Understanding Users’ Perception of Performance Disruptions at End-Hosts

Diana Joumblatt, Oana Goga, Renata Teixeira
Laboratoire LIP6 -- CNRS and UPMC Sorbonne Universités

Jaideep Chandrashekar, Nina Taft
Technicolor
Network performance disruptions are frustrating.
User experience vs. typical network performance metrics

- Current state of the art
  - Raw network performance metrics
  - User perspective only in niche applications

- What’s missing?

User perspective of the overall online experience
Why user perspective matters?

- **Automated diagnosis**
  - Avoid unnecessary diagnosis when users don’t care
  - Improve diagnosis when network performance seems ok and the user is annoyed

- **Performance management**
  - Adapt application performance to user needs
Goal of our work

Automatically detect network performance disruptions as perceived by end-users
Challenges in measuring user perception

- User perception varies
  - Per user, per environment, per application
  - For a given user according to external factors

- Imbalance in number of samples
  - Can’t collect frequent user feedback (~10 per day)
  - Orders of magnitude more network measurements (~$10^3/10^4/10^5/\ldots$)

- End-host data collection raises issues
  - Privacy
  - Machine overhead
Outline

- Measure network performance with user feedback
- Characterize end-host performance across networking environments
- Correlate network performance with user feedback
Outline

- Measure network performance with user feedback
- Characterize end-host performance across networking environments
- Correlate network performance with user feedback
HostView overview

End-host

- User feedback
- System performance
- Network traffic
- Application context
- Network environment
- Traces
- Privacy
- Upload traces

Privacy

Every 4 hours
100 kbps max.

Front-end server

Network environment

Transfer to backend

Back-end server

Traces

Transfer to backend

Network environment

Traces
User feedback mechanisms

- **System Trigged feedback**
  - Experience sampling methodology (ESM)
  - Triggered based on state of machine
  - 5 short questions about network performance
  - At most 3 times a day

- **User Triggered feedback**
  - “I’m annoyed” button 😞
  - Same questions as in ESM
  - Can trigger as often as user wants
Example question

4. Did you experience any of the following problems, in the last five minutes? (Click all that apply)

- [ ] Can't connect to some sites or services
- [ ] Poor voice or audio quality
- [ ] Slow download or upload
- [ ] Slow browser
- [ ] Poor video quality
- [ ] Any other problem(s):
- [ ] None
Deployment

- Recruiting volunteers
  - Leaflets at IMC 2010 and CS Mailing lists
  - 50 USD Amazon gift cards
  - Real-time feedback about network connection

- Data: 40 users
  - Nov 2010 – Feb 2011
  - 26 Mac OS and 14 Linux
  - 14 countries
  - Most users ran tool for one month
Outline

- Measure network performance with user feedback
- Characterize end-host performance across networking environments
- Correlate network performance with user feedback
Network performance across environments

- **Goal**
  - Understand baseline network performance

- **Characterization questions**
  - How diverse are networking environments of a single user?
  - Does the set of applications vary across environments?
  - Does network performance vary across environments?
  - Which factors impact end-host performance?
Network environment

SSID1 – en0 – Free SAS – FR – Wireless – Home
SSID2 – en0 – TU Berlin campus – GR – Wireless – Conference
MAC1 – en1 – UPMC – FR – Ethernet – Work

(Home vs. Work)
(Home vs. Conference)
(Work vs. Conference)
Diversity of environments

- 5 interface types
- 10 user tags
- 12 countries
- 58 source ISPs
## Diversity of applications

<table>
<thead>
<tr>
<th>Multiple environments</th>
<th>Single environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook.com</td>
<td>Rhythmbox</td>
</tr>
<tr>
<td>Twitter.com</td>
<td>iTunes</td>
</tr>
<tr>
<td>Akamai.net</td>
<td>Hulu.com</td>
</tr>
<tr>
<td>Google.com</td>
<td>Netflix.com</td>
</tr>
<tr>
<td>Skype</td>
<td>VLC</td>
</tr>
<tr>
<td>Mail</td>
<td>Speedtest.net</td>
</tr>
<tr>
<td>Wordpress.com</td>
<td>uTorrent</td>
</tr>
<tr>
<td>ssh</td>
<td>openvpn</td>
</tr>
</tbody>
</table>
Method

- Performance metrics per connection
  - RTTs
  - Download data rates

- Application and environment per connection

- How different is performance in two environments?
  - Compare distributions of each metric between environments
Hellinger distance

**Definition**
- $f_{\text{home}}^{\text{down}}(x)$: Empirical probability distribution that a download data rate measurement will take value $x$ in environment home

$$
HD_{\text{home/work}}^{\text{down}} = \sqrt{1 - \sum_{x \in X} \sqrt{f_{\text{home}}^{\text{down}}(x) f_{\text{work}}^{\text{down}}(x)}}
$$

**Advantages**
- Measures distance between 2 distributions using the full density not just two points
- Bounded to [0,1]
\[ \text{HD}_{ij} = 0.17 \]

- If \( \text{HD}_{ij} < 0.1 \), then distributions are similar
- If \( \text{HD}_{ij} > 0.3 \), then distributions in \( i,j \) significantly different
Comparing performance across environments

- Select environments with at least 5000 samples
  - 21 users
  - 164 environment pairs

- Hellinger distance for all possible pairs of environments for a single user
  - Home vs. work
  - Airport vs. hotel
  - Home vs. friend’s home
  - Work vs. other workplace
  - ...
Does network performance vary per environment?

RTTs and data rates differ per environment.
Does network performance vary per environment?

17 out of 21 users experience high data rate variation across environments.

20 out of 21 users experience high RTT variation across environments.
Is the performance difference explained by the mix of apps?

- Date rates depend more on apps
- RTTs still differ per environment, regardless of the apps
Which factors impact end-host performance?

- Interface
- User tag for location
- Source ISP and country
- Destination ISP, city and country
- Country and ISP locality
- HTTP content type
- Dest. Port number
- Network load

RuleFit

RTTs

Data rates
### Average feature importance

<table>
<thead>
<tr>
<th>Data rates</th>
<th>RTTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network load (0.98)</td>
<td>Destination ISP (0.76)</td>
</tr>
<tr>
<td>Destination city (0.52)</td>
<td>Destination city (0.56)</td>
</tr>
<tr>
<td>Destination ISP (0.49)</td>
<td>Network load (0.54)</td>
</tr>
<tr>
<td>Content-type (0.35)</td>
<td>Source ISP (0.33)</td>
</tr>
<tr>
<td>Port number (0.27)</td>
<td>Content-type (0.12)</td>
</tr>
<tr>
<td>Source ISP (0.06)</td>
<td>Port number (0.06)</td>
</tr>
<tr>
<td>ISP locality (0.02)</td>
<td>Destination country (0.03)</td>
</tr>
</tbody>
</table>
Summary of characterization

- How diverse are networking environments of a user?
  - 75% users connect to multiple environments
  - Some users have over 10 environments

- Does the set of applications vary across environments?
  - True for everybody
  - Combination of applications affect data rates strongly

- Does network performance vary across environments?
  - RTTs vary dramatically

- Which factors impact end-host performance?
  - Source ISP for RTTs
  - Set of applications and content-type for data rates
Outline

- Measure network performance with user feedback
- Characterize end-host performance across networking environments
- Correlate network performance with user feedback
User vs. Network Reporting

- **User Perspective**
  - Good and poor performance epochs as flagged by the user

- **Network and System Perspective**
  - Good and poor performance epochs as flagged by lower-level metrics that trigger atypical (anomalous) behavior automatically

- **Question:** Do these co-occur?
Anomalous performance

- Metrics
  - RTTs
  - TCP retransmissions
  - Wireless noise level
  - Machine CPU load
  - Data rates
  - Wireless signal strength
  - Any instance of TCP reset

Above 95th percentile
Below 5th percentile
Can’t connect to some sites or services

Time in minutes

RTT  +
TCP RST ▲
TCP RETR ▼

1 = 95th%
0 = 5th%
Everything is good!
Matching performance reports

<table>
<thead>
<tr>
<th>Machine anomaly</th>
<th>User report</th>
<th>Yes</th>
<th>No</th>
<th>No feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>95  reports</td>
<td>82 reports</td>
<td>User doesn’t care</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>352 reports</td>
<td>313 reports</td>
<td>Missing user feedback</td>
</tr>
<tr>
<td>No</td>
<td>Not reporting correct system metrics</td>
<td>13 reports</td>
<td>39 reports</td>
<td>Missing user feedback</td>
</tr>
</tbody>
</table>
Summary of correlating user feedback with network performance

- Hard to get feedback from users
  - Many network performance samples without feedback
  - Users are diverse in how they report a problem

- Raw network metrics alone are not enough
  - Not all outliers affect the user perception
Next steps

- Incorporate user context to define anomaly
  - Application
  - Environment
  - What annoys the user

- Apply machine learning techniques
  - Use HostView data to train models of user perception

- Build online detector of poor user experience
We need your help!

- New release of HostView
  - Upload bandwidth capped at 100 kbps
  - IPv6 support

- Please download HostView at:
  http://cmon.lip6.fr/HostView

- Platforms: Mac OS 10.5, 10.6, 10.7 and Linux

Thank you!