Assessing The Real Impact of 802.11 WLANs: A Large-Scale Comparison of Wired and Wireless Traffic

Félix Hernández-Campos[a] Maria Papadopouli[a,b,c]

***Abstract*— We compared the traffic from hosts connected to the network via a wired or wireless interface, emphasizing the impact of 802.11 on packet delay and loss. Our study uses only passive monitoring techniques, namely, inference from TCP header traces. This enabled us to study a population of several thousand hosts in a real production environment, in which more than 31 million TCP connections were made. Our first contribution is methodological. Passive methods always have some degree of uncertainty, and we overcome this limitation by mostly relying on relative differences between wired and wireless traffic. Our analysis revealed that wireless clients experienced substantially higher packet delay variability than wired clients but their loss rates are surprisingly similar. We found that both the number of unnecessary TCP retransmissions and, even more substantially, the number of interrupted connections are higher for the wireless LAN than for the wired LAN. To the best of our knowledge, this is the first research effort to directly contrast wired and wireless traffic of a large production network.**

I. Introduction

The increasingly successful deployment of 802.11 wireless networks has motivated numerous research efforts in recent years. Wireless networking provides a plethora of novel research questions in topics such as mobility, power management and capacity planning. One of the current challenges in this area is to develop more accurate and realistic characterizations of production wireless networks and their performance. This will help to incorporate more representative assumptions in theoretical and simulation studies, perform more realistic testbed experiments, and design benchmarks.

The present paper focuses on characterizing packet level performance in wireless networks. There is a growing number of studies that examine this issue using controlled experiments in a small wireless network and conducting active measurement experiments on it [13], [5], [15]. In contrast, our study looks at a large production network using passive measurements (inferences from packet headers), providing a characterization of the network during normal operation rather than under artificial conditions. In particular, we analyze traffic from several thousand wireless hosts at the University of North Carolina at Chapel Hill (UNC). Furthermore, we not only consider traffic from wireless clients, but also traffic from wired clients. Figure 1 illustrates our measurement setup.

a. Department of Computer Science, University of North Carolina at Chapel Hill, USA.

b. Institute of Computer Science, Foundation for Research and Technology-Hellas, Greece.

c. This work was partially supported by the IBM Corporation under an IBM Faculty Award 2004/2005 grant.

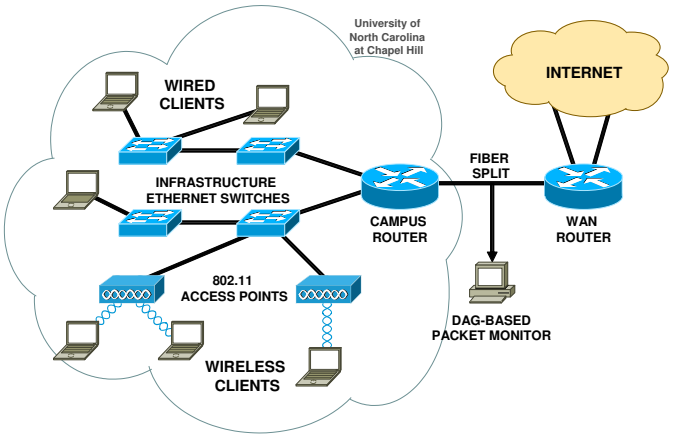

Fig. 1. Our data collection took place in the Gigabit Ethernet border link between UNC and the Internet. UNC has a large campus network with almost 600 wireless access points and several thousand Ethernet switches that form the backbone of the university's LAN.

Our data (packet header traces) comes from the border link between the university and the rest of the Internet. Both wireless and wired clients must use this link in order to communicate on the Internet. Our study is made possible by the fact that university relies on DHCP [3] to dynamically assign IP addresses to network hosts. Every time a host joins the network, either by using a wired or a wireless interface, it receives an IP address from the pool of network prefixes that the university reserves for this purpose. There are two distinct large subsets of prefixes, each used by the wireless and wired clients, respectively. This makes it possible to collect packet headers and unambiguously distinguish wired client connections from wireless client connections. Since our monitoring takes place very close to the LAN, we are able to get a rather accurate picture of packet dynamics before they are distorted by the WAN. Note also that the DHCP network prefixes we monitored were never used by university servers, which use static IP addresses. We therefore study connections from university clients that are likely to use the network in similar ways, modulo the impact that access technology may have on user behavior. To the best of our knowledge, this is the first research effort to directly contrast wired and wireless traffic of a large production network.

II. Data Set

Our data set consists of a total of 175 GB of packet header traces collected from the link between the University of North Carolina at Chapel Hill and the rest of the Internet (see Figure

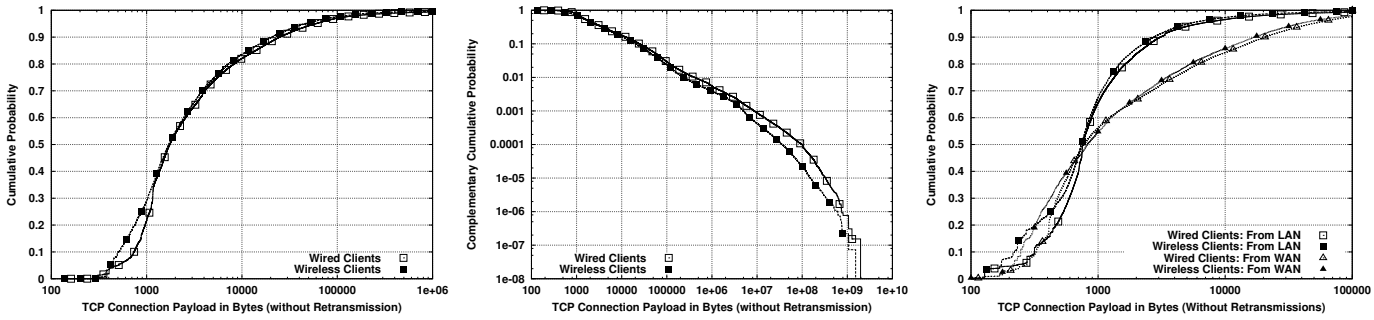

Fig. 2. Distributions of connection sizes in bytes: breakdown by client type (CDFs on the left, CCDFs in the middle) and by direction (CDFs on the right).

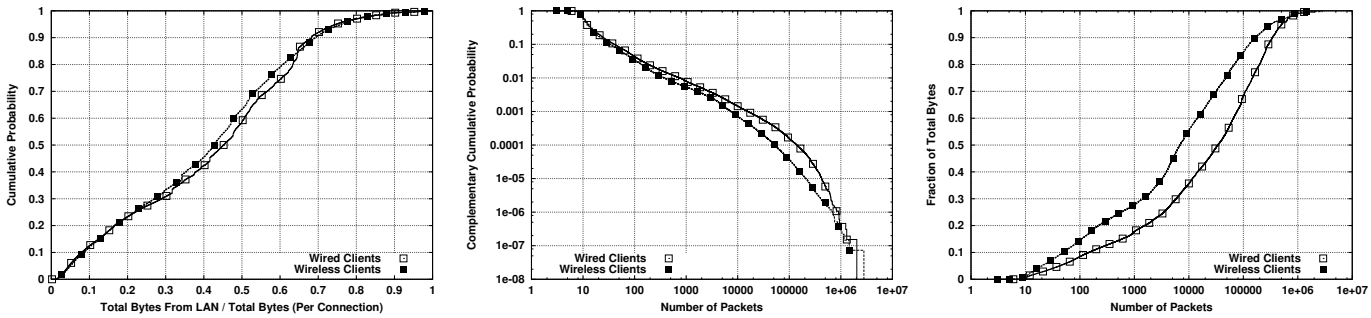

Fig. 3. Distributions of connection directionality ratio (left), connection sizes in number of segments (middle), and cumulative fraction of bytes associated to connections with a given number of segments (right).

1). The data collection took place between 12:06 PM on Wednesday April 13rd, 2005, and 22:18 PM on Wednesday, April 20th, 2005, resulting in a continuous trace of 178.2 hours. Packet headers were acquired using a high-precision monitoring card (Endance's DAG 4.3 GE) attached to the receiving end of a fiber split. The card was installed in a high-end FreeBSD server. Neither the server nor the card's driver reported any failures or packet drops during the monitoring.

The analysis of our data revealed at total of 9,766,507 TCP connections from wired clients and 21,396,174 TCP connections from wireless clients. Significant fractions of these connections (33.66% and 36.47% respectively) corresponded to pathological cases where no useful payload was exchanged between the two end points. Examples of these cases are attempts to connect to offline servers or hosts behind a firewall, where we only observed SYN segments in one direction, and attempts to connect to closed port, where we observed SYN segments in one direction and reset segments in the opposite one. Some of these connections are perfectly legitimate (but unsuccessful) while other are malicious network scans and port scans. While the fraction of connections without any useful payload is significant, it represents only a tiny fraction of the traffic, 0.09% of the bytes from wired clients, and 0.16% of the bytes from wireless. We do not perform any further analysis of these pathological connections in the rest of these study.

TCP connections from wired and wireless clients carried a total of 500.53 GB and 567.66 GB respectively. Given the substantially smaller number of connections from wired clients, we infer the average size of the connections from wired clients is substantially larger. However, as the left plot in Figure 2 shows, the distributions of connections sizes had very similar bodies. The cause of the higher average is in the tails of these distributions (middle plot of Figure 2): While both distributions are quite heavy, the tail of the distribution for the wired LAN is significantly heavier[1]. Therefore, it was somewhat more likely to encounter very large connections (50+ MB and above) in the wired LAN, but very large connections were also observed in the wireless LAN. We also observed a higher utilization of port numbers associated with peer-to-peer applications (BitTorrent and Gnutella) by wired clients.

The right plot in Figure 2 shows the bodies of the distributions of connection sizes for the two directions of traffic, separately. The distributions of bytes sent from the LAN (*i.e.*, from the university) are quite different from the distributions of bytes sent from the WAN, but there is a remarkable agreement between the wired and wireless clients. The tails of this distributions (not shown) are consistent with the the middle plot, *i.e.,* the distributions for the wired clients are heavier than those for wireless clients. The left plot of Figure 3 examines the directionality in the flow of data. The distribution for wireless clients shows only a small tendency to sent more bytes than received. The distributions of the number of packet per connection are shown in the middle plot of Figure 3 using a CCDF. As in the case of the bytes, most connections had a small number of packets. This has important implications when applying packet inference methods to the traces. These methods are more reliable for connections with more packets, which provide more samples. However, the right plot in Figure 3 shows that while there was a relatively small number of connections with a large number of packets, they carried the majority of the bytes. For example, connections with 100 packet or more represented less than 5% of the connection, but they carried 85%-90% of the bytes.

[1]The reader should not be confused by the seemingly small difference between the two curves. The sizes in the tails represent 10s/100s of MBs, so even a modest increase in the fraction of large sizes can easily skew the average of a distribution where 99% of the sizes are between 1 and 10 KB.

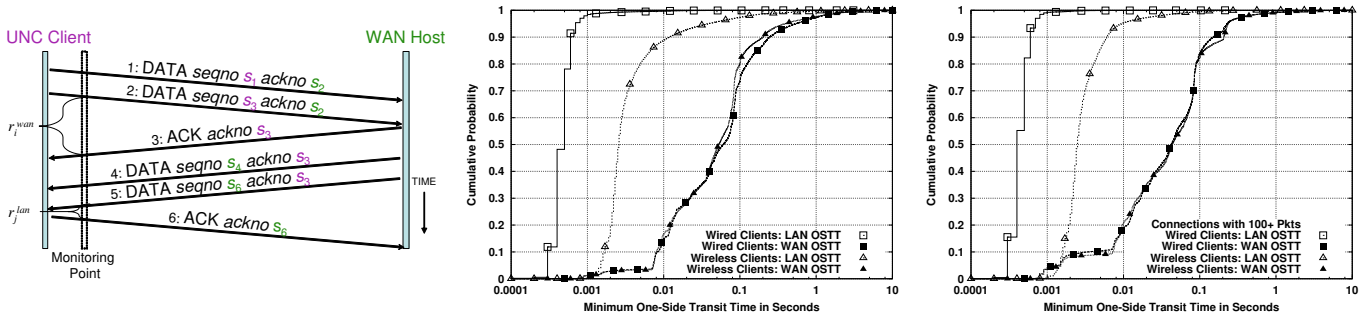

Fig. 4. Basic illustration of OSTT measurements (left) and distributions of minimum OSTTs for all connection (middle) and those with 100 packets or more (right).

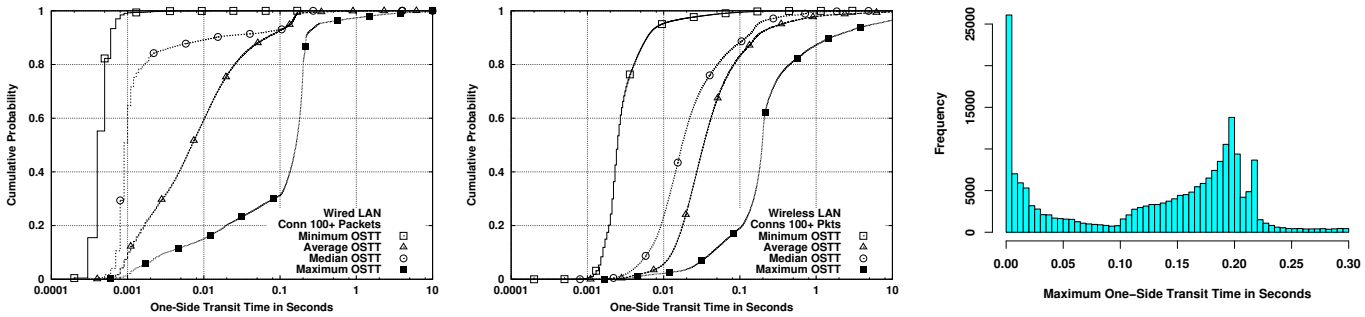

Fig. 5. Distributions of four different statistics of the connection LAN OSTTs for wired (left) and wireless (middle) clients; Histogram of maximum LAN OSTTs from wired clients (right).

III. Packet Delay

One of the salient features of 802.11 networks is the use of positive acknowledgments to detect and correct packet losses. Any packet sent from the client to access points (or vice versa) is expected to be acknowledged right after its reception. If no acknowledgment is received, the host sending the packet retries again up to five times [4]. The rationale of this behavior is to ameliorate the high loss rates encountered in wireless communication. In practice, this also means that the time required by a packet to go from one host to another in the wireless network is not only determined by the propagation but also by the number of times the packets has to be retransmitted. It is therefore expected to observe higher delays in wireless LANs than in wired LANs. In this section, we use a TCP-based passive measurement technique to quantify how much higher this delays are in the real world. In order to avoid confusion, we will refer to the 802.11 acknowledgment as LL-ACKs (link-layer acknowledgments) and to the TCP acknowledgments simply as ACKs.

The measurement method employed in this section to study delays is illustrated in the left plot of Figure 4. Our intend was to study delay on the LAN and WAN sides of connections independently. Given the location of the monitor, packets sent from the university were observed right after they traversed the local area network. Those packets sent from hosts outside the university were observed after they traversed their entire Internet path up to the university ingress point. We can then say that this monitoring setup divided connections into two sides, a LAN side (the university's local area network) and a WAN side (the rest of the Internet). The basic method for measuring delays from packet header traces is to couple TCP data segments and corresponding TCP-ACKs [1], [8], [6]. In the figure, the difference between the arrival times of segments 2 and 3 a measurement d_i^{wan} of the time required to travel from the monitoring point and back on the WAN side of the connection. Similarly, coupling a TCP data segments sent in the opposite direction with its corresponding TCP-ACKs provides a delay measurement on the LAN side of the connection, as illustrated with the last two segment in the figure. Due to the way these measurements split end-to-end round-trip time (RTT) into two parts, we will refer to type of delay observations as *one-side transit times* (OSTTs). It is important to note that packet reordering and retransmissions create ambiguous cases (see [9] and [14]) from which no reliable delay measurement can be made, and we carefully filter out these cases from our measurements.

Our analysis of the packet headers observed for each TCP connection results in a set of OSTT measurement for the LAN side $\{d_1^{lan}, d_2^{lan}, \ldots, d_n^{lan}\}, n \geq 0$, and another set of measurement for the WAN side $\{d_1^{wan}, d_2^{wan}, \ldots, d_m^{wan}\}, m \geq 0$. In general, n and m become larger for connections with a larger number of segments, since they provide more opportunities observe OSTTs by coupling TCP data and ACK segments. Given that the number of observation per connection is variable, studying the population of connection delays requires to the computation of a statistic (*e.g.*, minimum, average) in order to summarize the information in each set of OSTT observations. The middle plot in Figure 4 shows the distribution of minimum OSTTs for wired and wireless clients and for the LAN and the WAN sides. As expected, WAN delays are much larger than LAN delays. The two WAN OSTT distributions are very similar, and have the bulk of their values between 6 and 250 milliseconds. We therefore observe little dependency between the types of access technology (wired and wireless) and the range of delays to the hosts contacted by the university clients.

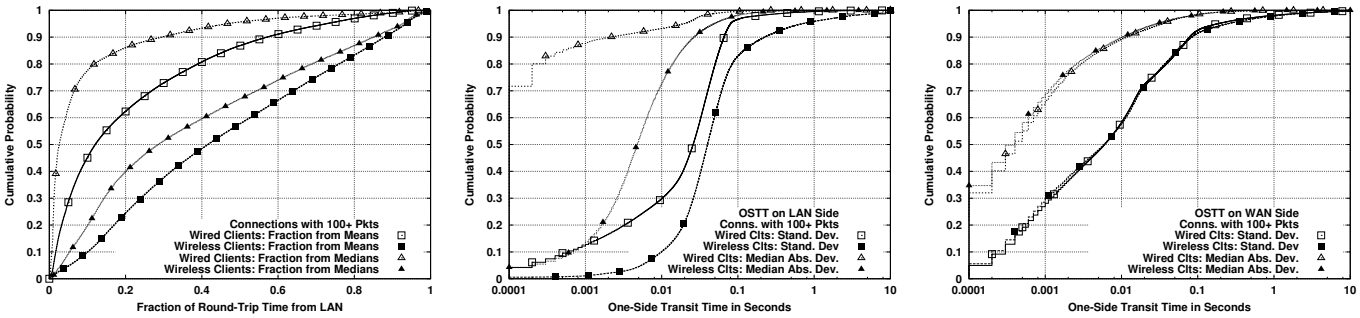

Fig. 6. Distributions of the fraction of RTT from wired and wireless LANs (left) and distributions from two measures of OSTT dispersion for LAN (middle) and WAN (right) sides.

The distributions of LAN OSTTs appear less alike. Almost all of the connections from the wired LAN had minimum OSTTs between 0.7 and 1 ms. In contrast, the distribution shows two regions, one with high slope between 1 ms and 7 ms, and another one with a lower slope between 7 and 250 ms. We hypothesize that the first region corresponds to pairs of packets that suffered no link-layer retransmissions, while those in the second region come from connections where no samples that were unaffected by retransmission was available. This brings up an important issue. A statistic computed from a set of OSTTs becomes a more reliable summary of the conditions experienced by a TCP connection as the number of observations becomes larger. The right plot of Figure 4 shows the same distributions of minimum OSTTs but only for connections with 100 or more packets. The distributions are generally quite similar, although the fraction of observations in the [7, 250] region of the wireless LAN OSTTs became substantially smaller. In any event, the plots illustrate that connections with 100 packets or more experienced a similar range of delays to that of the full set of connections.

The left of plot of Figure 5 shows the distributions of several OSTT statistics (minimum, average, median and maximum) computed for connections using the wired LAN with 100 or more packets. The middle plots shows the same distribution for connections using the wireless LAN. One striking observation about the left plot is the heavy distribution of maximum OSTT. Seventy percent of the values were above 100 milliseconds, which seems counterintuitive for a wired LAN. Upon close examination, we realized that these large values come from TCP's delayed ACK mechanism [2], which can add an extra delay between a data segment and its ACK. Most TCP implementations introduce extra delays between 100 and 200 milliseconds, and this range of values is clearly reflected in the histogram of wired LAN maximum OSTTs shown in the right plot of Figure 5. This issue, which has received little attention in the literature, makes non-robust statistics, such as the average, poor representatives of the network delay experienced by a connection. The left plot in Figure 5 illustrates this point, The distribution of median OSTTs is far lighter than that of average OSTTs The average is far more affected than the median by the OSTT observations from delayed ACKs. In the rest of this study, we will rely on robust statistics in order to reduce the impact of delayed ACKs on our results[2].

Comparing the distributions of median OSTTs for wired and wireless clients, shown in Figure 5, reveals that packets experienced significantly higher delays in the wireless LAN. For example, while there is only a small fraction of median OSTTs above 10 milliseconds for connections on the wired LAN, on the wireless LAN 80% of the median OSTTs are above 10 milliseconds, and 40% are above 25 milliseconds. One interesting question is whether these higher OSTTs observed on the wireless LAN have a significant impact on round-trip times. The left plot in Figure 6 sheds light on this question by showing the distribution of the fraction of the round-trip time represented by the LAN side of each connection. We calculated this fraction using two definition of round-trip time: sum of the averages of $\{d_i^{lan}\}$ and $\{d_i^{wan}\}$, and sum of the medians. Given the measurement artifact created by TCP's delayed ACKs, the sum of the medians provides a much more reliable estimate of the network-layer round-trip time of a connection. The plot shows that the fraction of the RTT due to the wireless LAN is far more significant than that due to the wired LAN. For example, the LAN represented at most 10% of the RTT for 80% of the connections from wired clients, while this number goes down to 49% of the connections for wireless clients.

The middle plot of Figure 6 provides a good illustration of the much larger delay variability that packets experienced in the wireless LAN. The median absolute deviation, MAD [10], shows a very substantial difference between the wired and the wireless LANs. 80% of the connections from the LAN had a MAD of at most 100 microseconds, while this was only the case for 5% of the connections from the wireless LAN. More importantly, we observe that 13% of the connections show a MAD above 10 milliseconds, and even 50% of them above 40 milliseconds. These are already rather substantial numbers, which demonstrate that the wireless LAN has far higher network jitter than the wired LAN. Furthermore, this variability is substantially larger than the one observed for the WAN side of connections, as shown in the right plot of Figure 6. Note that the two measures of dispersion provide very consistent results for both types of clients, and that traversing the wireless LAN introduced substantially more variability than traversing the WAN.

[2]We are also working on a refi ned data-ACK coupling algorithm that can explicitly fi lter out observations from delayed ACKs.

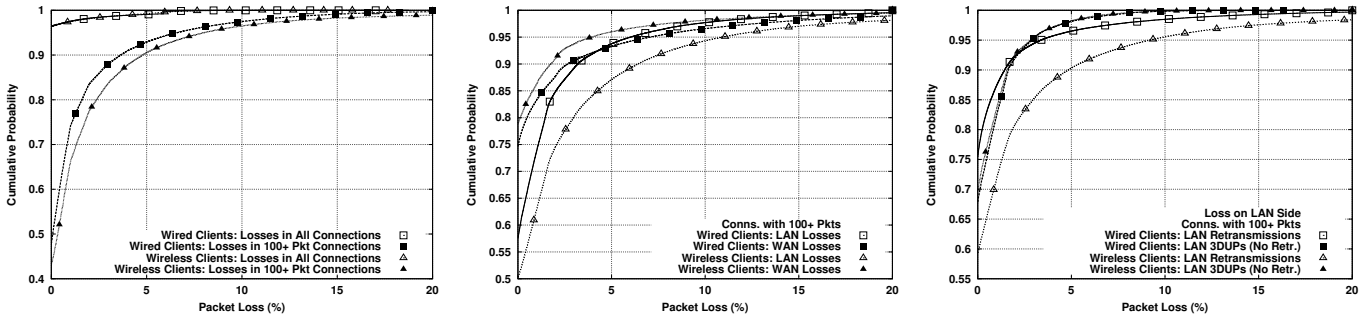

Fig. 7. Distributions of overall loss rates (left), and loss rates broken down by connection side (middle) and by type of loss event (right).

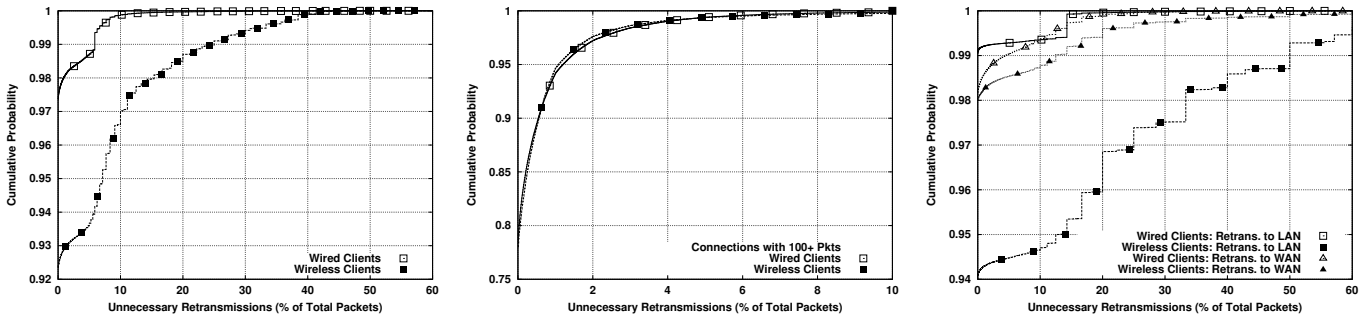

Fig. 8. Distributions of unnecessary TCP retransmissions in all connections (left) and in connections with 100 packets or more (middle), and breakdown by direction of the retransmission (right).

IV. Packet Loss

Wireless networks are expected to have far higher loss rates than their wired counterparts. However, 802.11 makes use of LL-ACK mechasnim to reduce the loss rates observed by higher network layers. The obvious question is how successful this mechanism is in the real world. As in Section III, we can rely on TCP dynamics to study the loss rates observed in our dataset, and compare wired and wireless clients. While more sophisticated methods for loss inference exist (see [8] for example), in this section we rely on two basic indicators of loss: retransmissions and triple duplicate ACKs. In general, retransmissions of TCP data segments are due to packet losses, and therefore the percentage of retransmitted data segments in a TCP connection is expected to be highly correlated with its loss rate. Both the use of cumulative acknowledgments and of carefully estimated retransmission timers [11] are intended to avoid most, if not all, unnecessary retransmissions. Counting retransmissions is a very effective loss estimator when monitoring takes place on the endpoints (as in [12]). However, in our case, we observed packets either after the LAN or after the WAN. This means that loss may occur prior to our monitoring point, making the retransmission count less than the true number of losses. There are a number of ways of addressing this difficulty, but we have chosen to rely on a simple one. To correct the loss rate estimate from retransmissions, we counted the number of triple duplicate ACK events (3DUPs). Note that we used 3DUPs to increase the loss rate estimate only when no retransmission were observed, and that we only counted each sequence of three or more duplicate ACKs once. Our estimate of the loss rate is therefore the ratio between the number of retransmissions plus the number of 3DUPs events (without an observed retransmission), and the total number of segments in a connection. We certainly do not claim this is the ultimate loss estimate, but we believe it is appropriate for comparing the loss rates of connections from wired and wireless clients in a fair manner.

The overall results of our loss analysis are shown in the left plot of Figure 7, where we compare the measured loss rates for TCP connections from wired and wireless clients. The first two distributions came from all connections, and we can hardly find any difference between them. In contrast, the distributions for connections with 100 packets or more shows slightly higher loss rates for connections from wireless clients. For example, loss rates above 2% were observed for 17% of the connections from wired clients, while this was the case for 23% of the connections from wireless clients. One conclusion we can obtain from these numbers is that the wireless LAN increased loss rates very moderately. The 802.11 link-layer retransmission mechanism seems highly effective in our enviroment. This observation is consistent with the high delay variability discussed in Section III, which suggests that plenty of losses were recovered. It is also important to keep in mind that while the overall impact of the WLAN in terms of loss rates seems small, it may be much higher under specific conditions (*e.g.,* during roaming, for access points with substantial loads, *etc.*). This is the subject of our follow-up work.

The middle plot in Figure 7 studies the loss rates separately for the two directions of the connections. We counted the retransmissions of packets flowing from the WAN to the LAN (loss after the monitor), and the 3DUPs observed in the LAN to the WAN direction (loss before the monitor), as loss in the LAN. This is by no means always accurate, since loss of ACKs may confused this heuristic identification of the side in which the loss happened. Despite this shortcoming, the comparison of the distributions in plots reveals that significantly larger loss

rates on the LAN side of the connections from the wireless clients. The right plot in Figure 7 divides the LAN loss rates into retransmissions and 3DUPs, revealing that the measured higher loss rates for wireless clients come mostly from counts of retransmissions.

Previous work has identified the high delay variability that the 802.11 link-layer retransmission mechanism creates can have a very detrimental effect on TCP's retransmission timeout estimation [13], [5]. We have examined whether this phenomenon is of any significance in our production wireless network using packet header traces, and our results are shown in Figure 8. Our count of unnecessary retransmissions was the number of times we observed a retransmitted data segment with a sequence number already covered by the cumulative acknowledgments flowing in the opposite direction. Each of these cases represents, unambiguously, a retransmission that did not contribute anything to the communication between the two endpoints, since the receiving end had already received (and acknowledged) the data. The left plot shows that the number of unnecessary retransmissions is significantly larger for connections from the wireless LAN when we consider all of the TCP connections in our dataset. Note however that the middle plot (for connection with 100 packets or more) shows little difference between connections from the wired and from the wireless LAN. We hypothesize that these longer connections give TCP's RTO estimator enough samples to settle on a more conservative timeout estimate that significantly reduces the number of unnecessary retransmissions. This observation seems to suggest that efforts to improve RTO calculations in the presence of a wireless link should focus on short connections, and probably on TCP's slow start phase.

V. Incomplete Connections

The previous two sections compared the delay and loss characteristic during the lifetime of TCP connections. Despite the carefully engineered mechanisms in modern network protocols, it is possible to encounter situations in which this mechanisms fail to maintain communication, and connections have to be closed. An interesting question is therefore whether the use of wireless LANs results in a higher number of interrupted connections. In principle, a complete TCP connection must be terminated with a two-way exchange of FIN segments [7]. However, it is common to observe connections that do not terminate in this manner. This can be due to several reasons, but there are three particularly relevant ones for our discussion. First, TCP implementations often limit the maximum number of times a packet can be retransmitted. If this limit is reached, the TCP endpoint gives up and resets the connection. Second, if a host is suddenly disconnected from the network, the offline endpoint cannot perform connection termination. This is expected to be relatively frequent in wireless environment, where laptops and PDA can easily be powered down without giving TCP a chance to close any existing connections. Third, problems with access point radio coverage and 802.11 roaming mechanisms (which we have observed repeatedly) can also result in interrupted TCP connections. Given the plausible causes of connection interruption, we wondered whether our dataset unveiled a higher incidence of interrupted TCP connections for wireless clients.

Our analysis of the entire set of TCP connections revealed that only 0.35% of connections from wired clients and 0.34% of connection from wireless clients lacked FIN and RST segments. These cases come from the time boundaries of the tracing period, which necessarily results in some partially captured connections. This is not expected to happen in the three cases above, where one of the endpoints will generally be able to reset the connection after determining that it became unviable. We examined the connections in our dataset that were closed only in one direction (*i.e.*, at least one FIN or RST segments were sent by only one of the endpoints), and found much more significant percentages. Connections from wired clients were closed only from one direction in 14.01% of the cases, while those from wireless clients were similarly closed in 23.00% of the cases. These nine percentage points do seem significant. We further examined these connections and found the following:

- 747,571 and 1,670,204 connections were closed only from the LAN (*i.e.,* from UNC) for wired and wireless client respectively.
- 149,173 and 1,442,533 connections were closed only from the WAN for wired and wireless client respectively.

The second bullet is especially significant, since it would correspond to connections where the problematic endpoint is located at the university. The much higher number for wireless clients clearly indicates that the interrupted connections are frequent in wireless environments. Further analysis is needed to understand the real implications of these interrupted connections (*e.g.,* interrupting a idle persistent HTTP connection after the browsing is over has no real impact on user experience).

VI. Conclusions and Future Work

We presented a large-scale passive measurement study of the characteristics of TCP connections, in terms of their volumes, delays, losses and lack of termination. Our goal was to draw a contrast between connections from wired clients and those from wired clients. In general, we found that the wireless network introduced substantially higher delay variability, but that its loss rates were only marginally above those observed for the wired LAN. We also quantified unnecessary retransmissions, which are significantly more frequent for wireless clients. Finally, we analyzed the termination of the connections in our dataset, and found a large number of connections for which the wireless client did not take any action terminate the connection.

The focus of our current work is to identify more precisely the operational causes of the results found in this paper. First, we are working on correlating the performance of individual connections with additional data that was collected concurrently. In particular, we used SNMP and the syslog service to gather information about AP load, client mobility, *etc*. Second, we are conducting active measurement experiments in which we try to verify some of the hypothesis put forward in our paper. We are specially interested in the evaluating the effectiveness of existing mobility mechanisms, and how they affect the network, transport and application layers.

REFERENCES

[1] Jay Aikat, Jasleen Kaur, F. Donelson Smith, and Kevin Jeffay. Variability in tcp round-trip times. In *IMC '03: Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, pages 279–284, New York, NY, USA, 2003. ACM Press.

[2] M. Allman, V. Paxson, and W. Stevens. RFC 2581: TCP congestion control, 1999.

[3] R. Droms. RFC 2131: Dynamic host confi guration protocol, March 1997.

[4] Matthew S. Gast. *802.11 Wireless Networks: The Definitive Guide*. O'Reilly & Associates, Inc., Sebastopol, CA, USA, 2002.

[5] Andrei Gurtov. Effect of delays on tcp performance. In *PWC '01: Proceedings of the IFIP TC6/WG6.8 Working Conference on Emerging Personal Wireless Communications*, pages 87–108, Deventer, The Netherlands, The Netherlands, 2001. Kluwer, B.V.

[6] F. Hern´andez-Campos. *Generation and Validation of Empirically-Derived TCP Application Workloads*. PhD thesis, University of North Carolina at Chapel Hill, 2005.

[7] Information Sciences Institute. RFC 793, 1981. Edited by Jon Postel. Available at http://rfc.sunsite.dk/rfc/rfc793.html.

[8] Sharad Jaiswal, Gianluca Iannaccone, Christophe Diot, and Donald F. Towsley. Inferring tcp connection characteristics through passive measurements. In *INFOCOM*, 2004.

[9] P. Karn and C. Partridge. Improving round-trip time estimates in reliable transport protocols. In *SIGCOMM '87: Proceedings of the ACM workshop on Frontiers in computer communications technology*, pages 2–7, New York, NY, USA, 1988. ACM Press.

[10] National Institute of Standards and Technology. Engineering statistics handbook, 2005.

[11] V. Paxson and M. Allman. RFC 2988: Computing TCP's Retransmission Timer, November 2000.

[12] Vern E. Paxson. *Measurements and Analysis of End-to-End Internet Dynamics*. PhD dissertation, University of California, Lawrence Berkeley National Laboratory, April 1997.

[13] B. Rathke, M. Schlaeger, and A. Wolisz. Systematic measurement of tcp performance over wireless lans. Technical Report TKN-98-001, Telecommunication Networks Group, Technische Universitt Berlin, December 1998.

[14] Richard W. Stevens. *TCP/IP Illustrated, Volume 1: The Protocols*. Addison-Wesley, 1994. StEV w 94:1 1.Ex.

[15] M. Yarvis, K. Papagiannaki, , and W.S. Conner. Characterization of 802.11 wireless networks in the home. In *1st workshop on Wireless Network Measurements (Winmee)*, April 2005.