Assessment of VoIP Quality over Internet Backbones

IEEE INFOCOM 2002
Presented by 唐崇實
Outline

- Introduction
- VoIP System
- VoIP Quality Assessment Methodology
- Internet Measurements
- Numerical Results
- Conclusion
Introduction

- VoIP: via managed IP network or Internet?
- Objective: to assess to what extent today’s Internet stands up to the toll-quality expectation

Characteristics of the approach

- Use delay and loss measurements collected by sending probes
- Use appropriate voice quality measures that take into account various transmission impairments
- Take into account the effect of different components of VoIP system, with emphasis on playback buffer
VoIP System

Components of the VoIP system

- **Encoding & packetization delay**
  - **Encoder:** G.711, G.723.1, G.726, G.729
  - **Voice Activity Detection (VAD)**

- **Propagation, transmission & queueing delay**
  - **Packetizer:** encapsulate speech samples into packet of RTP/UDP/IP

- **Buffering & decoding delay**
  - **Playback buffer:** two operation modes
    - Fixed scheme
    - Adaptive scheme
    - $p = d_{av} + 4v$
  - **Depacketizer**
  - **Decoder**
  - **Packet Loss Concealment (PLC)**
Assessment Methodology

- Choose voice quality measures
  - Network performance
  - Mean Opinion Score (MOS)
  - E-model (R-value)
- Combine recent studies to rate an entire voice call, consisting of multiple short speech segments
E-model

- A computational model, ITU-T G.107/G.108
- Represented by psycho-acoustic scale
  \[ R = (R_o - I_s) - I_d - I_e + A \]
  - \( R_o \): effect of noise; \( I_s \): loudness and quantization; intrinsic to voice signal itself and not depend on the transmission over the network
  - \( I_d \): capture the effect of delay, quality degradation due to one-way or “mouth-to-ear” (m2e) delay
  - \( I_e \): capture the effect of signal distortion, called the “Special Equipment Impairment factor”
  - \( A \): advantage factor
Delay impairment $I_d$

$\begin{align*}
I_d &= I_{dte}(m2e, EL_2) + I_{dle}(m2e, EL_1) + I_{dd}(m2e) \\
\end{align*}$

- $I_{dte}, I_{dle}$: impairments due to talker and listener echo
- $EL_1, EL_2$: echo losses at the point of reflection
- Tasks (not captured in $I_d$): types of conversation, “Task 1” (most stringent) “Task 6” (most relaxed)
Loss impairment $I_e$

- $I_d$ captures the distortion of the original signal due to low-rate codec and packet loss
- Impairment increase
  - 4 unit in R scale per 1% packet loss for codec with PLC
  - 25 units per 1% packet loss for codec without PLC
VoIP call quality

- Calculating the average loss rate and the average delay only give a rough estimate

Approach
- Divide the call duration into fixed time interval and assess the quality of each interval independently
- Independent MOS(t) rating of each short interval [9]
- Take into account the “recency effect”[10]
- Use variable length intervals to handle the burstiness of packet loss, defined bursts and gaps [4] [7]
- Assessment of an entire call use both final rating [4] and the worst quality experienced during a call
Instantaneously perceived $I_e$

- $I_e(loss)$ for a gap or burst, exponential curve with
  - $T_{bad} = 5\ sec$ for the high loss periods
  - $T_{good} = 15\ sec$ for the low loss periods
Measurements Experiments

- Delay and loss measurements by probes
  - Probes were sent by and collected at measurement facilities in 5 major US cities, 43 paths by 7 providers

- Probes were 50 bytes each and sent every 10ms from 2001/6/27 19:22~2001/6/29 00:50
Trace Description (1/2)

- 43 paths are classified into 5 types

<table>
<thead>
<tr>
<th>Type</th>
<th>Num</th>
<th>Example path</th>
<th>Delay (in msec)</th>
<th>Loss usual loss events</th>
<th>long outages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>min 50% 98% 99%</td>
<td>avg clip (msec) #clips per hour</td>
<td>duration (sec) times per day</td>
</tr>
<tr>
<td>A</td>
<td>11</td>
<td>EWR $P_6$ ASH short</td>
<td>3.4 3.6 3.7 3.72</td>
<td>20 1-5</td>
<td>5-12 1-2</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>ASH $P_4$ EWR short</td>
<td>6.8 7.2 120 200</td>
<td>20 2-3</td>
<td>12-25 1</td>
</tr>
<tr>
<td>C</td>
<td>16</td>
<td>SJC $P_5$ EWR long</td>
<td>32.7 32.8 33.5 34.5</td>
<td>0 0</td>
<td>2 1</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>SJC $P_4$ ASH long</td>
<td>45.1 45.4 170 225</td>
<td>10 1-2</td>
<td>15-25 2-3</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>THR $P_1$ ASH long</td>
<td>77.8 78.2 100 210</td>
<td>10 2-20</td>
<td>1 1</td>
</tr>
</tbody>
</table>

- Observation of loss events
  - Only 3 out of 43 paths had consistently no loss
  - 6 out of 7 providers experienced outage periods 10-220 sec for 1-2 times per day
  - 0.5-2sec loss durations were correlated with delay spikes
  - The number of out-of-order packets was negligible
Trace Description (2/2)

- Observation of delay
  - Paths of type A and C have practically no queueing
  - Paths of type B and D have in general low queueing, except for clustered delay spikes
  - Paths of type E: queueing component is high and delay varies

![Example trace, type E](image)

(a) Example trace, type E

![Example trace, type D](image)

(b) Example trace, type D
Numerical Results (1/3)

- An example call trace of type \( E \)
  - Large delay variation and a period of sustained loss
  - There exists a tradeoff between loss and delay
Numerical Results (2/3)

- Many calls uniformly spread over an hour
  - 150 short (3.5 min mean): business calls
  - 50 long (10 min mean): residential calls

choice of fixed value is critical:
150 ms is acceptable (6% below 3.6)
Numerical Results (3/3)

- All other paths
  - Both short (type A) and long (type C) distance achieve an excellent MOS at all times except for rare cases when network drop occurs.
  - Path of type B and D exhibit periodically clusters of high spikes: packets are dropped at the playback buffer, whether fixed or adaptive.
Discussion

- On the performance of the backbone networks
  - Highly loaded paths (type E) as well as some over-provisioned paths exhibiting frequent delay spikes (type B and D) have poor VoIP performance

- On the playout buffer
  - An appropriate choice of the fixed playout buffer is the one that leads to maximum MOS
  - The need of adaptive playout comes when
    - The delay is high and no margin for overestimating
    - Delay is unknown and receiver doesn’t know how to select an appropriate fixed value
  - The adaptive playback was useful on loaded network (type E) that exhibited high and slowly varying delays but fails on others.
Conclusion

- This paper assess the ability of Internet backbones to support VoIP considering realistic configuration.
- In general, backbone networks are over-provisioned, but poor VoIP performance on a large number of ISP backbone networks is observed.
- Action for improving today’s VoIP performance to reach toll-quality standards can be taken
  - Inside the network: marking the voice traffic to give it preferential treatment
  - At the receiver, playout buffer scheme should be carefully chosen to match the delay pattern