Signal Processing for Packet Voice Telephony

IEEE OEB ComSoc, San Ramon
May 20, 2004

Krishna V V
Principal Solutions Architect
DSP Products Development

LSI Logic
Agenda

• About LSI Logic, briefly ...
• Packet Voice Networks Overview
• Signal Processing Layer Internals
  – Voice-specific functions
    • Speech codecs; packet loss concealment; echo cancellation; silence suppression; …
  – Other related functions
• Processing needs for different contexts
• A few trends & issues
Voice Networks: Past
Voice Networks: Present
Telling Quotes

“If you don’t do it, next year or the one after that you won’t be playing in the game”

Richard Notebaert (CEO, Qwest)

(while announcing Qwest’s plan to offer low-cost IP based phone service, 2003).
Telling Quotes

“As of three weeks ago, all the long-distance (voice) traffic in Italy is carried over the IP network”

Stefano Pileri (Head, Domestic Network, Telecom Italia)

Telling Quotes

“Telecom may be heading the way of DRAMs, where the price is set by the most idiotic competitor. ... It is a race to the bottom, and the bottom in this case is free service”

Robert Lucky

Not so SIPle News

Packet Pathways

- **Wired**
  - Cable Modems
  - DSL Networks
  - Corporate Ethernet LANS
  - Managed / Public Wide-area Networks (WANs)

- **Wireless**
  - WLANS (WiFi)
  - Satellite IP Networks (including VSATs)
  - BWA (WiMax) (?)
Protocol Menu

- Protocols for VoIP Inter-operability:
  - MGCP (CableLabs / PacketCable)
  - H.323 (ITU, videoconferencing)
  - H.248 /Megaco (ITU)
  - SIP (IETF)
- Others key protocols (Quality, …):
  - RTP, RTCP
  - UDP
  - RSVP, DiffServ, …
Packet Voice “Boxes”

Increasing voice channels (1X to ~ 50,000X)
## VOIP Market: DSP Slice

<table>
<thead>
<tr>
<th></th>
<th>2001</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Growth (voice channels):</td>
<td>12 Million</td>
<td>560 Million</td>
</tr>
<tr>
<td>DSP Revenue:</td>
<td>US$129M</td>
<td>US$1400M</td>
</tr>
<tr>
<td>ASP (DSP + S/W)</td>
<td>$25</td>
<td>$225</td>
</tr>
<tr>
<td>Channels / DSP</td>
<td>&lt; 5</td>
<td>~ 88</td>
</tr>
<tr>
<td>Average cost / channel</td>
<td>$5.70</td>
<td>$2.60</td>
</tr>
</tbody>
</table>

## Packet Voice: Key Hurdles

<table>
<thead>
<tr>
<th></th>
<th>Typical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Delay</strong></td>
<td>Typical end-to-end delays around 100-200ms</td>
</tr>
<tr>
<td><strong>Packet Jitter</strong></td>
<td>Typical arrival time jitter around 20-50ms</td>
</tr>
<tr>
<td><strong>Packet Loss</strong></td>
<td>Typical losses around 1-2%</td>
</tr>
</tbody>
</table>
## Delay: G.114 Guidelines

<table>
<thead>
<tr>
<th>One-way Delay</th>
<th>ITU-T Classification (with echo “adequately controlled”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 150ms</td>
<td>Mostly acceptable.</td>
</tr>
<tr>
<td>150-400ms</td>
<td>Acceptable (maybe).</td>
</tr>
<tr>
<td>&gt; 400ms</td>
<td>Unacceptable (in general).</td>
</tr>
</tbody>
</table>

### TYPICAL DELAYS
- Terrestrial, national long distance PSTN: < 50ms
- Terrestrial, international PSTN: ~ 100ms
- Cellular: Mobile to PSTN: ~ 150ms
- Cellular: Mobile to Mobile: ~ 300 – 400ms
# Delay in Packet Networks

Overall delay break-up:

<table>
<thead>
<tr>
<th>Component</th>
<th>Range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech Codec</td>
<td>0.2 – 68ms</td>
<td>Low delay for PCM, ADPCM, G.728, …</td>
</tr>
<tr>
<td>Packetization</td>
<td>5 – 30ms</td>
<td></td>
</tr>
<tr>
<td>Interleaving</td>
<td>0 – 60ms</td>
<td>Optional</td>
</tr>
<tr>
<td>Transmission</td>
<td>25 – 150ms</td>
<td></td>
</tr>
<tr>
<td>Jitter Buffer</td>
<td>50 – 100ms</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>~80 – 400ms</td>
<td><strong>Typical: 100 – 200ms</strong></td>
</tr>
</tbody>
</table>

Handling Delay (Echo)

- One-way delays in packet voice networks > 100ms
- As recommended in ITU-T G.131, a network echo canceller (EC) is required.
- EC required only for:
  - PSTN interfaces on voice gateways
  - Analog phone (SLIC) interfaces on CPEs
- EC not required for digital IP phones (AEC is a different option)
- EC tail length – a much misused parameter
- ITU-T G.168 EC was initially developed for PSTN. Can it be applied as-is for packet voice networks?
Tackling Echo: ITU Standards

CCITT → ITU
EC for Packet Voice

Question: Why EC in the Gateway?
Echo Level and Delay

EC – A Black-box View

\[
ERL = L_{\text{RIN}} - L_{\text{ECHO}}
\]

\[
A_{\text{COM}} = L_{\text{RIN}} - L_{\text{SOUT}} \text{ (near-end signal absent)}
\]
G.168 EC Internals

- Control Logic (Adaptation, NLP)
- V.25 Tone Detector; Holding-band Logic
- Nonlinear Processor (NLP); Comfort Noise (CNI)
### Some EC Design Options

<table>
<thead>
<tr>
<th>“Full tail”</th>
<th>“Tail independent” or “Floating window”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single filter with robust control</td>
<td>Double filter with simpler controls</td>
</tr>
<tr>
<td>Time domain</td>
<td>Transform domain</td>
</tr>
</tbody>
</table>
Full Tail / Floating Window

Actual echo path

Full tail solution

2-window solution

12 ms 12 ms
Handling Packets

Packets assembled and ready to go …

Average end-to-end delay

Packets arriving at destination with jitter.

Packets arriving at destination with jitter and out of order (likely on public networks).
Jitter Buffer

• Evidently, jitter buffer is a crucial module in the receiver.
• Out-of-order packet arrivals can be sorted based on RTP time stamp.
• Trade-off of voice quality versus latency.
  – A small buffer helps minimize the extra latency, but drops packets that arrive too late.
• Adaptive jitter buffer that grows or shrinks as needed, is one solution.
## Packet Losses

- In addition to packet drops by jitter buffer, packet losses are likely due to
  - UDP (does not offer guarantee of delivery)
  - Network congestion (bandwidth)
  - Router overload (packet throughput)

- Up to 5% (or more?) packet losses considered likely
  - Even 1% packet loss degrades voice quality significantly
  - Packet Loss Concealment (PLC) is yet another essential module in the receiver
  - PLC is not sufficient to handle certain tone signals (DTMF digits, V.25 tone for EC disabling, etc.)
Tone Relay

- Helps in reliable transfer of DTMF digits and other signaling tones (packet losses)
- Fast DTMF detection also avoids possible leakage problems
  - Fast detection particularly important with low bit rate voice codecs such as G.723.1 or G.729.
- Q: Does tone relay use UDP or TCP?
Dealing With Packet Loss

- **Network Level (transparent to DSP)**
  - QoS protocols
  - Call Admission Control
- **Other Non-transparent Means**
  - Adaptive Jitter Buffer
  - Interleaving
  - Transmit Redundant Packets
  - Silence Suppression

Quality gained at the cost of extra latency

Indirect approach – reduce network congestion
Packet Voice: Key Blocks

- EC required only if echo can occur in ROUT / SIN loop.
- EC (if deployed) should be properly aligned with any I/O buffering delays in the loop.
• If an EC is present, how should its NLP + CNI module be used? Preferable to integrate this function with the VAD.
Packet Voice: RCV Path Detail

Voice RIN (to EC) / ROUT to listener

From Jitter Buffer

Tone Generator

CNG

Voice Decoder

PLC
Speech Coding Criteria

Depending on the end-use, each criterion requires a different weightage.
Speech Coding Approaches

Hybrid Codecs employ tools such as vector quantization to gain quality at low bit rates.

Hybrid codecs are more costly to implement.
Voice Quality

G.107 → 94
90
80
70
60
50
0

MOS
Mean Opinion Score

R
E-Model Rating Value

Mean Opinion Score (MOS)
5.0
4.4
4.3
4.0
3.6
3.1
3.0
2.6
2.0
1.0
0
### Some Voice Codec Options

<table>
<thead>
<tr>
<th>Codec</th>
<th>Bit Rate (kbits/s)</th>
<th>Delay</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.711 (PCM)</td>
<td>64</td>
<td>0.125ms</td>
<td>Toll quality</td>
</tr>
<tr>
<td>G.726 (ADPCM)</td>
<td>32</td>
<td>0.25ms</td>
<td>Toll Quality</td>
</tr>
<tr>
<td>G.728</td>
<td>16</td>
<td>0.625ms</td>
<td>Licensing required</td>
</tr>
<tr>
<td>G.723.1</td>
<td>5.3 / 6.3</td>
<td>30ms (7.5ms)</td>
<td>Licensing required</td>
</tr>
<tr>
<td>G.729 / G.729A</td>
<td>8</td>
<td>10ms (5ms)</td>
<td>Licensing required</td>
</tr>
<tr>
<td>GSM (RPE-LTP)</td>
<td>13</td>
<td>20ms</td>
<td>Patents? Low Quality</td>
</tr>
<tr>
<td>iLBC</td>
<td>13.4 / 15.2</td>
<td>30ms / 20ms</td>
<td>Royalty-free / IETF</td>
</tr>
<tr>
<td>Speex</td>
<td>2.2 - 44</td>
<td>30ms</td>
<td>Open source</td>
</tr>
</tbody>
</table>

NOTE: Delay column shows frame size $T_F$ and look-ahead buffer duration $T_{LA}$, if any. Total codec processing delay is $2T_F + T_{LA}$.
## Voice Codec Costs

<table>
<thead>
<tr>
<th>Codec</th>
<th>Bit Rate (kbits/s)</th>
<th>DSP Processing Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.711 (PCM)</td>
<td>64</td>
<td>&lt; 1 MHz (with PLC)</td>
</tr>
<tr>
<td>G.726 (ADPCM)</td>
<td>32</td>
<td>4 – 9 MHz (with PLC)</td>
</tr>
<tr>
<td>G.728</td>
<td>16</td>
<td>20 – 30 MHz</td>
</tr>
<tr>
<td>G.723.1</td>
<td>5.3 / 6.3</td>
<td>12 – 20 MHz</td>
</tr>
<tr>
<td>G.729 / G.729A</td>
<td>8</td>
<td>12 – 20 / 6 – 10 MHz</td>
</tr>
<tr>
<td>GSM (RPE-LTP)</td>
<td>13</td>
<td>3 – 8 MHz</td>
</tr>
<tr>
<td>iLBC</td>
<td>13.4 / 15.2</td>
<td>8 – 12 MHz</td>
</tr>
<tr>
<td>Speex</td>
<td>2.2 - 44</td>
<td>?</td>
</tr>
</tbody>
</table>

**NOTE:** Program / data memory requirements of each codec should also be considered.
R-degradation: Packet Loss

Source: Luis F Ortiz (Brooktrout Technology), RTC Magazine, July 2001
R-degradation: Packet Loss

Source: Luis F Ortiz (Brooktrout Technology), RTC Magazine, July 2001
Packet Header Overheads

- Packetization overheads can be significant.
- Header compression (cRTP) is possible for IP/UDP/RTP. Use of cRTP needs low round-trip delays (for header repair requests). This is primarily useful for (low-speed) local links.
- Large packets amortize overheads at the cost of extra latency.
- Other overheads (such as RTCP) are not counted here.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>E’NET</td>
<td>~18</td>
</tr>
<tr>
<td>IP</td>
<td>20</td>
</tr>
<tr>
<td>UDP</td>
<td>8</td>
</tr>
<tr>
<td>RTP</td>
<td>12</td>
</tr>
</tbody>
</table>

Speech Payload
(Depends on codec and frame size)
Bit Rates With Overheads

Voice Payload (Bytes) :

<table>
<thead>
<tr>
<th>Codec</th>
<th>Frame Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10ms</td>
</tr>
<tr>
<td>G.711</td>
<td>80</td>
</tr>
<tr>
<td>G.726</td>
<td>40</td>
</tr>
<tr>
<td>G.729A</td>
<td>10</td>
</tr>
</tbody>
</table>

Channel Bit Rate (kbits/s) :

<table>
<thead>
<tr>
<th>Codec</th>
<th>Frame Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10ms</td>
</tr>
<tr>
<td>G.711</td>
<td>110.4</td>
</tr>
<tr>
<td>G.726</td>
<td>78.4</td>
</tr>
<tr>
<td>G.729A</td>
<td>54.4</td>
</tr>
</tbody>
</table>

Unless packet header overheads are reduced, benefits of low bit-rate codecs are not fully utilized.
## Signal Processing Options

<table>
<thead>
<tr>
<th>Low-channel CPEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>One General purpose μP alone</td>
</tr>
<tr>
<td>One DSP alone</td>
</tr>
<tr>
<td>Single-chip IP Phone SoC (μP + DSP + I/O)</td>
</tr>
<tr>
<td>One uP (host) + one or more DSP(s)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Large CPEs, Gateways</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple hosts + DSP Farms</td>
</tr>
<tr>
<td>Hosts + DSPs with HW Accelerators</td>
</tr>
<tr>
<td>Few hosts + SoC (Processors + HW + I/O)</td>
</tr>
</tbody>
</table>
Signal Processing Costs

- DSP MHz numbers below typically scale up by 1.5X to 3.0X on general purpose processors.
- Only functions that contribute to peak processor load are listed.
- Memory usage (not presented) is often a critical factor.

<table>
<thead>
<tr>
<th>Functions</th>
<th>DSP MHz / voice channel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codec (VAD-CNG, PLC)</td>
<td>3 – 30 (depends on codec choice)</td>
</tr>
<tr>
<td>G.168 EC</td>
<td>3 – 10 (depends on EC design)</td>
</tr>
<tr>
<td>DTMF Tx + Rx</td>
<td>1 – 4</td>
</tr>
<tr>
<td>Caller ID Tx</td>
<td>1 – 2</td>
</tr>
<tr>
<td>Other Functions</td>
<td>4 – 14 (Jitter buffer, packet processing, I/O handling, task scheduling)</td>
</tr>
</tbody>
</table>

Total: 12 – 60 MHz
Trends & Issues

- Wide band (7 kHz) voice codecs
- Stereo audio (conferencing) (?!?)
- Improved multi-party conferencing support
  - Conference bridges with multi-casting?
- Improved QoS
- Improved security
- Lower power consumption

In particular, the emerging VoWLAN (or VoWiFi) market needs this support.
Packet Voice References

• Books
  – D. J. Wright, "Voice over Packet Networks", Wiley (2001)
• IETF
• Cable Labs (PacketCable)
• ATM Forum
• Frame Relay + MPLS Forum
• Historical papers (packet voice problem not new):
  – W. A. Montgomery: "Techniques for packet voice synchronization" IEEE JSACS, SAC-1, 6, Dec 1983