OneClick --
A Framework for Measuring Network Quality of Experience

Kuan-Ta Chen, Cheng-Chu Tu, Wei-Cheng Xiao
Institute of Information Science, Academia Sinica

(Presenter on INFOCOM’09: Polly Huang from National Taiwan University)
QoS and QoE

- QoS (Quality of service)
  - The quality level of system performance metric
    - Communication networks: delay, loss rate
    - DBMS: query completion time

- QoE (Quality of experience)
  - The quality of how users “feel” about a service
    - Subjective: Mean Opinion Score (MOS)
    - Objective: PSNR (picture), PESQ (voice), VQM (video)
Relationship between QoS and QoE

QoE

QoS, e.g., network bandwidth

Comfort range

Too bad to perceive

Marginal benefit is small

IEEE INFOCOM 2009
Institute of Information Science, Academia Sinica
Knowing the Relationship is Important!

- So we know
  - How to adapt voice/video/game data rate (QoS) for user satisfaction (QoE)

- So we really know
  - How to send multimedia data over the Internet
Measuring QoS and QoE

- QoS (A great body of work)
  - Measure network loss, delay, available bandwidth
  - Inference topology
  - Estimate network capacity
  - etc

- QoE (Some work)
  - Objective: PSNR (picture), PESQ (voice), VQM (video)
  - Subjective: MOS (general)

Still not quite the human experience which is multi-dimensional

What’s left!
MOS (Mean Opinion Score)

<table>
<thead>
<tr>
<th>MOS</th>
<th>Quality</th>
<th>Impairment</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Excellent</td>
<td>Imperceptible</td>
</tr>
<tr>
<td>4</td>
<td>Good</td>
<td>Perceptible but not annoying</td>
</tr>
<tr>
<td>3</td>
<td>Fair</td>
<td>Slightly annoying</td>
</tr>
<tr>
<td>2</td>
<td>Poor</td>
<td>Annoying</td>
</tr>
<tr>
<td>1</td>
<td>Bad</td>
<td>Very annoying</td>
</tr>
</tbody>
</table>

Problems

1. Slow in scoring (think/interpretation time)
2. People are limited by finite memory
3. Cannot capture users’ perceptions over time
4. MOS is coarse in scale granularity
5. Dissimilar interpretations of the scale among users
Our Ambition

Identify a **simple** and yet **efficient** way to measure users’ satisfaction
The Idea: Click, Click, Click

- Web surfing
  - Click on a link
  - You wait, and you refresh the link
  - You wait, and you refresh the link again, and again, and ...

- Knocking at someone’s door
  - Knock on the door
  - You wait, and you knock on the door again
  - You wait, and you knock on the door again and again, and ...
Introducing OneClick

- Simple instruction to users:
  - Click when you feel dissatisfied
  - Click multiple times when you feel even less satisfied
- Clicking rate as the QoE
Nice Things about OneClick

- Natural
  - We are already doing it to show *lost of patience* all the time
- Bad-memory proof
  - Real-time decisions
  - No need to “remember” past experience
- Time-aware
  - Capture users’ responses at the time of the problems
  - Useful to study recency, memory access, and habituation effect
Easy to Implement

- As a plug-in to your network applications
  - Flash version done!
- Co-measurement of QoS and QoE
Talk Progress

- Overview
- Methodology
  - Pilot Study
- Validation
- Case Studies
- Conclusion

OneClick: A Framework for Measuring Network Quality of Experience
Human as a QoE Rating System

Application QoS

User

Application QoE

Network Setting

Click Events

affect

vary this:

observe this:

reflect
QoE ↔ QoS Modeling

- Click events as a counting process
- Poisson regression

\[ \log(C(t)) = \alpha_0 + \alpha_1 N_1(t) + \ldots + \alpha_k N_k(t) \]

- **C(t): QoE**
  - Clicking rate at time t
- **N_1(t), N_2(t), ... : QoS**
  - Network conditions at time t
- **\alpha_i : Regression coefficients**
  - Derived from the maximum likelihood method
Wait a Minute...

- **Response delays?**
  - Users may not be able to click immediately after they are aware of the degraded quality

- **Clicking rate of a user consistent?**
  - Does a subject give similar ratings in repeated experiments?

- **Clicking rate consistent across users?**
  - Different subjects may have different preference on click decisions.
Pilot Study

- An 5-minute English song
- Audio quality of AIM Messenger with various network settings
Test Material Compilation

- For each network setting
  - Play the song
  - Record the song
- K settings → K recordings

A test material = Random non-overlapping segments from K different recordings

A Random test material = non-overlapping segments from K different recordings
Response Delays

- Try Poisson regression on $C(t+x)$ to $N_1(t)$, $N_2(t)$, ...
- Varying $x$
- Show the goodness of fit per $x$
1-2 Seconds Delay

Response delay calibration needed!
Our Solution

- Shift the click event process by time $d$
- $d$ is decided by model fitting
  - Let $d$ be the $x$ such that the goodness of fit is the best
  - Let $d$ be the $x$ such that the residual deviance is the min
Consistency of $C(t+d)$ from Same User
Consistency of $C(t+d)$ from Different Users

Cross-user normalization needed!
Calibration and Normalization Added

OneClick Measurement

Response Delay Calibration

Regression Modeling With Normalization

Poisson Regression Model for Click Rate and Network Factors
Talk Progress

- Overview
- Methodology
- Pilot Study
- Validation
- Case Studies
- Conclusion
Rationale

- Direct: get people to do OneClick and MOS

- Indirect: get people to do OneClick and PESQ/VQM

Exact problem we are trying to solve 😞
PESQ-based Validation

- **PESQ**: Perceptual Evaluation of Speech Quality

- OneClick vs. PESQ to evaluate the audio quality of three VoIP applications
  - AIM
  - MSN Messenger
  - Skype

- Network factors
  - Loss rates (0% – 30%)
  - Bandwidth (10 Kbps – 100 Kbps)
[Validation]

**Qualitative Comparison**

**Network Loss Rate**

**Bandwidth**
VQM-based Validation

- VQM: Video Quality Measurement

- OneClick vs. VQM to evaluate video quality of two video codecs
  - H.264
  - WMV9 (Windows Media Video)

- Factors
  - Compression bit rate (200 Kbps – 1000 Kbps)
[Validation]

Qualitative Comparison

![Graph showing VQM and Click rate vs Bit rate for H.264 and WMV9]
Talk Progress

- Overview
- Methodology
- Pilot Study
- Validation
- Case Studies
- Conclusion
Case Studies

- Evaluation of applications’ QoE
  - VoIP applications
    - AIM
    - MSN Messenger
    - Skype
  - First-person shooter games
    - Halo
    - Unreal Tournament
Varying Bandwidth

- MSN Messenger is generally the worst
- Skype is the best if bw < 80 Kbps, otherwise AIM is the best
[Case Study]

Contour Lines of Click Rates

- **Slope of contour line**
  - Application’s sensitivity to loss vs. bandwidth shortage
  - AIM is relatively more sensitive to network losses
[Case Study]

**Comfort Region**

- Comfort Region: a set of network configurations that leads to satisfactory QoE
- Skype is the best in bw-restricted scenarios (< 60 Kbps) when loss rate is < 10%
Talk Progress

- Overview
- Methodology
- Pilot Study
- Validation
- Case Studies
- Conclusion

OneClick: A Framework for Measuring Network Quality of Experience
Nice about OneClick

- Natural & fast
  - We are already doing it to show lost of patience all the time
- Bad-memory proof
  - No need to “remember” past experience
- Time-aware
  - Capture users’ responses at the time of the problems
- Fine-grain
  - The score can be 0.2, 3.5, or even 12.345
- Normalized user interpretation
  - Different interpretations are normalized
- Easy to implement
  - http://mmnet.iis.sinica.edu.tw/proj/oneclick/
OneClick Online

Welcome to OneClick
OneClick is a part of research for VoIP voice quality. Please go through the following tutorial for playing OneClick in a most convenient way.

Listen and Click
After clicked the "Click to Start" button, the sound will be played, listen to it carefully. When you feel the voice quality is not good enough (delay), please click on the big "quality is bad" button, or press space key from your keyboard.

Graph:
- Your click rate
- Inverse PESQ
On-Going Work

- Large-scale experiments (by crowdsourcing)
  - http://mmnet.iis.sinica.edu.tw/proj/oneclick/
- Click rate vs. MOS
- QoE-centric multimedia networking
  - As an example, Tuning the Redundancy Control Algorithm of Skype for User Satisfaction, IEEE INFOCOM 2009.
Thank You!

Kuan-Ta Chen

http://www.iis.sinica.edu.tw/~swc