schemes for atm lans. In *Proceedings of IEEE INFOCOM*, 1994.

[31] R. Jain, A. Durresi, and G. Babic. Throughput fairness index: An explaination. In *ATM Forum/99-0045*, 1999.

[32] V. Kawadia and P. Kumar. A cautionary perspective on cross layer design. *IEEE Wireless Communication Magazine*, pages 3–11, February 2003.

[33] J. Liu and S. Singh. ATC: TCP for mobile ad hoc networks. *IEEE Journal on Selected Areas in Communications*, 19(7):1300–1315, July 2001.

[34] H. Luo, S. Lu, and V. Bharghavan. A new model for packet scheduling in multihop wireless networks. In *Proceedings of ACM MobiCom*, 2000.

[35] S. Mascolo, C. Casetti, M. Gerla, M. Y. Sanadidi, and R. Wang. TCP westwood: Bandwidth estimation for enhanced transport over wireless links. In *Proceedings of IEEE ICNP*, 2002.

[36] K. Nahm, A. Helmy, and C.-C. J. Kuo. TCP over multihop 802.11 networks: Issues and performance enhancement. In *Proceedings of ACM MobiHoc*, 2005.

[37] R. Ramanathan. Challenges: A radically new architecture for next generation mobile ad hoc networks. In *Proceedings of ACM MobiCom*, 2005.

[38] A. Raniwala and T. cker Chiueh. Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network. In *Proceedings of IEEE INFOCOM*, 2005.

[39] J. H. Saltzer, D. P. Reed, and D. D. Clark. End-to-end arguments in system design. *ACM Transactions on Computer System*, 2:277–288, November 1984.

[40] K. Sanzgiri, I. D. Chakeres, and E. M. Belding-Royer. Determining intra-flow contention along multihop paths in wireless networks. In *Proceedings of Broadnets Wireless Networking Symposium*, 2004.

[41] P. Sinha, T. Nandagopal, N. Venkitaraman, R. Sivakumar, and V. Bharghavan. WTC: a reliable transport protocol for wireless wide-area networks. In *Proceedings of ACM MobiCom*, 1999.

[42] J. So and N. H. Vaidya. Multi-channel MAC for ad hoc networks: Handling multi-channel hidden terminals using a single transceiver. In *Proceedings of ACM MobiHoc*, 2004.

[43] K. Sundaresan, V. Anantharaman, H.-Y. Hsieh, and R. Sivakumar. ATP: a reliable transport protocol for ad-hoc networks. In *Proceedings of ACM MobiHoc*, 2003.

[44] C. Tschudin and E. Osipov. Estimating the ad hoc horizon for TCP over IEEE 802.11 networks. In *Proceedings of Med-Hoc-Net*, 2004.

[45] K. Xu, M. Gerla, L. Qi, and Y. Shu. TCP unfairness in ad hoc wireless networks and a neighborhood RED solution. In *Proceedings of ACM MobiCom*, 2003.

[46] S. Xu and T. Saadawi. Does the IEEE 802.11 MAC protocol work well in multihop wireless ad hoc networks? *IEEE Communication Magazine*, 39(6):130–137, June 2001.

[47] Y. Yi and S. Shakkottai. Hop-by-hop congestion control over a wireless multi-hop network. In *Proceedings of IEEE INFOCOM*, 2004.