

Understanding End-user Perception of Network Problems

J. Scott Miller
Northwestern University
jeffrey-
miller@northwestern.edu

Amit Mondal
Northwestern University
a-
mondal@northwestern.edu

Rahul Potharaju
Northwestern University
rahull@northwestern.edu

Peter A. Dinda
Northwestern University
pdinda@northwestern.edu

Aleksandar Kuzmanovic
Northwestern University
akuzmal@northwestern.edu

ABSTRACT

It is widely assumed that certain network characteristics cause end-user irritation with network performance. These assumptions then drive the selection of quality of service parameters or the goals of adaptive systems. We have developed a methodology and toolchain, SoyLentLogger, that employs user studies to empirically investigate such assumptions. SoyLentLogger collects client-centric network measurement data that is *labeled* by the end-user as being associated with irritation at perceived network performance (or not). The data collection and labeling occurs in real-time as the user normally uses the network. We conducted a study that tracked 32 ordinary users over a period of 3 weeks and then used that data to test common assumptions about network sources of user irritation.

Categories and Subject Descriptors

C.2.0 [Computer-communication Networks]: General;
H.1.2 [Information Systems]: Models and Principles—
User/Machine Systems

General Terms

Measurement, Human Factors

Keywords

User Studies, Network Performance, Empathic Systems

1. INTRODUCTION

What network characteristics lead to end-user irritation? There is surprisingly little work that addresses this question, though recent work [5] has suggested that user irritation is frequent and pervasive. Engineers typically try to minimize irritation through the use of application- or network-level

This project is made possible by support from the National Science Foundation (NSF) via grant CNS-070691.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

W-MUST'11, August 19, 2011, Toronto, Ontario, Canada.
Copyright 2011 ACM 978-1-4503-0800-7/11/08 ...\$10.00.

quality of service measures, but these are based on assumptions, *rules of thumb*, for what network characteristics lead to irritation.

We have *empirically* addressed this question, and evaluated a range of rules of thumb, through a careful user study. In this paper we summarize this work. A detailed supporting technical report is also available [7]. Our study produced client-centric network measurement data that was *labeled*, in real-time by the participants in our study, as being irritating or not. For any given assumption or rule of thumb, we can conduct a hypothesis test via a query over the labeled network data. As far as we are aware, ours is the first study to explicitly correlate user irritation due to perceived network problems with low-level network measurements.

This work is complementary to efforts to understand user-perceived quality in particular application domains (e.g., [8]), to create better methodologies for finding mean opinion scores for particular domains (e.g., [1]), or to introduce user-perceived quality into network-level decisions (e.g., [6]). We seek to provide more information to help understand the relationship between user irritation and network characteristics in the broader context of the extant Internet.

2. SoyLentLogger

Users in our study ran SoyLentLogger on their personal Windows-based laptop computers. We developed SoyLentLogger to measure the network and the local machine at regular intervals, and in response to user irritation events. SoyLentLogger is designed for use on Windows XP/Vista/7 with the .NET 2.0 platform, and thus maximizes the number of possible participants in any study. The core of the system runs as a Windows service, and the system can update itself, allowing its use for different studies.

SoyLentLogger periodically uploads data to our server when the size of the local log exceeds a limit. The 95th percentile network overhead, which includes uploading study data and checking for updates, is less than 10 kbps in both the upstream and downstream directions. The 95th percentile memory and CPU overhead are also negligible, and well within the comfort ranges found in an earlier study [3].

Figure 1 illustrates the network, host, and user measurements made in SoyLentLogger. SoyLentLogger allows users to express their irritation using a globally mapped key (F8) or a “system tray” menu item. We refer to a user’s expression of irritation as an “irritation event.” When an irritation event occurs, SoyLentLogger immediately collects data that would otherwise be collected periodically. This allows us to

Factor	Period	Details
CPU Utilization	1 second	Aggregate system load.
Process Statistics	30 seconds	Per-process performance and resource utilization data including cumulative time in execution, working memory size, and number of page faults.
Application Focus	10 seconds	The name and title of the current window in focus.
User Activity	30 seconds	Total mouse movement, button clicks, and keystrokes in the last interval.
Offered Throughput	5 seconds	The average offered throughput of each flow during the collection period.
Application RTT	Continu.	The elapsed time between the transmission of a TCP packet and the arrival of the acknowledging packet.
Receiver Signaling	Contin.	TCP packets for which the receiver window is set to zero, potentially indicating load on a remote server.
Duplicate Packets	Contin.	TCP packets repeated by either the sender or receiver.
Web Traffic	Contin.	The URL and method of each HTTP request.
Link	5 seconds	The properties of the current network interface being used, such as IP and link speed.
Wireless Interface	5 seconds	Additional information on the wireless link including base station MAC address and signal strength.

Figure 1: The hardware, OS, network, and user metrics logged by SoylentLogger.

capture some transient events that might go undetected by the periodic monitoring. SoylentLogger also collects information that captures the user context, such as the user’s current task (e.g., foreground window) and attentiveness (e.g., mouse travel). Among other things, this information lets us focus on network conditions that occur when the user is *physically present* at the machine. Rapid-fire sequences of key-presses are aggregated in our analysis to avoid a single actual irritation event appearing to be multiple events.

SoylentLogger examines all outgoing and incoming packets from the user’s network interface. From them, we derive connection-level behavior, such as the instantaneous round-trip-time, throughput, and loss rate. We further tag each connection with the the parent application that owns its socket. In addition to measuring the user’s traffic, SoylentLogger also periodically initiates ping and traceroute probes to the 25 most frequently accessed IP addresses to assess the end- and remote-host’s connectivity.

3. USER STUDY AND DATA

Our user study was approved by Northwestern’s IRB, allowing us to recruit across the university. We recruited 32 participants from a pool of students, staff, and faculty using a combination of broad flyer and email advertising. The study group, primarily undergraduate students, received network connectivity via both the university and a number of regional ISPs. After signing up for our study, participants visited our lab to have SoylentLogger installed on their personal machines by the investigators. Each filled out a background questionnaire. Our subjects use a variety of network services. They are generally non-technical users, with only a few having a CS, CE, or EE.

The participants read an overview of the study¹ that included the following instructions: “We ask that you press (the irritation button) when you are uncomfortable or dissatisfied with the network service being provided to the applications you are using.” The document also gave several examples of network performance issues ranging from slow-loading web content to interrupted video and audio streaming. Our document stresses that the users are to signal irritation with the *performance* of the network applications being used and not the *content* of those applications. This was reinforced, and the participant continually reminded of

¹The materials used during the study can be found at <http://empathicsystems.org/SoylentLogger>

the study, by affixing a small sticker to the participant’s laptop next to the relevant key that read “Press F8 when irritated by the network.”

The period of the study was three weeks. During this time, the participants used their laptops as normal, with the addition of registering irritation. After completing the study, each participant was paid \$25 for his or her time.

The 32 users in our study generated 899 irritation events and 20.0 GB of data over the three week period. This is about 1.2 events per user per day, although there is considerable variation in irritation rate across users. Irritation events were dispersed across the three weeks for all users. The interarrival time distribution for irritation events from an active individual user appears to fit a power law. End-users are apparently frequently dissatisfied with their network usage—50% of irritation events occur within 17 minutes of a previous irritation event.

4. HYPOTHESES

In evaluating hypotheses we use a common methodology for correlating factors with irritation events. If a factor correlates with irritation, we assume it changes near the irritation event. To make the meaning of “near” precise, we define two parameters that together specify a window of time around an irritation event in which the factor may change and be considered correlated with the event. The window is ω seconds long and is displaced from the irritation event by an interval τ . ω can be thought of as how long the change must persist to cause irritation, while τ is the delay between the cause and the user’s reaction. Parameters ω and τ define the “irritation window” associated with an irritation event.

Our methodology considers the user’s presence: we only analyze log data for which we know the user is present, based on the keyboard and mouse activity we recorded. We filter out data logged outside of a 60 second window centered around each instance of user activity. Generally, we evaluate a hypotheses by comparing the distribution of a given factor outside of all irritation windows with the distribution of it occurring within irritation windows.

We evaluated the sensitivity to ω and τ , although we omit the presentation of the results for τ , as the sensitivity is much weaker than to ω . In our presented results, we set τ to 1 second.

Hypothesis 1: Users can distinguish between local and network sources of irritation.
Result: Supported by our evidence.

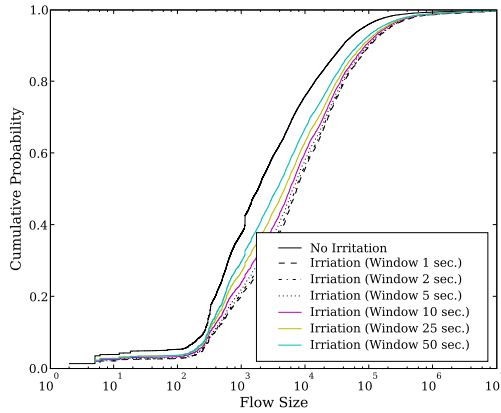


Figure 2: Flow size distribution for a range of window sizes. Irritation events are associated with larger flows, on average, than flows not associated with irritation. However, the absolute size of these flows is not dramatically different.

For each of CPU utilization, page fault rate, and aggregate network throughput utilization, we compared its distribution during irritation windows and during times outside of irritation windows. While the CPU and page fault distributions did not significantly vary between these two categories, the distribution of aggregate network throughput did. The median throughput during irritation windows is an order of magnitude higher than the median throughput outside of those windows. Note further that given our methodology, we are considering throughput behavior that *precedes* the irritation event; higher throughput is followed by irritation. Our technical report [7] contains a detailed analysis, which is also supported by our earlier work that considered irritation due to CPU, memory, and I/O load [3].

Of course, there are other possible explanations, but the result certainly can be readily explained if users can indeed successfully distinguish between local and network sources of irritation.

Hypothesis 2: Most irritation is associated with small flows.

Result: Supported by our evidence. Further observations.

It is widely assumed that small (low byte count) flows are critical to the end-user experience and that the poor performance of small flows dominantly affects users’ perception of the network service. As a result, the performance of small flows has traditionally been one of the key QoS metrics (e.g. [2] driving the development of techniques to optimize behavior according to remaining bytes in a flow (e.g., [11]). We find that while the majority of the connections associated with irritation are quite small, connections present during irritation skew to larger sizes and longer durations.

Figure 2 compares the distributions of flow sizes both during irritation and not. The median flow size is 2.8 times larger during irritation, although the absolute size of these flows is still less than 10 KB. As shown in Figure 3, the flow *duration* during irritation is considerable longer, with the median duration 34.6 times larger during irritation. User irritation is most closely associated with small flows that are *long-lived*, which might be termed the *lethargic mice*.

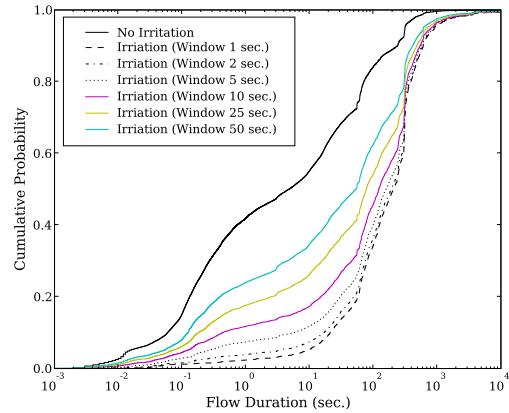


Figure 3: Flow duration distribution for a range of irritation window sizes. When considering window sizes less than 10 seconds, the distribution of flow durations is substantially different during irritation events.

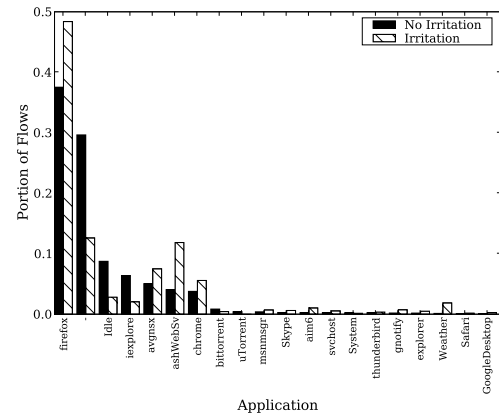


Figure 4: The portion of non-irritation and irritation traffic associated with each application. Firefox, Chrome, and the Avast Internet Security Suite (ashWebSV) are associated a higher proportional of flows during irritation than not, while idle system activity is less likely to associated with irritation.

Hypothesis 3: User irritation is dependent on the application and services with which that user interacts.

Result: Supported by our evidence.

Figure 4 plots the overall percentage of flows attributed to each of the common applications seen in our study, along with the percentage of flows associated with irritation attributed to each application. Here “-” indicates that we were unable to determine the application before the flow terminated. Clearly some applications have flows that are disproportionately associated with irritation events. Almost 40% of the flows seen in our study are generated from Firefox, and 75.8% of the flows in our study come from web traffic. Surprisingly, we find that Internet Explorer has a lower rate of irritation as compared the other browsers. We hypothesize that this is due to participants in our study using different browsers for different sites and services. As we will

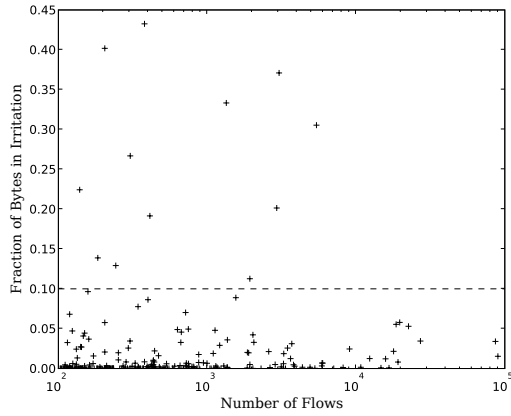


Figure 5: For each destination AS, the number of flows seen to the AS is plotted versus the fraction of bytes associated with irritation events. Note that even for destinations for which we have considerable data, the rate of irritation can be very high.

show, the choice of service plays an important role in user irritation.

Given the volume of the web traffic, we looked deeper at how irritation varied by the destination AS or service provider. To determine the web application in use, we used the IP to Autonomous System Number (ASN) maps provided by Cymru [9], which cover 96% of our flows.

To compare the amount of irritation associated with each autonomous system (AS), we consider the amount of data associated with irritation. As before, we associate a flow with an irritation event if that flow is present during the irritation window specified by that event. If a flow is associated with irritation, we consider *all* bytes of that flow as irritation bytes. Thus, the irritation rate is the number of bytes in all such flows divided by the total number of bytes transferred to or from that AS.

Figure 5 plots, for each of the 284 ASes that have over 100 flows, the number of flows seen with that AS as a *destination* and the fraction of all bytes transferred to or from that AS that are associated with an irritation event. We find that there are a large number of ASes with substantial per-byte irritation rates. 12 of these ASes have over 10% of their traffic associated with irritation, which implies a disproportionate amount of user irritation.

Figure 6(a) shows the top 5 ASes in terms of traffic. There is both considerable data for all of these providers and visible stratification among them. For example, while Level 3 Communications and Limelight Networks provide similar content delivery services, the two have very different irritation rates, with Level 3’s irritation rate being 69 times greater. Figure 6(b) shows the top 3 ASes (of those with more than 1000 flows) in terms of irritation rate. These hosts show very high rates of irritation emanating from a small number of large flows. While these ASes represent less than 5.1% of traffic, they make up 48.9% of all bytes associated with irritation.

Hypothesis 4: Users are more sensitive to the network while using streaming applications.

Result: Not supported by our evidence.

To detect streaming, we consider flows that are associ-

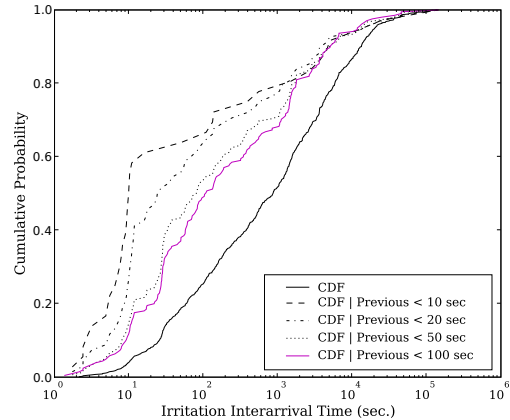


Figure 7: The distribution of per-user irritation interarrival times given that the previous interarrival time is less than some threshold.

ated with a web browser, that use remote port 80, and are larger than 1 MB. Next, we filter this set using a traffic classification method described elsewhere [10]. We then validate these flows by mapping their destination IP addresses to ASNs. We find that more than 90% of the flows we select have remote IP addresses that belong to content delivery networks or content-generating companies.

We compare the irritation events per hour for these streaming flows to the rate of non-streaming flows. We associate a user irritation event with streaming if it overlaps with a streaming flow associated with that particular user. We thus count the total number of irritation events that occur during streaming. We also compute the aggregated streaming time by summing up individual streaming flow durations. We ensure that we count the overlapping periods only once. Finally, we divide the total number of irritation events associated with streaming by the total streaming time to get the metric, the irritation rate during streaming. To compute the irritation rate for the non-streaming case, we divide the number of irritation events that are not associated with streaming by the total amount of time the users were active during which there were no active streaming flows.

The irritation rate during streaming is 0.41/hour, which is about half that seen outside of streaming (0.81/hour). This is quite surprising. While network behavior does indeed induce considerable irritation during streaming, it apparently induces far more irritation for other kinds of flows. It is important to note further that there is also very high variation across users both in the time that a user spends using streaming applications and the irritation rate for streaming and non-streaming flows.

Hypothesis 5: User irritation is stateful.

Result: Supported by our evidence.

To characterize the extent to which user irritation is stateful, we consider triples of sequential irritation events. We plot the distribution of the interarrival time between the second and third events given that the interarrival time of the first and second is below some threshold. We then vary this threshold for a range of values. If user irritation is indeed stateless, then these distributions should not differ.

Figure 7 plots the distributions. It is clear that irritation rate is indeed influenced by the prior knowledge of the

Host	No Irritation		Irritation		Total Traffic (MB)	% Bytes in Irritation
	Traffic (MB)	Flows	Traffic (MB)	Flows		
Google Inc.	8402.24	85133	295.85	1376	8698.08	3.40
Comcast Cable Communications Inc.	6475.09	7084	0.04	3	6475.13	< 0.01
Northwestern University	4242.61	88970	66.10	908	4308.71	1.53
Level 3 Communications	3988.20	18024	234.10	582	4222.30	5.54
Limelight Networks Inc.	3155.00	14608	2.51	110	3157.51	0.08

(a) Top-5 ASes by traffic volume.

Host	No Irritation		Irritation		Total Traffic (MB)	% Bytes in Irritation
	Traffic (MB)	Flows	Traffic (MB)	Flows		
Advanced Video Communications Inc.	767.08	3032	452.35	10	1219.43	37.10
Global Crossing Ltd.	480.51	1325	240.20	19	720.71	33.33
NTT America Inc.	559.78	5379	246.13	45	805.91	30.54

(b) Top-3 ASes by irritation rate.

Figure 6: Traffic quantity and irritation rates for a selection of ASes. ASes with seemingly identical responsibilities show considerable variation in their irritation rates. A small number of ASes are associated with large amounts of irritation.

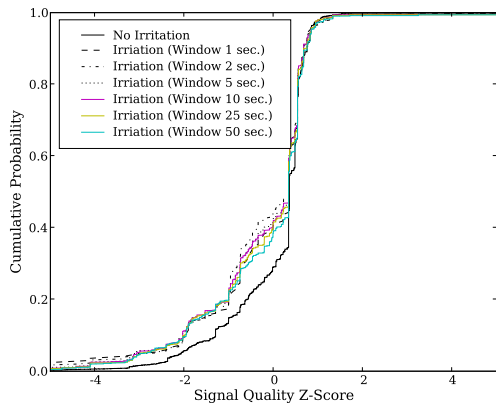


Figure 8: Windows “Signal Quality” metric as we sweep through a range of window sizes.

user’s irritation rate, implying that user irritation is stateful. When a user expressed irritation twice in the preceding 10 seconds, the next irritation event is likely to be generated within 10 seconds in 60% of the cases. This implies that irritation, once caused by the network, tends to persist for the user.

Hypothesis 6: RSSI and link quality indicators predict user irritation on wireless networks.
Result: Not supported by our evidence.

We next explore the correlation between irritation events generated during the study and the condition of the local wireless network. To do this, we consider two metrics: the “Signal Quality” metric provided by Windows and the received signal strength indicator (RSSI) exposed by the 802.11 hardware. Overall we find that irritation events are most correlated with the “Signal Quality” metric and that disruptions are not transient.

Figure 8 plots the distribution of the normalized Windows “Signal Quality” metric both in the presence of irritation and not for a range of ω . The difference between the distributions is maximized when $\omega = 2$ seconds, though the signal quality metric is not sensitive to ω . This, along with the fact that signal quality is consistently lower near irritation

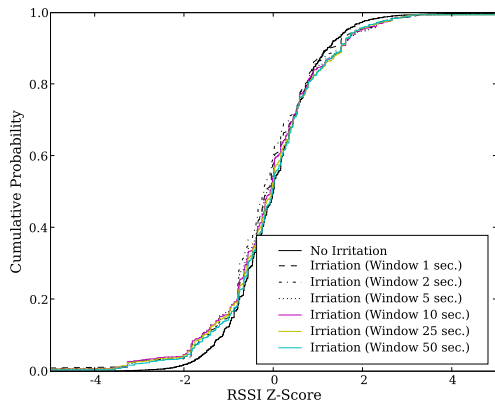


Figure 9: The distribution of normalized received signal strength reported by the wireless interface during irritation. Irritation events associated lower signal strength represent < 20% of irritation events at wireless access points.

events, implying that while users are more likely to express irritation if signal quality is decreased, it does not appear that this is caused by transient physical-layer disruptions.

In Figure 9 we plot a distribution of normalized² RSSI values around irritation along with a baseline. We see the largest difference when the RSSI value is more than one standard deviation below the average RSSI value for that hardware. However, fewer than 20% of the measured RSSI values during irritation events fall into this category. Once again, we see little sensitivity to the choice of ω value, implying that the irritation is unlikely to be caused by transient changes in signal strength.

It may appear that “Signal Quality” is a better metric because of its stronger correlation with user irritation. However, when we consider how well signal quality *predicts* irritation, the measure does little good. In Figure 10 we show the false-positive and false-negative rate for a threshold-based predictor using various signal quality values. The figure

²Normalization was needed as 802.11 does not define the units of RSSI [4]. We use a z-score normalized for each user.

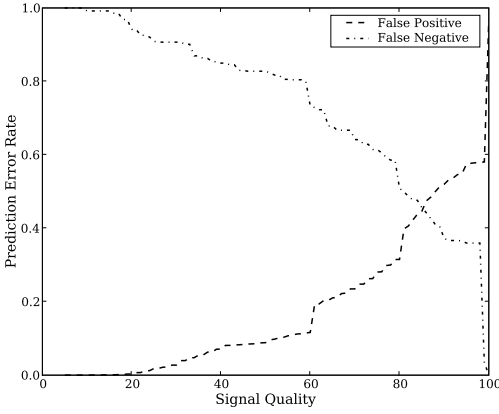


Figure 10: The predictive power of the Windows “Signal Quality” for user irritation, using a threshold-based predictor, as a function of the threshold. There is no threshold which provides simultaneously low false negative and positive rates.

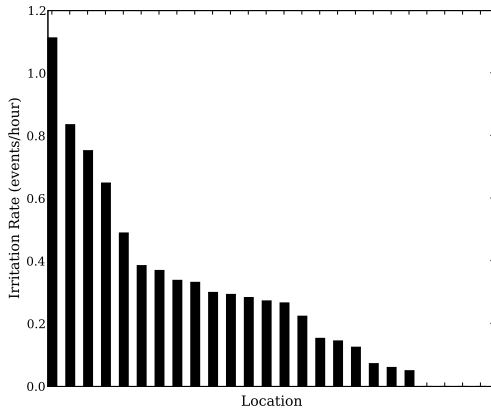


Figure 11: The rate of irritation events for the top-25 most frequently visited access points, sorted in order of decreasing irritation rate.

shows that there is no threshold which provides low false negative and false positive rates simultaneously. Thus, while important, signal quality is not a strong predictor of user irritation by itself. We are *not* claiming that wireless performance is irrelevant to user irritation, but that these metrics are not good predictors of it.

Hypothesis 7: User irritation is affected by user location.
Result: Supported by our evidence.

Finally, we consider the extent to which irritation is associated with wireless access points. Figure 11 shows the rate of irritation for the 25 most frequently visited access points, each having at least 5 hours of user activity. If each access point were equally likely to be associated with user irritation, we would expect a uniform distribution; however, this is not the case. Also, across all access points for which we have more than 1 hour of trace data, the top 20% of locations in terms of irritation rate are responsible for 64% of the overall irritation rate. Improving service at a small sub-

set of locations may result in a disproportionate reduction in total irritation.

5. CONCLUSIONS

We presented a tool and a methodology for collecting and studying end-user irritation with the network “in the wild.” We used the data we collected from an extensive user study to test a range of assumptions or rules of thumb that are commonly made in network control systems or adaptive applications. The most important implications of our work so far are that users are able to appropriately assign blame to the network when they are irritated, and that a small number of sources seem to disproportionately contribute to the irritation experienced by those users.

6. REFERENCES

- [1] CHEN, K.-T., TU, C. C., AND XIAO, W.-C. OneClick: A framework for measuring network quality of experience. In *Proceedings of IEEE INFOCOM 2009* (2009).
- [2] CHRISTIANSEN, M., JEFFAY, K., OTT, D., AND SMITH, F. Tuning RED for Web traffic. *IEEE/ACM Transactions on Networking* 9, 3 (2001), 249–264.
- [3] GUPTA, A., LIN, B., AND DINDA, P. A. Measuring and understanding user comfort with resource borrowing. In *Proceedings of the 13th IEEE International Symposium on High Performance Distributed Computing (HPDC 2004)* (June 2004).
- [4] IEEE COMPUTER SOCIETY. Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, June 2007. IEEE Standard 802.11, 2007 Edition.
- [5] KARAGIANNIS, T., ATHANASOPOULOS, E., GKANTSIDIS, C., AND KEY, P. HomeMaestro: Order from chaos in home networks. Tech. Rep. MSR-TR-2008-84, Microsoft Research, 2008.
- [6] LANGE, J. R., MILLER, J. S., AND DINDA, P. A. Emnet: Satisfying the individual user through empathic home networks. In *Proceedings of the 29th IEEE Conference on Computer Communications (INFOCOM)* (March 2010).
- [7] MILLER, J. S., MONDAL, A., POTHARAJU, R., DINDA, P. A., AND KUZMANOVIC, A. Network monitoring is people: Understanding end-user perception of network problems. Tech. Rep. NWU-EECS-10-04, Department of Electrical Engineering and Computer Science, Northwestern University, March 2010.
- [8] RIX, A. Perceptual speech quality assessment - a review. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on* (May 2004), vol. 3, pp. iii–1056–9 vol.3.
- [9] TEAM CYMRU. The IP to ASN Mapping Project. (web page). <http://www.team-cymru.org/Services/ip-to-asn.html>.
- [10] TRESTIAN, I., RANJAN, S., KUZMANOVIC, A., AND NUCCI, A. Unconstrained endpoint profiling: Googling the Internet. In *Proceedings of ACM SIGCOMM '08* (Seattle, WA, Aug. 2008).
- [11] YANG, S., AND DE VECIANA, G. Size-based adaptive bandwidth allocation: Optimizing the average QoS for elastic flows. In *Proceedings of IEEE INFOCOM '02* (New York, NY, June 2002).