Voice over the Internet (the basics)
Outline

• Basics about voice encoding
• Packetization trade-offs
• Architecture of basic VoIP tool
• Playback buffer (jitter buffer)
  • Adaptive playback buffers?
• How to deal with packet losses and late packets?
Voice over the Internet

- Includes computer2computer voice applications (like Skype, VoIPBuster, etc)
- + VoIP services
- + Telephony Routing over IP (TRIP)
- Includes “off-net” calls (calls to PSTN phones)
Reading-1

• “Voice over Internet Protocol (VoIP)” by Bur Goode, published at IEEE Proceedings, Sep’02
It all starts from an analog signal
## Codecs

<table>
<thead>
<tr>
<th>Codec</th>
<th>Algorithm</th>
<th>Frame Size/ Lookahead</th>
<th>Usual Rate</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.711</td>
<td>PCM</td>
<td>0.125 ms/0</td>
<td>64 Kb/s</td>
<td>Universal use</td>
</tr>
<tr>
<td>G.722</td>
<td></td>
<td>0.125 ms/1.5 ms</td>
<td>48, 56 or 64 Kb/s</td>
<td>Wideband coder</td>
</tr>
<tr>
<td>G.726</td>
<td>ADPCM</td>
<td>0.125 ms/0</td>
<td>32 Kb/s</td>
<td>High quality, low complexity</td>
</tr>
<tr>
<td>G.728</td>
<td>LD-CELP</td>
<td>0.625 ms/0</td>
<td>16 Kb/s</td>
<td>High quality in tandem; Recommended for cable</td>
</tr>
<tr>
<td>G.729(A)</td>
<td>CS-ACELP</td>
<td>10 ms/5 ms</td>
<td>8 Kb/s</td>
<td>Widespread use</td>
</tr>
<tr>
<td>G.729e</td>
<td>Hybrid CELP</td>
<td>10 ms/5 ms</td>
<td>11.8 Kb/s</td>
<td>High quality/complexity; Recommended for cable</td>
</tr>
<tr>
<td>G.723.1(6.3)</td>
<td>MPC-MLQ</td>
<td>30 ms/7.5 ms</td>
<td>6.3 Kb/s</td>
<td>Video conferencing origin</td>
</tr>
<tr>
<td>G.723.1(5.3)</td>
<td>ACELP</td>
<td>30 ms/7.5 ms</td>
<td>5.3 Kb/s</td>
<td>Video conferencing origin</td>
</tr>
<tr>
<td>IS-127</td>
<td>RCELP</td>
<td>20 ms/5 ms</td>
<td>Var. 4.2 Kb/s avg.</td>
<td></td>
</tr>
<tr>
<td>AMR</td>
<td>ACELP</td>
<td>20 ms</td>
<td>Var. 4.75-12.2 Kb</td>
<td>Compatible w. No. Amer. &amp; Japanese digital cellular, WCDMA (not CDMA2000); Nokia IPR</td>
</tr>
</tbody>
</table>
How does PCM work?

• Voice spectrum extends to about 3-4KHz

• According to Nyquist’s rate, a sampling frequency of 8KHz should be enough to completely reconstruct the original voice signal from the sampled signal

• PCM uses 8 bits per sample (64kbps)

• Frame size?
  • G.711 uses 125msec (too large for packet voice)
  • G.729 uses 10msec
Listen to the various codecs and judge for yourself


(look at bottom of this page)
Popular recent codecs for VoIP

- See GlobalIPSound

(http://www.gipscorp.com/products/demos.php)

- **Wide band codecs (50-8,000 Hz)**
  - iLBC (packetization: 20 and 30 msec, bitrate: 15.2 kbps and 13.3 kbps)
    - Free, open-source
    - No error propagation when lost frame (problem with LPC)
  - iSAC (proprietary – best codec currently?)
    - PACKET SIZE Adaptive, 30 - 60 ms
    - BIT RATE Adaptive and variable, range 10 - 32 kbps
    - SAMPLING RATE 16 kHz
    - AUDIO BANDWIDTH 8 kHz
MOS scores

- Also look at the effect of “codec concatenation” (e.g., G.729*3)
Effects of transcoding
Packetization tradeoffs

• R: encoding rate (bps)
• H: header size per packet (bits)
  • E.g., 40B for RTP/UDP/IP packet
• S: packetization period or sample duration (sec)
• BW: voice transmission requirement
  • BW = R + H/S
  • How can you decrease BW?
  • Lower R means more complex codec, more correlations across successive packets
  • Higher S means more delay at sender and larger sensitivity to packet losses
Fig. 5. The varying bands, from top to bottom, represent the following VoIP bandwidth requirements (40-byte headers): 120–140, 100–120, 80–100, 60–80, 40–60, 20–40, and 0–20.
Fig. 6. From top to bottom, varying bands represent the following VoIP bandwidth requirements (4-byte headers): 70–80, 60–70, 50–60, 40–50, 30–40, 20–30, 10–20, 0–10.
Control Mechanisms for Packet Audio in the Internet

Jean-Chrysostome Bolot        Andrés Vega-García
INRIA
B. P. 93
06902 Sophia-Antipolis Cedex
France
{bolot, avega}@sophia.inria.fr
Figure 1: Structure of the audio tool
Network effects

• One-way delay between sender/receiver
  • Includes encoding, packetization, transmission, propagation, queueing, jitter compensation, decoding
  • Typically, acceptable if < 150msec for domestic calls and < 400msec for international
    • Depends on call’s interactivity
  • What can we do to reduce packet delay?
Network effects (cont’)

• Packet losses
  • Low-bitrate codecs are very sensitive to packet losses (why?)
  • Should we do retransmissions?
  • Should we do Forward-Error-Correction?
  • Or just, packet loss concealment? How?

• Delay variation or jitter
  • Jitter compensation buffer at receiver
  • How large should this buffer be?
  • Losing vs discarding packets
  • Delay budget calculations

• Insufficient network capacity
  • Rate adaptation (use multiple codecs)
Adaptive Playout Mechanisms for Packetized Audio Applications in Wide-Area Networks

Ramachandran Ramjee, Jim Kurose, Don Towsley
Department of Computer Science
Univ. of Massachusetts
{ramjee,kurose,towsley}@cs.umass.edu

Henning Schulzrinne
GMD-Fokus
Berlin, Germany
hgs@fokus.gmd.de
Figure 1: Generation and reconstruction of packetized voice
## Delay budget

<table>
<thead>
<tr>
<th>Delay Source (G.729)</th>
<th>On-net Budget (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device Sample Capture</td>
<td>0.1</td>
</tr>
<tr>
<td>Encoding Delay (Algorithmic Delay + Processing Delay)</td>
<td>17.5</td>
</tr>
<tr>
<td>Packetization/ Depacketization Delay</td>
<td>20</td>
</tr>
<tr>
<td>Move to Output Queue/Queue Delay</td>
<td>0.5</td>
</tr>
<tr>
<td>Access (up) Link Transmission Delay</td>
<td>10</td>
</tr>
<tr>
<td>Backbone Network Transmission Delay</td>
<td>Dnw</td>
</tr>
<tr>
<td>Access (down) Link Transmission Delay</td>
<td>10</td>
</tr>
<tr>
<td>Input Queue to Application</td>
<td>0.5</td>
</tr>
<tr>
<td>Jitter Buffer</td>
<td>60</td>
</tr>
<tr>
<td>Decoder Processing Delay</td>
<td>2</td>
</tr>
<tr>
<td>Device Playout Delay</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>121.1 + Dnw</strong></td>
</tr>
</tbody>
</table>
Figure 3: Example illustrating playout mechanisms
A Survey of Packet Loss Recovery Techniques for Streaming Audio

Colin Perkins, Orion Hodson, and Vicky Hardman
University College London
Figure 3. A taxonomy of sender-based repair techniques.
Repair using parity FEC.
Figure 5. Repair using media-specific FEC.
Figure 6. *Interleaving units across multiple packets.*
Figure 7. A taxonomy of error concealment techniques.
Figure 8. Rough quality/complexity trade-off for error concealment.