Resource Assignment for Integrated Services in Wireless Multimedia Networks

Victor Bahl
bahl@microsoft.com
http://research.microsoft.com/~bahl

Microsoft Research
Redmond, Washington

August 8, 1997
Presentation Outline

1. Introduction and Problem Description
2. Bandwidth Allocation, Reservation, and Utilization
   → Connection level *Intra-Frame Statistical Multiplexing*
3. Bandwidth Partitioning
   → Priority Sharing with Restrictions (PSR)
   → The Smart Allocate Algorithm
4. Conclusions

Current Research Activities
Introduction and Problem Description

- Different broadband sources (traffic classes) require different amounts of bandwidth and have different transmission priorities
  - Bandwidth: Video > Voice > Data
  - Priorities: Voice > Video > Data

- Challenge for the network designer:
  Develop techniques that can assign bandwidth to the different traffic classes in manner such that
  - a guarantee on the level of service can be provided to a portion of the time-critical traffic classes, and
  - the needs of the system (i.e. maximize the # of supported connections) are balanced against the needs of the applications

(Bandwidth prediction and reservation for bursty sources such as VBR video is an unsolved problem, so how do we allocate bandwidth for such sources)
**Possible Approaches**

Adapt the application:
- Use an appropriate (joint source-channel) video codec
  - very **low bit rate** with **high average PSNR** and **acceptable frame rate** (subband based, MPEG-4, H.263L etc.)
- Use a CBR voice connection (e.g. GSM speech CODEC)

Adapt the communication layer:
- Carry out appropriate resource management (bandwidth reservation, allocation and utilization)
- Implement a “smart” bandwidth partitioning strategy
- Implement a complimentary MAC protocol
Bandwidth Reservation, Utilization, and Partitioning

Question
Can performance guarantees be provided to VBR video without significantly under-utilizing the bandwidth?

Answer
Yes! Use Intra-frame statistical multiplexing
Statistical Multiplexing: Connection Level and System Level

Connection-level Multiplexing

Region 1 / Subband 1
Region 2 / Subband 2
Region n / Subband n

Data Source

Audio Source

Video Source

MUX

System-level multiplexing

Laptop

Laptop

Laptop

August 8, 1997  Victor Bahl
Effect -- Improved Bandwidth Utilization

Bandwidth usage with and without intra-frame statistical multiplexing
Bandwidth Partitioning (Systems Perspective)

Possible Strategies:
- Complete Sharing
- Complete Partitioning (or Mutually Restricted Access)
- Partitioning with Priority Borrowing (or Partial sharing)

Strategy 1: Complete Sharing (CS)
Strategy 2: Complete Partitioning (CP)
Strategy 3: Priority Sharing (PS)
## Sharing Strategies -- Pros and Cons

<table>
<thead>
<tr>
<th></th>
<th>CP</th>
<th>CS</th>
<th>PS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROS</td>
<td>• Blocking level for each class easy to adjust</td>
<td>• Blocking level among classes not adjustable</td>
<td>• May be tunable to variable blocking requirements</td>
</tr>
<tr>
<td></td>
<td>• Complete Protection from overload of other classes</td>
<td>• No protection from overload of other classes</td>
<td>• May offer protection against overload from other classes</td>
</tr>
<tr>
<td></td>
<td>• Requires too much bandwidth</td>
<td>• Better Bandwidth usage than CP</td>
<td>• More efficient bandwidth usage than CP and CS</td>
</tr>
</tbody>
</table>
Proposal -- Priority Sharing with Restrictions (PSR)

<table>
<thead>
<tr>
<th>Requester</th>
<th>Data</th>
<th>Voice</th>
<th>Video-Dynamic</th>
<th>Video-Static</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>-</td>
<td>BP</td>
<td>BP</td>
<td>BP</td>
</tr>
<tr>
<td>Voice</td>
<td>B</td>
<td>-</td>
<td>B</td>
<td>BP</td>
</tr>
<tr>
<td>Video-Dynamic</td>
<td>B</td>
<td>BP</td>
<td>-</td>
<td>BP</td>
</tr>
<tr>
<td>Video-Static</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>-</td>
</tr>
</tbody>
</table>

B --- Borrow  BP --- Borrow, but Preemption possible
Comparison of Bandwidth Partitioning Schemes

Voice + Data Traffic

Voice Traffic Load = 30 Erlangs
Data Traffic Load = 10 Erlangs

Video Traffic

Compression = Region-based H.263
Avg. Frame Rate = 7.5 Hz
Avg. Video Bit rate = 24 kbps
Image Dimensions = QCIF (176 x 144)
# of Regions/Image = 5
Peak for primary Region = 20 kbps
Avg. PSNR = 30.4 dB
Comparison of Bandwidth Partitioning Schemes

Bandwidth Partitioning and VBR Video

- voice = 60 erlangs
- data = 10 erlangs

Average PSNR

- PB
- PSR
- CS

Offered Video Traffic Load (Erlangs)
Optimal Bandwidth Partitioning

Problem
Can performance guarantees be provided to VBR video without significantly under-utilizing the bandwidth and can this be done in conjunction with minimizing the blocking probability for voice and data traffic?

Alternatively
How do we decide, how much of the available bandwidth should be allocated to each traffic class?

Solution
Allocate bandwidth so the maximum call blocking probability is minimized
Minimize the Maximum Blocking Probability

Problem Statement

Given the aggregate load $D_i$ for each traffic type $T_i$ and the total system bandwidth $B$, determine the allocated transmission capacity for each traffic type such that

$$\mathcal{O} = \max_i P(d_i(t) \geq B_i)$$

is minimized

subject to

$$\sum_{i=1}^{N} B_i = B$$

where $d_i(t)$ is the instantaneous demand.
Minimize

\[ \Theta = \max_i \mathbb{P}(d_i(t) \geq B_i) \]

Theorem

Let \( X_1, \ldots, X_N \) be independent random variables taking their values from the interval \([0, 1]\). Their probability distributions are otherwise arbitrary and not necessarily identical. Set \( X = \sum_i X_i \) and \( D = \mathbb{E}(X) \) then for any \( C \geq D \) the following estimation holds:

\[ \mathbb{P}(X \geq C) \leq \left( \frac{D}{C} \right)^C e^{C - D} \]

Furthermore, this estimation is best possible in the following sense: For any fixed \( \varepsilon > 0 \) and for any fixed \( D \) and \( C \) with \( C \geq D \) there exist infinitely many counter examples for which the reverse holds.
Minimize

$$\Theta = \max_i P(d_i(t) \geq B_i)$$

New Problem Statement

Given the aggregate load $D_i$ for each traffic type $T_i$ and the total system transmission bandwidth $B$, determine the allocated transmission capacity for each traffic type, such that

$$\Theta = \max_i \left\{ \left( \frac{D_i}{B_i} \right)^{B_i} e^{B_i - D_i} \right\}$$

is minimized

subject to

$$\sum_{i=1}^{N} B_i = B$$
Basis for Determining $B_i$s

Can prove: Allocation of $B_i$ is **asymptotically optimal if and only if**

$$
\left( \frac{D_1}{B_1} \right)^{B_1} e^{B_1 - D_1} = \ldots = \left( \frac{D_N}{B_N} \right)^{B_N} e^{B_N - D_N} \quad \text{holds!}
$$

Thus the problem is solved by letting,

$$
\left( \frac{D_i}{B_i} \right)^{B_i} e^{B_i - D_i} = \sigma
$$

and iteratively searching for a value $0 < \sigma \leq 1$ for which $\sum_{i=1}^{N} B_i = B$ holds within a given error bound $\varepsilon > 0$
Basis for Determining $D_i$

**Method**

If $X_i$ is the size of $i$th frame (or region or subband) in a video sequence which has $n$ frames, and if we define:

$$Y_n = \max \left( X_1, X_2, \ldots, X_n \right)$$

then by letting $Y = \lim_{n \to \infty} Y_n$ and deriving $f_{Y_n}(y)$

we can calculate the mean $\eta_y$ and determine $D_i$

$$D_i = M \times \eta_y$$

where $M$ is the # of video connections to be supported.
Statistical Characterization (VBR Video)

Inter-Frame Coding
“Headline News”

Avg. PSNR = 30.44 dB

Intra-Frame Coding
“Headline News”

Avg. PSNR = 31.52 dB

Bi-Normal distribution
“Amadeus”

QQ plot for “Headline News” (JPEG)
Using Weibull Distribution

August 8, 1997

Victor Bahl
### Statistical Characterization (VBR Video)

**Mean Square Error (inter-frame compressed, CNN sequence)**

- Normal: $0.042793$
- Gamma: $0.041567$
- Lognormal: $0.037322$
- Weibull: $0.070562$

**Captured & Compressed @ 15 fps**

- $\text{mean} = 5.17$
- $\text{par} = 9.89$
- $\text{cov} = 1.17$
- $\text{avg. dev.} = 4.12$

**Captured & Compressed @ 5 fps**

- $\text{mean} = 14.62$
- $\text{par} = 5.83$
- $\text{cov} = 0.57$
- $\text{avg. dev.} = 6.69$

---

August 8, 1997

Victor Bahl
Determining $f_{Y_n}(y)$ and $D_i(\eta_y)$

Results from distribution-based modeling of single VBR video sources

<table>
<thead>
<tr>
<th>Coding Technique</th>
<th>Request Video</th>
<th>Conference Video</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High Capture</td>
<td>High Capture</td>
</tr>
<tr>
<td></td>
<td>Low Capture</td>
<td>Low Capture</td>
</tr>
<tr>
<td>Intra-Frame</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Action</td>
<td>Normal</td>
<td>Low Action</td>
</tr>
<tr>
<td>Low Action</td>
<td>Normal/Gamma</td>
<td>High Action</td>
</tr>
<tr>
<td>Normal</td>
<td>Normal/Weibull</td>
<td>Low Action</td>
</tr>
<tr>
<td>Bi-Normal</td>
<td>Gamma/Lognormal</td>
<td>Bi-Normal</td>
</tr>
<tr>
<td>Bi-Normal</td>
<td>Gamma</td>
<td>Gamma-Pareto</td>
</tr>
<tr>
<td></td>
<td>Low Action</td>
<td></td>
</tr>
<tr>
<td>Inter-Frame</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notice, most video frame distributions can be described as:

$$F_x(x) = 1 - e^{-g(x)}$$  \text{where}  \quad g(x) \quad \text{is a increasing function of} \ X$$

then we can prove,  

$$F_Y(y) = \exp[-e^{-\alpha(y-u)}], \quad -\infty < y < \infty$$
Determining $f_{Y_n}(y)$ and $D_i(\eta_y)$

- Thus knowing $F_Y(y)$ the mean can be computed as:
  \[ \eta_y = u + \left( \frac{0.577}{\alpha} \right) \]
  where $u$ and $\alpha(>0)$ are the location and scale parameters

- For Voice connections:
  \[ D_a = M \times C \]
  where $C$ is CBR (e.g. GSM - 13 KHz, DECT - 32 KHz, …) and $M$ is the estimated number of voice connections

*Note: Can use $D_i$ s for Network Capacity Planning!*
The Smart Allocate Algorithm

Begin

1. \( l = 0, u = 1 \) Initialize

2. \( \sigma = \frac{l + u}{2} \) Midpoint of \([l, u]\)

3. Compute \( B_1(\sigma) \)

4. Compute Total Capacity \( \tilde{B} = \sum_{i=1}^{N} B_i(\sigma) \)

5. Check if Capacity is overused \( \tilde{B} > B \)

6. Check if Capacity is underused \( \tilde{B} < B - \varepsilon \)

7. Ease up on bound \( l = \sigma \)

8. Stricter bound \( u = \sigma \)

End

\( B_1 = B_1(\sigma), \ldots, B_N = B_N(\sigma) \)
illustrative Example

Assume TDM-based system with 99 BBU

Let the aggregate average demands be
\[ D_1 \text{ (voice)} = 20 \text{ Erlang} \text{ and } D_2 \text{ (video)} = 40 \text{ Erlangs} \]

Then the **Load Proportional** approach gives
\[ B_1 = 33, \ B_2 = 66, \text{ with Erlang’s blocking probability = 1 %} \]

**Smart Allocate** algorithm gives
\[ B_1 = 37 \text{ and } B_2 = 62, \text{ blocking probability = 0.57%} \]

**Improvement of 43% !**
Conclusion - Advantages of PSR + Smart Allocate Algorithm

- Accommodates all traffic classes without shutting out any single one

- Provides QoS guarantees for on-going VBR real-time video communications. This guarantee does not come at the expense of bandwidth and other traffic classes.

- Robust and insensitive to statistical assumptions. Does not require detailed knowledge of traffic, only aggregate average values (detailed statistical information is typically unavailable)

- Allocation based on minimizing a bound on the blocking probabilities that is proven to be asymptotically optimal - significant as it signifies that for large number of systems it is sufficient to know aggregate flow rates.
Current Projects at MSR
Peripatetic Computing for the Next Millenium

- Ad-Hoc multi-hop home networking
  - Protocols, Routing algorithms, architecture, roaming, locating, performance etc

- Hardware for hand-held communicators
  - RF issues, form factor, capacity, capability, user requirements, user interface etc.

- QoS in mobile multimedia
  - Content-Sensitive Video Coding (beyond MPEG-4), Resource assignment and management, etc.

- Operating system support for mobile users
  - caching, hoarding, prediction, data management, disconnected operations etc