Understanding End-User Perception of Network Problems

J. Scott Miller  Amit Mondal  Rahul Potharaju

Peter A. Dinda  Aleksander Kuzmanovic

Empathic Systems Project
empathicsystems.org

Northwestern University
Overview

• **Overall Question:** How does end-user *perception* of network performance correlate with low-level network measurements?

• **Approach:** User study that collects network, host, and user measurements on end-user machines, labeled by end-user with “irritation events”; comparison of labeled/unlabeled measurements

• **Context:** How to incorporate end-user satisfaction and guidance in computer systems, computer architecture (and network?) design?
SoylentLogger

- Windows service that monitors network, host, and user context and uploads to our server
  - Negligible network, cpu, and memory overheads on laptops
  - Packet-level inspection, connections tagged with applications
  - Periodic measurement, and irritation-driven measurement

- Measurements of
  - **User**: application focus, user activity, web traffic (URLs),
  - **Host**: CPU utilization, process statistics
  - **Network from the perspective of the host**: offered throughput, application RTT, receiver signaling duplicate packets, link properties, wireless interface properties, ping/traceroute probes
• User prep document focuses on network performance and states “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.”
User Study

- 32 users recruited using broad, IRB-approved advertising at Northwestern
  - Almost all non-technical users
  - $25 for participation
- SoylentLogger installed and tested by us on each user’s personal machine
- Controlled interaction with users
  - Users read standard preparatory document
  - Users told to use their machines normally
- Operation over same three week period
Data Set

• Immediately consecutive irritation events filtered
  • User mashing F8 repeatedly counts once

• Only data within 60 s of user activity considered
  • User must have opportunity to express irritation

• 20 GB of raw data

• 899 irritation events
  • ~1.2 events/user/day (varies across users, does not vary much across time)
  • Apparent power law interarrival times per user
  • 50% of irritation events occur within 17 minutes of a previous event
What does an irritation event label?

\( \omega \) : window of experience leading to event
\( \tau \) : delay from experience to keypress

We evaluate sensitivity of results to \( \omega \) and \( \tau \)

Most sensitivity is to \( \omega \)
Hypotheses Evaluated

**Supported**
- Users can distinguish between local and network sources of irritation
- User irritation is dependent on the applications and services with which the user interacts
- User irritation is stateful
- User irritation is affected by user location (wireless access point)

**Supported with other observations**
- Most irritation is associated with small flows

**Not supported**
- Users are more sensitive to the network when using streaming applications
- RSSI and link quality indicators predict user irritation on wireless networks
Hypotheses Evaluated

• **Supported**
  - Users can distinguish between local and network sources of irritation
  - User irritation is dependent on the applications and services with which the user interacts
  - User irritation is stateful
  - User irritation is affected by user location (wireless access point)

• **Supported with other observations**
  - Most irritation is associated with small flows

• **Not supported**
  - Users are more sensitive to the network when using streaming applications
  - RSSI and link quality indicators predict user irritation on wireless networks
Can users distinguish the network?

Greater difference than with page fault rate or CPU utilization

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

Result: Supported by our evidence.

We suspected that the network applications a user interacts with will vary in their association with irritation. This is a natural assumption as applications vary in their QoS requirements, resulting in some being more sensitive to disruptions in service. We also suspected that the content provider also plays a role in user irritation. For example, to enhance web browsing experiences, content distribution networks (CDN) move web content closer to clients by caching copies of web objects on thousands of servers worldwide. It has been demonstrated that this approach can help improve Web response times [58], and so it is assumed that CDNs positively impact users' perceived QoS. We find that this is not always the case as the rate of user irritation associated with greater difference than with page fault rate or CPU utilization.
Do Applications Matter?

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 the flows is still less than 10 KB. As shown in Figure 3, the flow duration during irritation is considerably 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 an 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 be 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 to 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
Does Destination AS Matter?

![Graph showing the relationship between the number of flows and the fraction of bytes in irritation.](chart.png)

**Destination ASes where >10% of flow bytes were associated with irritation events**

**Figure 5:** For each destination AS, the number of irritation bytes, given that there were at least 1000 flows associated with irritation. The x-axis represents the number of flows, and the y-axis represents the fraction of bytes in irritation.
Do WiFi Signal Quality Metrics Help?

- Not supported by our evidence
- NIC RSSI metric seems completely uncorrelated with user irritation events
- Windows signal quality metric is slightly correlated with user irritation events
But Windows Signal Quality Is A Poor Predictor...

No threshold detector provides simultaneously low false positive and false negative rates.
Yet Location (which WAP is used) Matters

Figure 10: The predictive power of the Windows "Signal Quality" for user irritation, using a threshold-based predictor, 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 subset 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


Conclusions

- Attempt to correlate user irritation due to perceived network problems with low level network measurement
  - Using a feedback mechanism that could be continuous employed
  - Contra QoE, OneClick, EmNet, Vienna Surfing, HostView, LRD,...
- More detailed technical report and study materials available online
- We are working on making data available
For More Information

• pdinda@northwestern.edu
• empathicsystems.org
• presciencelab.org

• Sponsored by NSF CNS-070691
Are short flows the critical ones?

Flow size (bytes) skews larger during irritation...

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 presented 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 considerably 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 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 to the other browsers. We hypothesize that this is due to participants in our study using different browsers for different sites and services.

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

Result: Supported by our evidence.
Are short flows the critical ones?

...but flow duration is where the action really is
# Top Five ASes By Traffic Volume

<table>
<thead>
<tr>
<th>Host</th>
<th>No Irritation</th>
<th>Irritation</th>
<th>Total Traffic (MB)</th>
<th>%Bytes in Irritation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>8402</td>
<td>85133</td>
<td>295</td>
<td>1376</td>
</tr>
<tr>
<td>Comcast</td>
<td>6475</td>
<td>7084</td>
<td>&lt;1</td>
<td>3</td>
</tr>
<tr>
<td>NU</td>
<td>4242</td>
<td>88970</td>
<td>66</td>
<td>908</td>
</tr>
<tr>
<td>Level 3</td>
<td>3988</td>
<td>18024</td>
<td>234</td>
<td>582</td>
</tr>
<tr>
<td>Limelight</td>
<td>3155</td>
<td>14608</td>
<td>3</td>
<td>110</td>
</tr>
</tbody>
</table>
## Top Three ASes By Irritation

<table>
<thead>
<tr>
<th>Host</th>
<th>No Irritation</th>
<th>Irritation</th>
<th>Total Traffic (MB)</th>
<th>%Bytes in Irritation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Video Commun.</td>
<td>767</td>
<td>3032</td>
<td>452</td>
<td>10</td>
</tr>
<tr>
<td>Global Crossing</td>
<td>480</td>
<td>1325</td>
<td>240</td>
<td>19</td>
</tr>
<tr>
<td>NTT America</td>
<td>560</td>
<td>5379</td>
<td>246</td>
<td>45</td>
</tr>
</tbody>
</table>

5.1% of observed traffic, but 48.9% of all bytes associated with irritation
Are Users Particularly Sensitive During Streaming?

- Not supported by our evidence

- Irritation events during times when at least one streaming flow exists: 0.41/hour

- Irritation events when no streaming flow exists: 0.81/hour

- We tag flows as “streaming” based on size, port, “Googling the Internet” technique, and destination ASN

- There are caveats
Is User Irritation Stateful?

Once the user is irritated he’s likely to stay irritated (or the irritating conditions are likely to persist)

Shorter conditioning intervals

Cumulative Probability

Irritation Interarrival Time (sec.)

[Graph showing cumulative probability distribution with different conditions]

Once the user is irritated he’s likely to stay irritated (or the irritating conditions are likely to persist)

Is User Irritation Stateful?

Once the user is irritated he’s likely to stay irritated (or the irritating conditions are likely to persist)

Shorter conditioning intervals

Cumulative Probability

Irritation Interarrival Time (sec.)

[Graph showing cumulative probability distribution with different conditions]