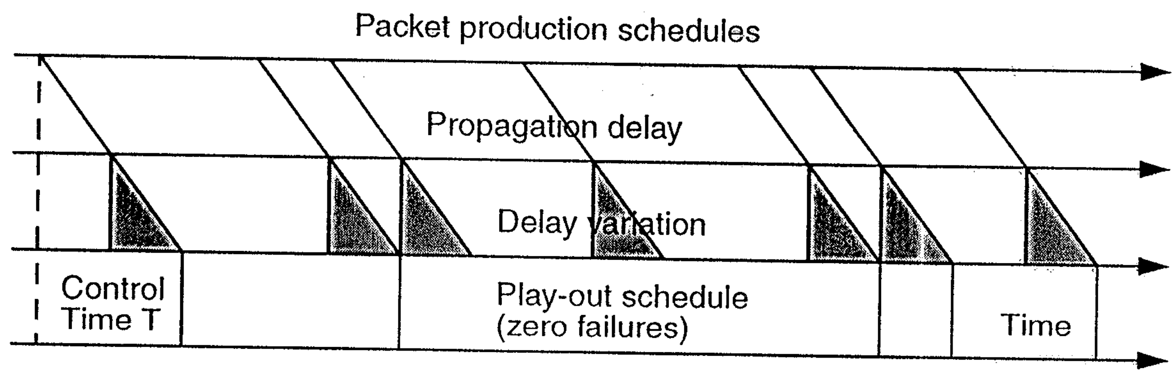
Network Randomness for Stored Multimedia Data



- The effect of network random delays on media streams
- Prefetching/control time requirements

Distributed Multimedia Networking

Objectives:

Identify suitable network resources and connection configuration so as to satisfy user's Quality of Presentation
——>Trade–off Analysis between Network Technology and Desired Quality of Service

Intelligent scheduling and switching of multimedia data streams ——> Synchronization Protocols

Class of Service Selection

- Due to the diverse nature of distributed multimedia applications, different protocols will be required for different types of traffic and control data

- User selection of a protocol profile that will provide a suitable service for each traffic type

- Extension of class of service selection to encompass more user oriented functions (to provide the required flexibility in the area of error control)

- Set of options :
  - error detection and indication
  - error detection and correction
  - error detection, correction, and indication

Quality Based Delivery of Multimedia Documents

- Synchronizations
  - Server Based ← ATM (Fuji, IBM, AT&T)
  - Resource Allocation ← Fast Ethernet PC Intel
  - Feedback Synchronization
  - MPEG
- Filters
  - Hierarchical Filtering
  - Selective Filter
  - Codec Filter

Admission Control

- This is a subjective policy matter
- Continuing sessions have precedence
- QoP of both new and continuing clients may be adjusted

Network Resource Allocation

Support multimedia traffic:

- Switch bandwidth can be bottleneck of current systems
- Multimedia systems: concurrent streams
- Need to manage admission control and bandwidth allocation of multimedia streams
  - should take advantage of multimedia characteristics (OCPN)
  - should adapt to changing demands
  - should be fair among clients
  - should be efficient (Given threshold)
- Periodic application of bandwidth allocation is done

Switch Capacity Allocation for a VP

Our Approach to Synchronous Retrieval of Multimedia Data:

- Divide the total switch capacity among logical sub-channels, with a sub-channel for each connection, and
- Design an switch controller that has intelligence to make allocation decisions

QoP Quantization

- Percentage of tolerable data loss is bi-level
  - recommended values: (*soft* QoP)
  - required values: (*hard* QoP)
- Procedures for specified levels of decrease
  - Allowable transformations
  - Effect on bandwidth

Switch Bandwidth Allocation

Scheme I: Fairness Policy

- choose bandwidths to make QoP values as close to each other as possible

J_i = Proportion of stream data dropped (assigned by system)
q_i = Quality of presentation parameter (requested by client).

In this study q_i is the reliability parameter.

Switch Bandwidth Allocation

Scheme I (cont.):

Formulate the bandwidth allocation policy as a nonlinear program:

$$\begin{aligned} &\text{Minimize} && \sum_{p,q=1}^{n} \sum_{p<q} (J_p - J_q)^2 \\ &\text{Subject To} && \sum_{i=1}^{n} J_i s_i = R - C \qquad\qquad (\mathbf{A}) \\ & && 0 \leqslant J_i \leqslant 1 - q_i;\ \ i = 1, \ldots, n \end{aligned}$$

- If all $J_i s$ are equal, objective function is zero, and constraints are satisfied.
- If $J_i s$ are not equal, they are still driven close to each other.
- The above NLP may not have a solution.

QoP Quantization

- Percentage of tolerable data loss is bi-level
  - recommended values: (*soft* QoP)
  - required values: (*hard* QoP)
- Procedures for specified levels of decrease
  - Allowable transformations
  - Effect on bandwidth

Switch Bandwidth Allocation

Scheme I (cont.):

If (A) has no solution, and QoP values are soft, there is a solution between these two feasibility regions:

$$\sum_{i=1}^{n} J_i s_i = R - C$$

$$0 \leqslant J_i \leqslant 1 - q_{\max}; \; i = 1, \ldots, n$$

((**infeasible**))

$$\sum_{i=1}^{n} J_i s_i = R - C$$

((**alwaysfeasible**))

$$0 \leqslant J_i \leqslant 1; \; i = 1, \ldots, n$$

$q = \frac{C}{R}$, for all clients.

If QoP values are hard, we may still be able to fit a proportion of the clients.

Scheme I: Procedure

$\mathcal{N}$: set of new clients; $\mathcal{C}$: set of continuing clients

1. If $\mathcal{N}$ may fit without droppage, allocate s_i.
2. If $\mathcal{N}$ does not fit, then

2.1 Solve NLP($\mathcal{N}$, $QoP_{\mathcal{N}}$)

2.2 If no solution, solve NLP($\mathcal{N} \cup \mathcal{C}$, $QoP_{\mathcal{N} \cup \mathcal{C}}$)

2.3 If no solution, reduce QoP in $\mathcal{C}$ to their minimum values, and admit partial clients from $\mathcal{N}$, if possible

Note: Extractable slack is given by:

E $= \Sigma(a_i - q_i s_i)$, for $i \in \mathcal{C}$.

Switch Bandwidth Allocation

Scheme II:

Disadvantages of the NLP approach::

- Making $J_i s$ close to each other may be inappropriate, and penalize clients with high QoP requirements
- Large number of situations may not admit a solution
- $O(n^3)$ may be expensive

Multi-Tiered Scheme:

- Bundle media traffic with similar QoP into a single tier
- Form search spaces for each tier
- Impose proportional penalties for each tier

Scheme II: Implementation

Case structure is same as Scheme I. However instead of forming NLPs, we form QoP intervals for each tier:

$[\alpha_1, 1], \cdots, [\alpha_k, 1]$

Final QoP values are determined by bisection.

$$\sum_{p=1}^{k} J_p S_p = size(\mathcal{N}) - avail(T^-)$$

$$0 \leqslant J_p \leqslant 1 - \alpha_p,\ p = 1 \cdots k.$$

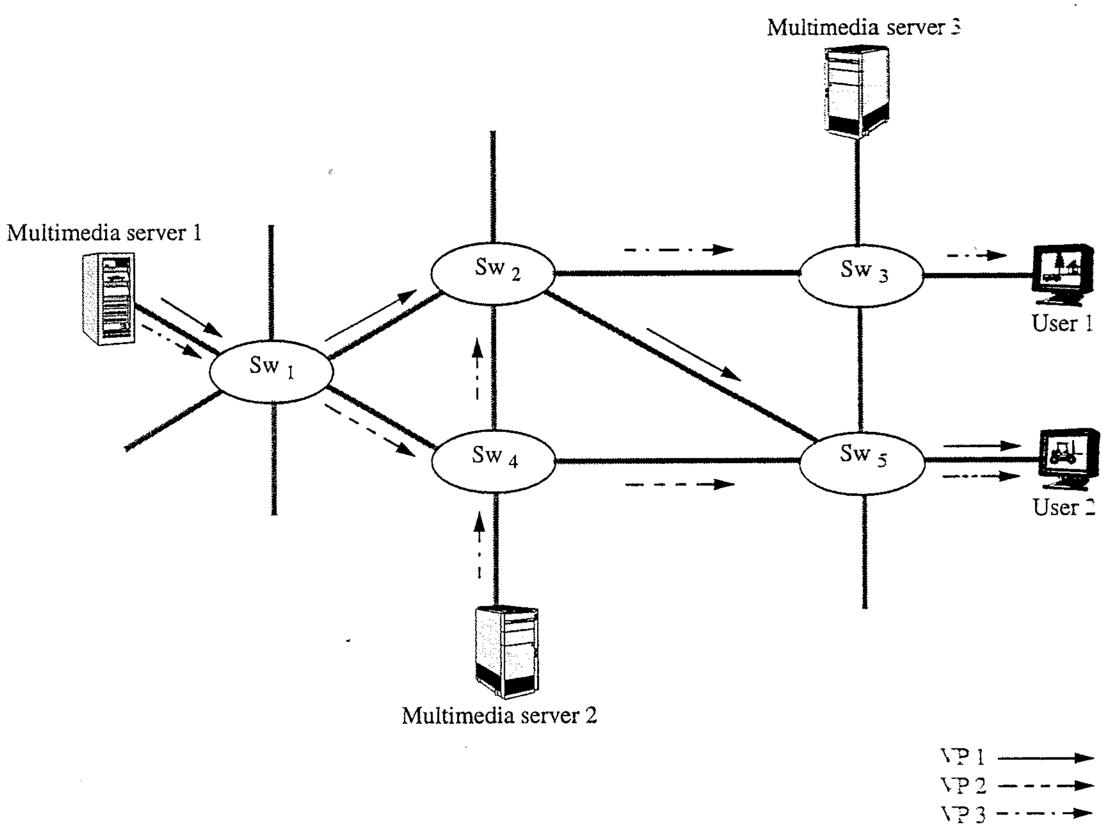
Complexity: $O(k \log \frac{(1-q_{min})}{\epsilon})$, or $O(n)$, where ϵ is the tolerance level to terminate bisections.

Network Resource controlled Synchronization

Resource allocation guarantees synchronous delivery of multimedia streams

- **allocate resources efficiently for maximum utilization**
- **Dynamic/Static capcaity & buffer allocation**
  - **Allocation is treated as an optimization problem**
- **Effective in resource constrained environments**
- **Requires processing at intermediate network nodes.**
- **High computational complexity**
- **Low client-server computational overhead**

QoP-Based VP Establishment



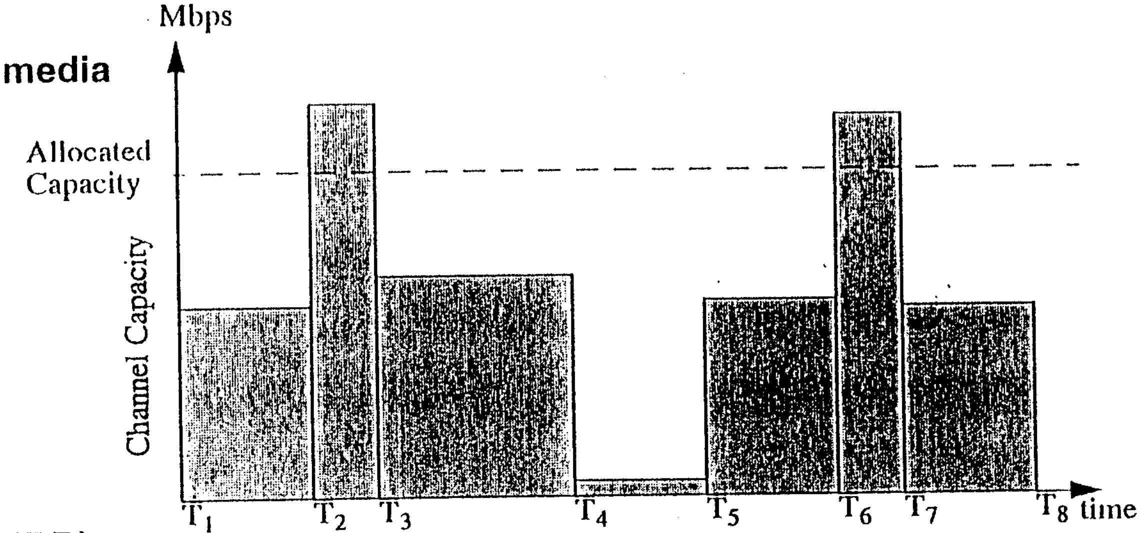
Network Resource Allocation

Support multimedia traffic:

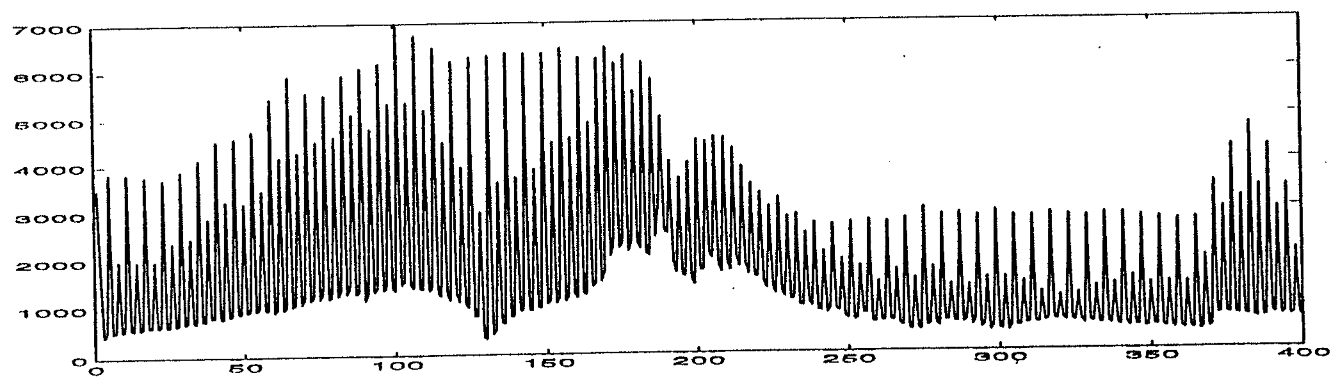
- Switch bandwidth can be bottleneck of current systems
- Multimedia systems: concurrent streams
- Need to manage admission control and bandwidth allocation of multimedia streams
  - should take advantage of multimedia characteristics
  - should adapt to changing demands
  - should be fair among clients
  - should be efficient
- Periodic application of bandwidth allocation is done

Traffic Characteristics of Media Streams in a Typical Multimedia Session

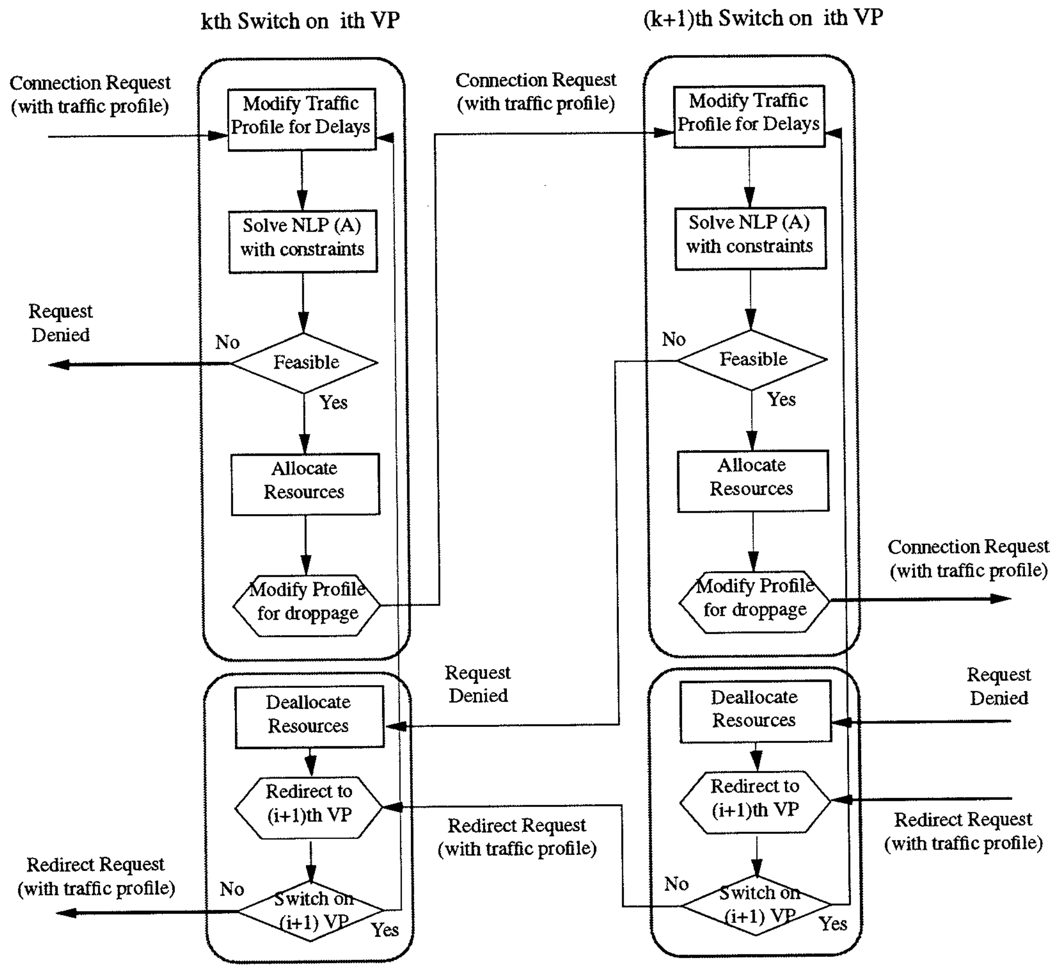
- **Complex Multimedia Information**



- **MPEG Video (VBR)**

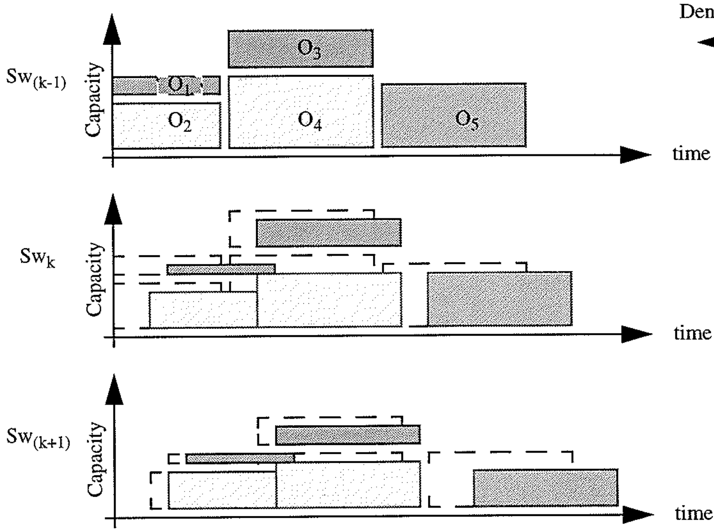


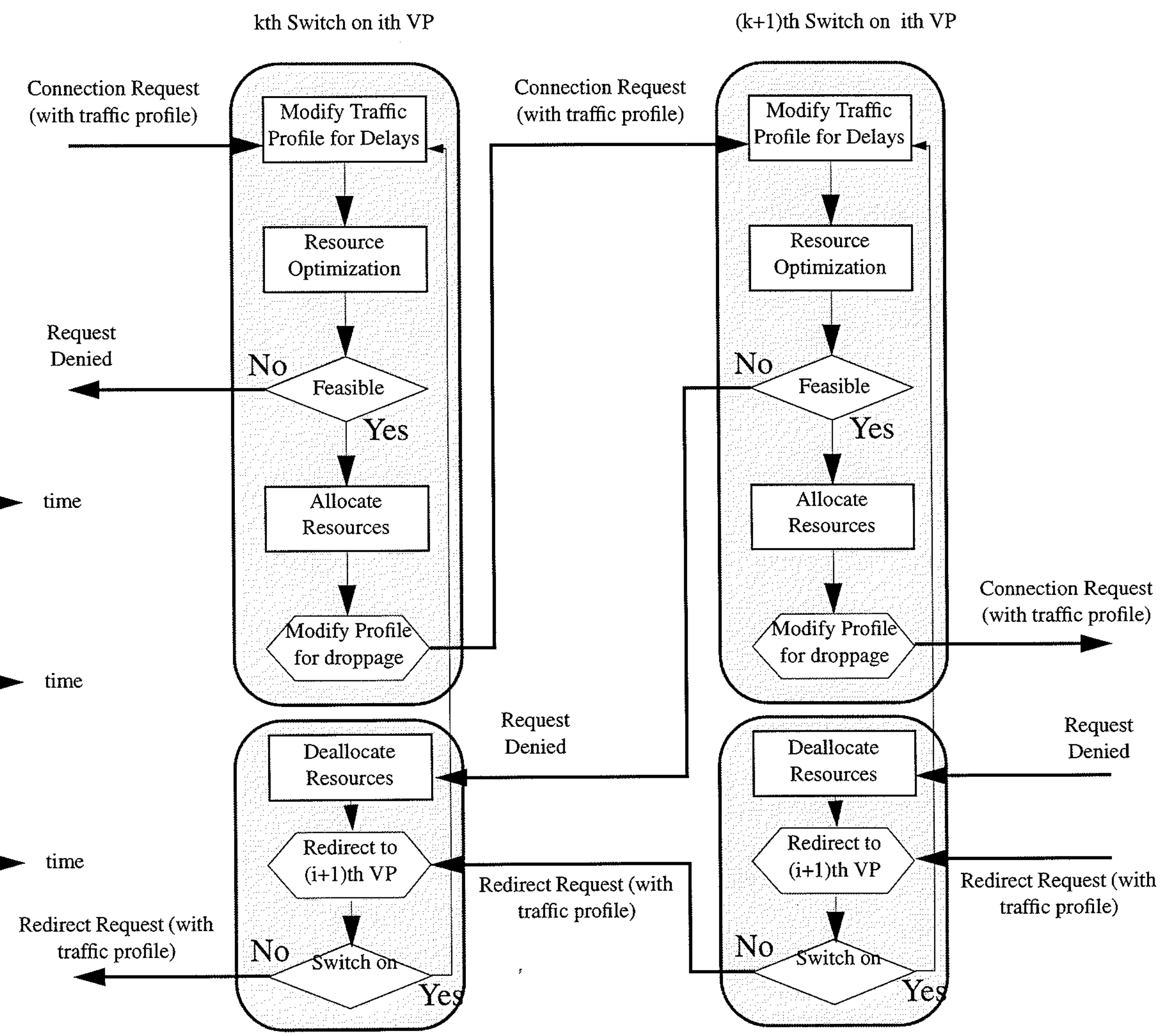
Resource Reservation Protocol



Candidate Protocols: RSVP, UPC, SRP, etc.

Dynamic Resource Allocation and Signalling Protocols





Static/Dynamic Resource Controlled Synchronization

Characteristics	Static	Dynamic
Resource Allocation Complexity	Low	High
Protocol/Signalling Complexity	Low	High
Resource Utilization	Low	High
Client-Server Participation	High	Low

End-to-End Synchronization*

Objective is to schedule the transmission for synchronized retrieval of multimedia data elements by optimizing some objective function

Mechanism:

- Network bandwidth is statically partitioned into multiple channels with capacities that may be unequal. This allows different types of media to be retrieved at different rates.
- The users' requested multimedia data is retrieved based on their deadlines

**IEEE Journal on Selected Areas in Communications*, September, 1996.

Server Based Synchronization

Given network resources, it uses a deadline based scheduling approach

- **Prefetch**
- **Delayed Presentation**

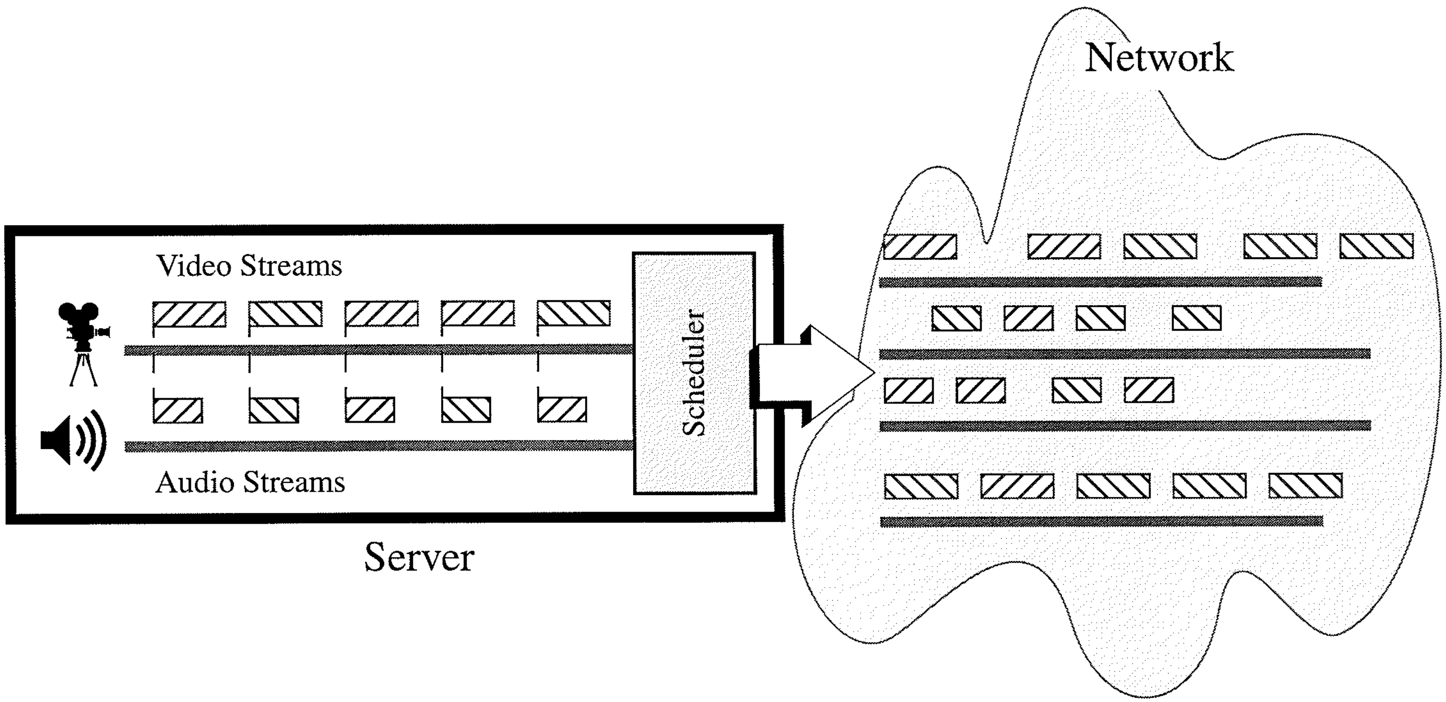
Objective is to

- **minimize presentation delay**
- **minimum network resource requirement**

Characteristics

- **Optimization is a NP hard problem**
- **Efficient algorithms exist**

Server Based Synchronization



Scheduler needs to optimize transmission scheduling decisions for media streams

Formal Temporal Specification Models

1. A storage model : Object Compostion Petri-net Model (OCPN)
   - Augmented timed, marked Petri-net
   - Able to capture all the necessary temporal relations among multimedia objects
2. A Synchronization and Communication Model : Extended OCPN
   - Transmitter-XOCPN : Specify transmission schedule at the server for each established channel
   - Receiver-XOCPN : Specify playout schedule and synchronization among multiple streams at the client site

Purdue University

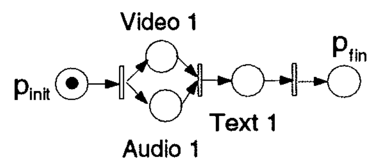
A Synchronization Mechanism at the Server

1. Identify a set of channels with propoer network QoS's that is required to transfer a multimedia document.
2. Issue connection requests for the required channels.
3. Classify a network system according to the established (virtual) channels.
4. Generate Transmitter-XOCPN: Find a set of transmission schedules of multimedia data for the established channels using server-based scheduling scheme.
   (Each schedule corresponds to a thread in the Transmitter-XOPCN)

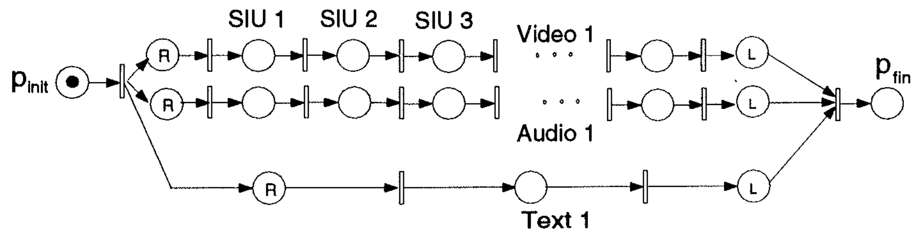
Scheduling for Synchronization

- A multimedia stream consists of smaller units that are significant to the play-out process. These allow fine-grained synchronization and are referred to as *Synchronization Interval Units* (SIUs)
- Availability of SIUs at the client should be timely
  - late SIUs: jitter or data loss
  - early SIUs: buffer overflow and data loss
- Timely delivery of SIUs may be posed as scheduling problem:
  - schedule n SIUs over m channels in order to optimize an objective function.

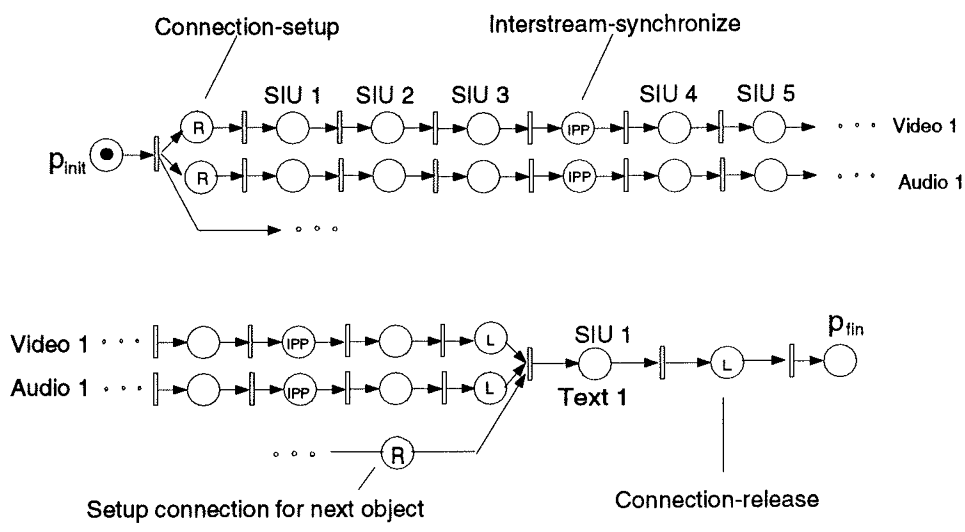
Temproal Synchronization Models



(a) OCPN



(b) Transmitter XOCPN at the server site



(c) Receiver XOCPN at the client site

IEEE Network, vol. 8, pp. 52–61, January/February 1994.

Scheduling Related Parameters

- Data Unit i;
  - Size : s_i
  - Playout deadline : d_i
- Channel j;
  - Throughput rate j : c_j
  - Average data unit reassembly delay : Δ_j

Scheduling Problem

- Suppose data unit i is scheduled for transmission on channel j at time S_j^i according to some scheduling policy.

- Average arrival time A_i of data unit i at the client site:

$$A_i = S_j^i + \frac{s_i}{c_j} + \Delta_j$$

- The *tardiness* of data unit i with respect to its playout deadline

$$T_i = \max\{0, A_i - d_i\}$$

Objective

Minimize maximum tardiness

$$T_{\max} = \max_{1 \le i \le n} \{T_i\} = \max_{1 \le i \le n} \{\max(0, A_i - d_i)\}$$

- Problem: $(Q/\Delta_j/T_{\max})$
  Given a multimedia document that consists of n data units, and m channels, find a m-channel schedule for data units that yields the minimum $T_{\max}$.

A Condition for Optimality

There exists an optimal schedule for the $Q/\Delta_j/T_{\max}$ problem in which the data units scheduled on each channel are in earliest due date (EDD) order.

Multimedia Scheduling Objectives

Various Objective Functions for Optimization of Scheduling of Multimedia Frames Include the Following:

- Minimize the completion time (makespan), C_{max}
- Minimize the number of late SIUs (ΣU_i)
- Minimize the number of early SIUs
- Minimize early and late SIUs
- Minimize total weighted tardiness, $\sum_j T_j$
- Minimize maximum lateness, $L_{max} = \max_j \{L_j\}$
- Minimize average lateness, $avg\{L_j\}$
- Minimize total weighted tardiness, $\sum_j w_j T_j$

Complexity of SIU Scheduling

- Determination of 2-Channel C_{max} is NP-hard. has been shown in previous work by relating to Uniform Parallel Processor scheduling problem (Garey and Johnson, 1979).
- C_{max} reduces to ΣU_i (Lageweg, 1982, CACM)

Therefore scheduling problems under our consideration are all NP-hard. Since efficient polynomial time solutions for such scheduling problems are not known, we look for efficient heuristic solutions for SIU scheduling.

Hueristic Scheduling Algorithms

- **Sort multimedia objects according to their presentation deadlines**
- **Schedule an object on a channel that can deliver it in the most timely manner**

Characteristics of these algorithms

- **Greedy in nature**
- **Low complexity - suitability for real-time applications**
- **Most of them have worst case bounds within a factor of 2 of the optimal solution**

C_{max} Heuristic (A)

- Sort SIUs in the order of their non-decreasing playout deadlines
- Schedule SIUs by ECT (earliest completion time) strategy. Thus we select channel j such that

$$j = \arg \min_{1 \leqslant k \leqslant m} \{L_k + \frac{s_i}{c_k} + \Delta_{ik}\}.$$

We wish to minimize $L_k + \frac{s_i}{c_k} + \Delta_{ik}$ among all channels.

- Scheduled retrieval time of SIU i is L_j
- Arrival time in display memory is $A_i = L_j + \frac{s_i}{c_j} + \Delta_{ij}$

ΣU_i Heuristic (B)

- Sort SIUs in the order of their non-decreasing playout deadlines
- Schedule to reach *at* their deadline
- Scheduled retrieval time,
  $r_i = \max\{L_j, (d_i - \frac{s_i}{c_j})\}$
- Arrival time at display is
  $A_i = \max\{(L_j + \frac{s_i}{c_j}), d_i\} + \Delta_{ij}$

ΣU_i Heuristic (C)

- Sort SIUs by increasing playout deadlines
- Schedule at the deadlines
- Find maximum lateness
- Advance schedule by a proportion of maximum lateness, L_{max}.

ΣU_i Heuristic (D)

- Same as Heuristic (B), with the following differences.
- Postpone the scheduling of an SIU if it is late.
- Append the late SIUs to the end of any schedule.

Evaluation of Proposed Heuristics

We perform a quality-based evaluation. The relevant quality parameters are:

- Percentage of deadline miss. Deadline misses cause inter-stream and intra-stream synchronization failures.
- Percentage of data data loss due to buffer overflow. This is a reliability parameter.

Filter: Effect of droppage against transmission rates

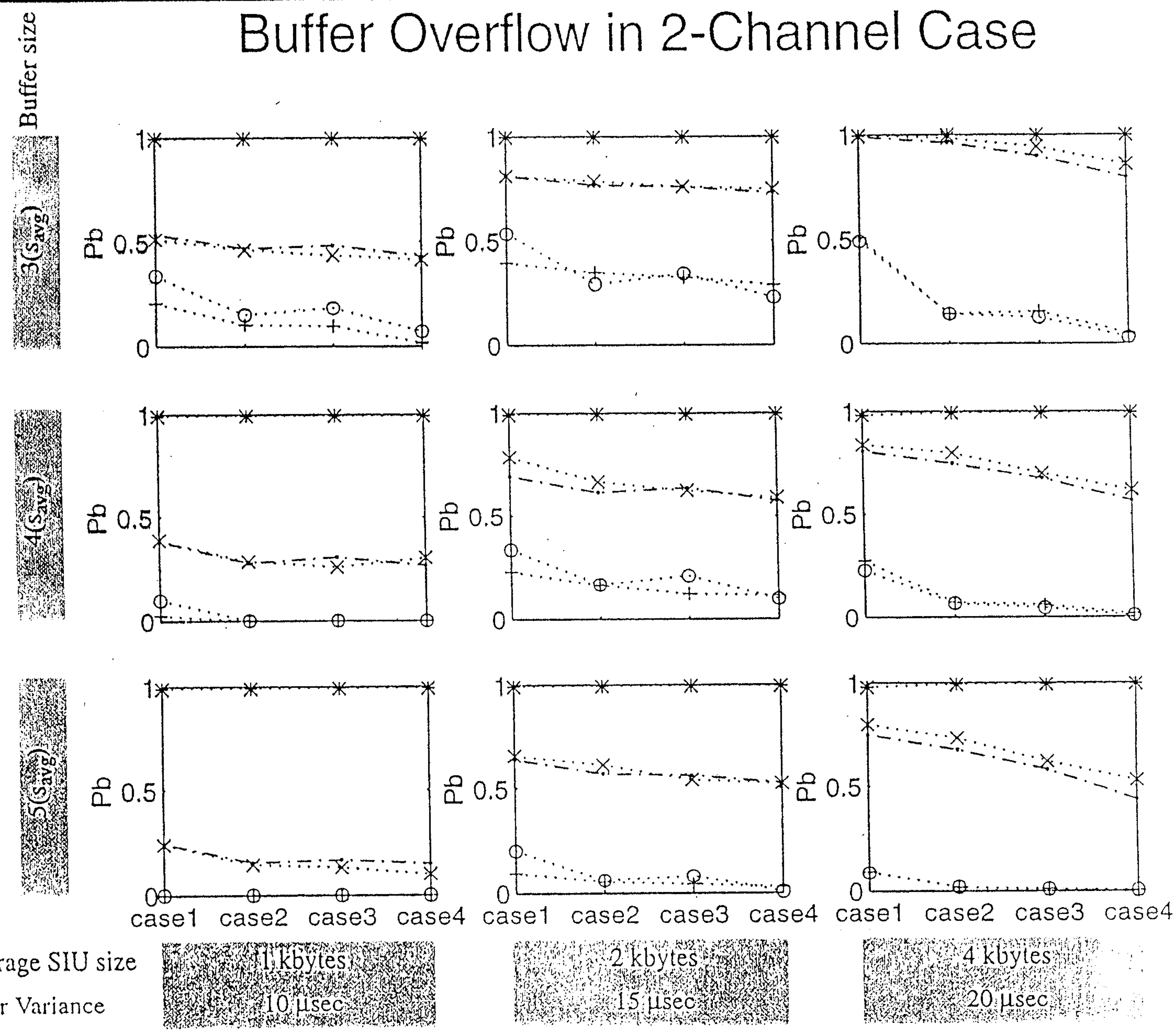
Packet Size	Unfilter Trans Rate (Mb/s)	Dropp. Ratio	Filtered Trans. Rate	Reduc. in Rate
512	4.45	0.07	4.29	0.036
1024	9.14	0.04	8.34	0.0875
1408	11.97	0.13	11.25	0.06

Simulation Environment

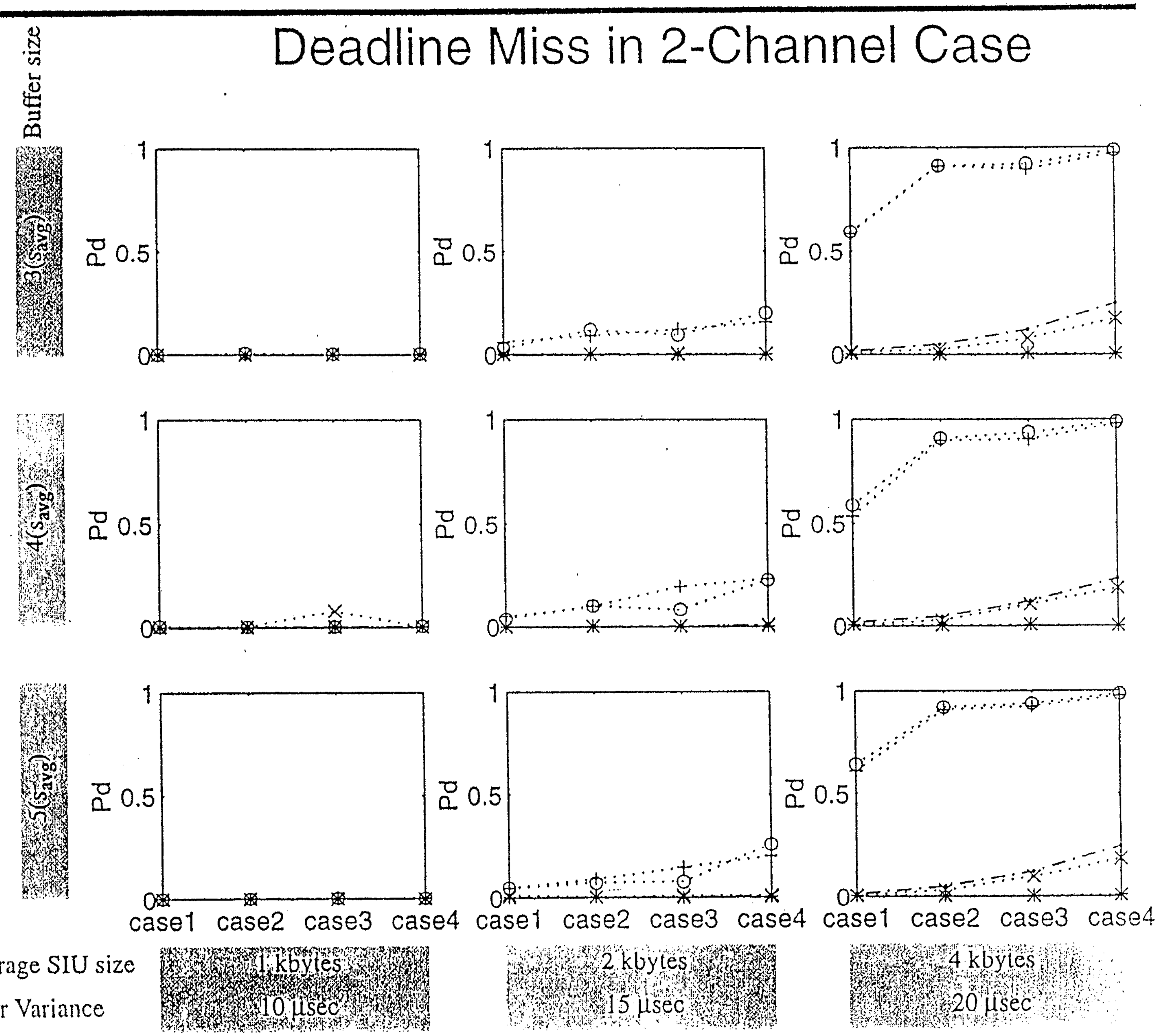
Multimedia Objects Parameters			
Object	Number of SIUs	Mean SIU Size s_{avg} (kbytes)	Range of Variations in kbytes (Truncated Normal Distributions)
Video SIUs	2500	1	$0.5 \leqslant s_i \leqslant 1.5$
		2	$1.0 \leqslant s_i \leqslant 3.0$
Audio SIUs	2500	0.268	constant

ATM Network Environment					
Delay Characteristics					
	ATM Cell Jitter Variance			10 μsec	15 μsec
Destination Buffer					
	Buffer Size		$3 \cdot s_{avg}$	$4 \cdot s_{avg}$	$5 \cdot s_{avg}$
Channel Capacity					
	Channels	Case 1 Mbps	Case 2 Mbps	Case 3 Mbps	Case 4 Mbps
2 Channels	$\mathcal{C}_1$	0.6	1.0	1.0	1.0
	$\mathcal{C}_2$	0.6	0.6	1.0	1.5
	$\mathcal{C}_1$	0.4	0.6	1.0	1.5
3 Channels	$\mathcal{C}_2$	0.4	0.4	0.6	1.0
	$\mathcal{C}_3$	0.4	0.6	0.4	0.6

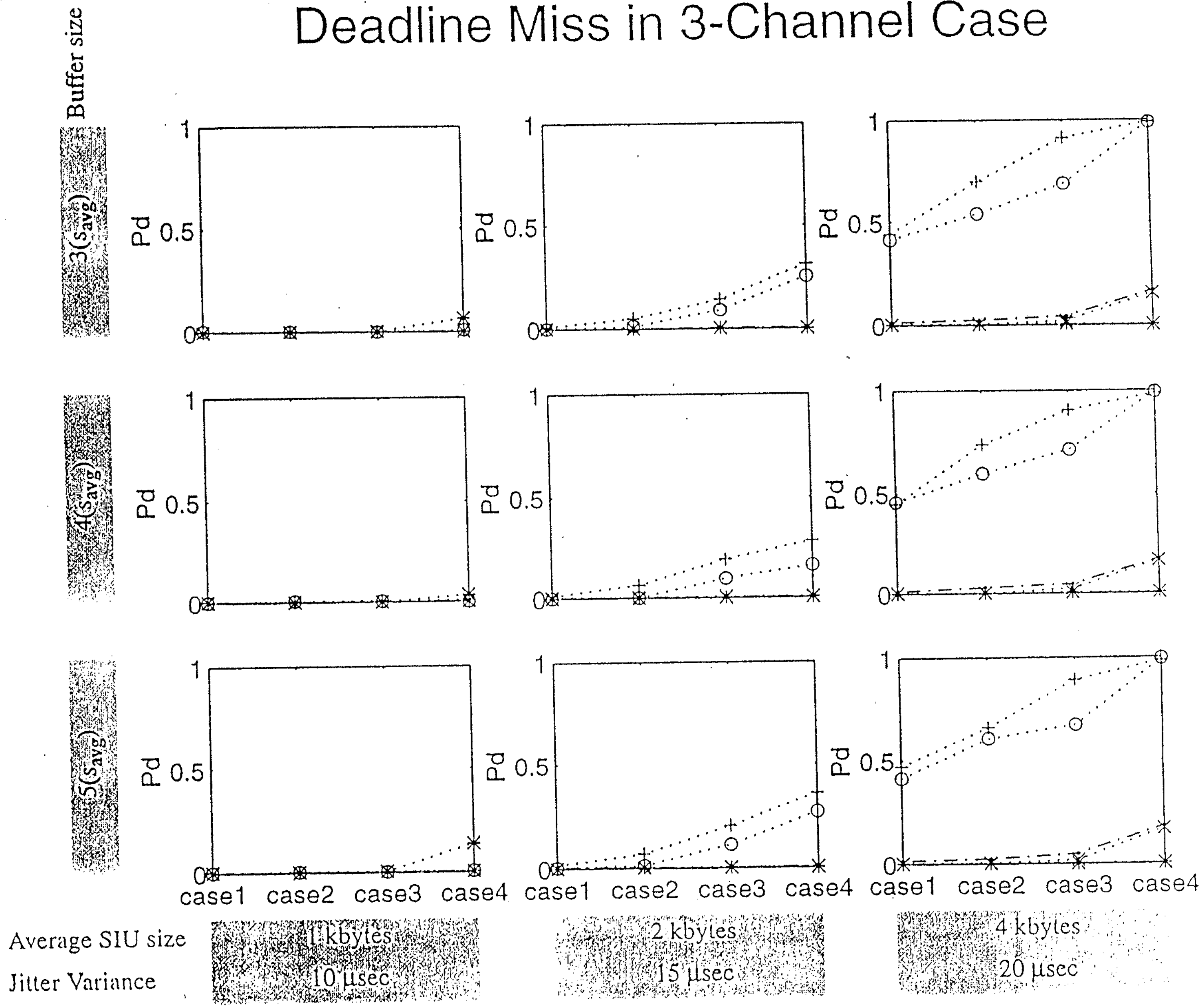
Buffer Overflow in 2-Channel Case



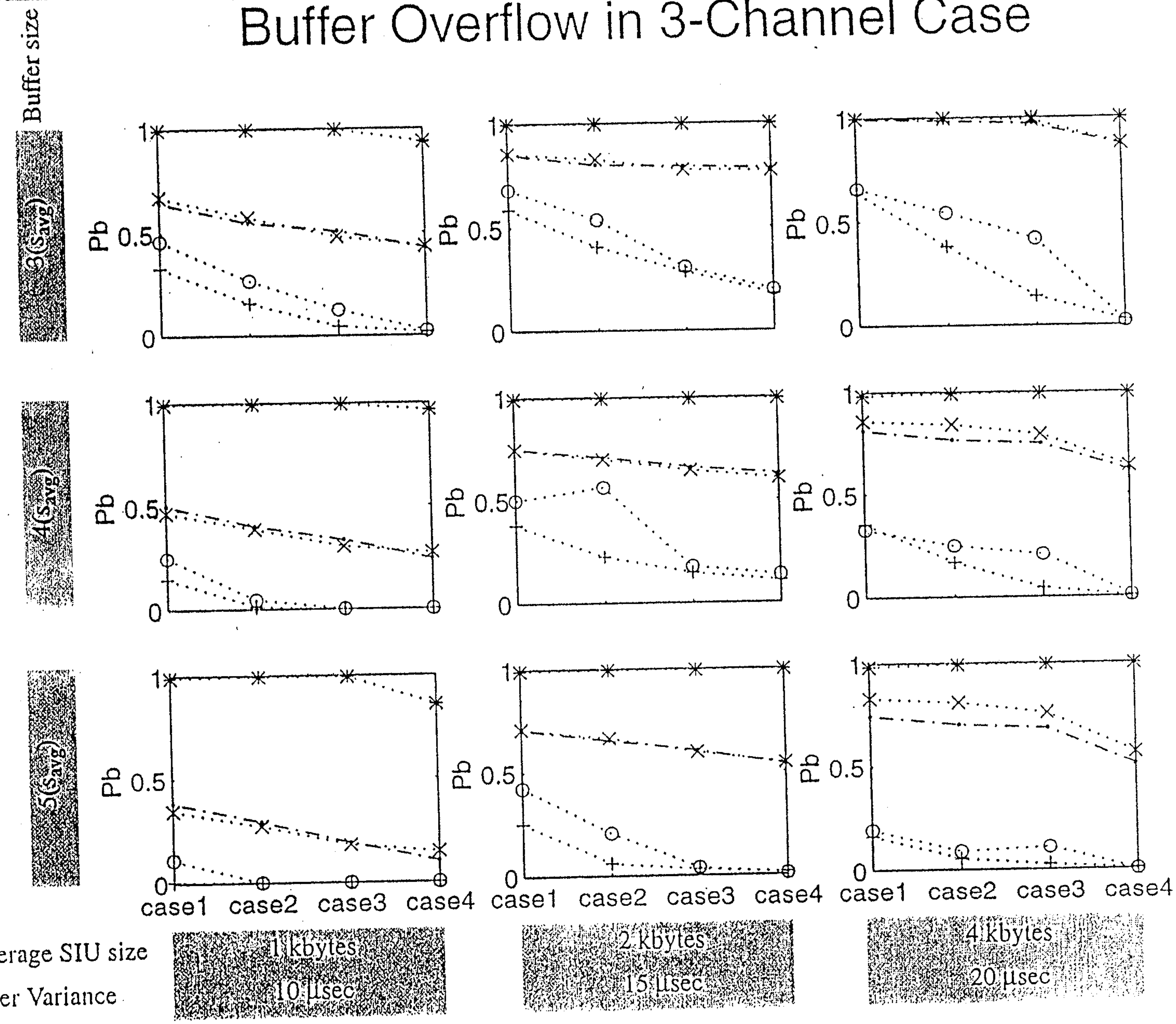
Deadline Miss in 2-Channel Case



Deadline Miss in 3-Channel Case



Buffer Overflow in 3-Channel Case



SIU Scheduling Simulation Results

- Heuristic for makespan minization has almost no deadline misses, but large memory overflows
- Heuristic for in-time scheduling has more deadline misses, and less memory overflow
- Other trends:
  - larger memories reduce buffer overflow, but do not seem to affect the level of deadline misses
  - higher random delays (δ) cause more deadline misses and more buffer overflow

Behavior of Algorithms

- Makespan minimization heuristic schedules whenever a channel becomes available. As a result deadlines misses are few. Since early SIUs need to be buffered in memory, there are large memory overflows.

- In-time scheduling heuristic schedules SIUs close to the deadlines. Early SIUs miss few deadlines, but need to stay in memory for longer time, until they are consumed.

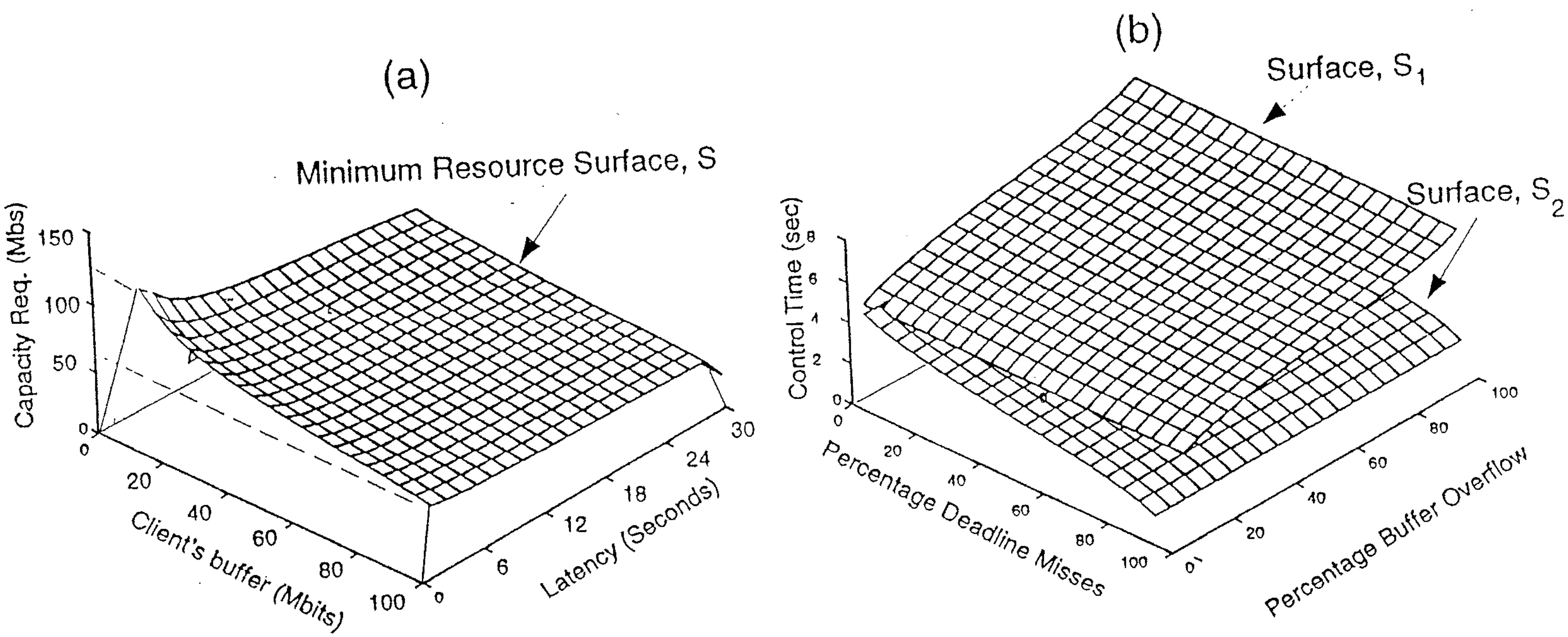
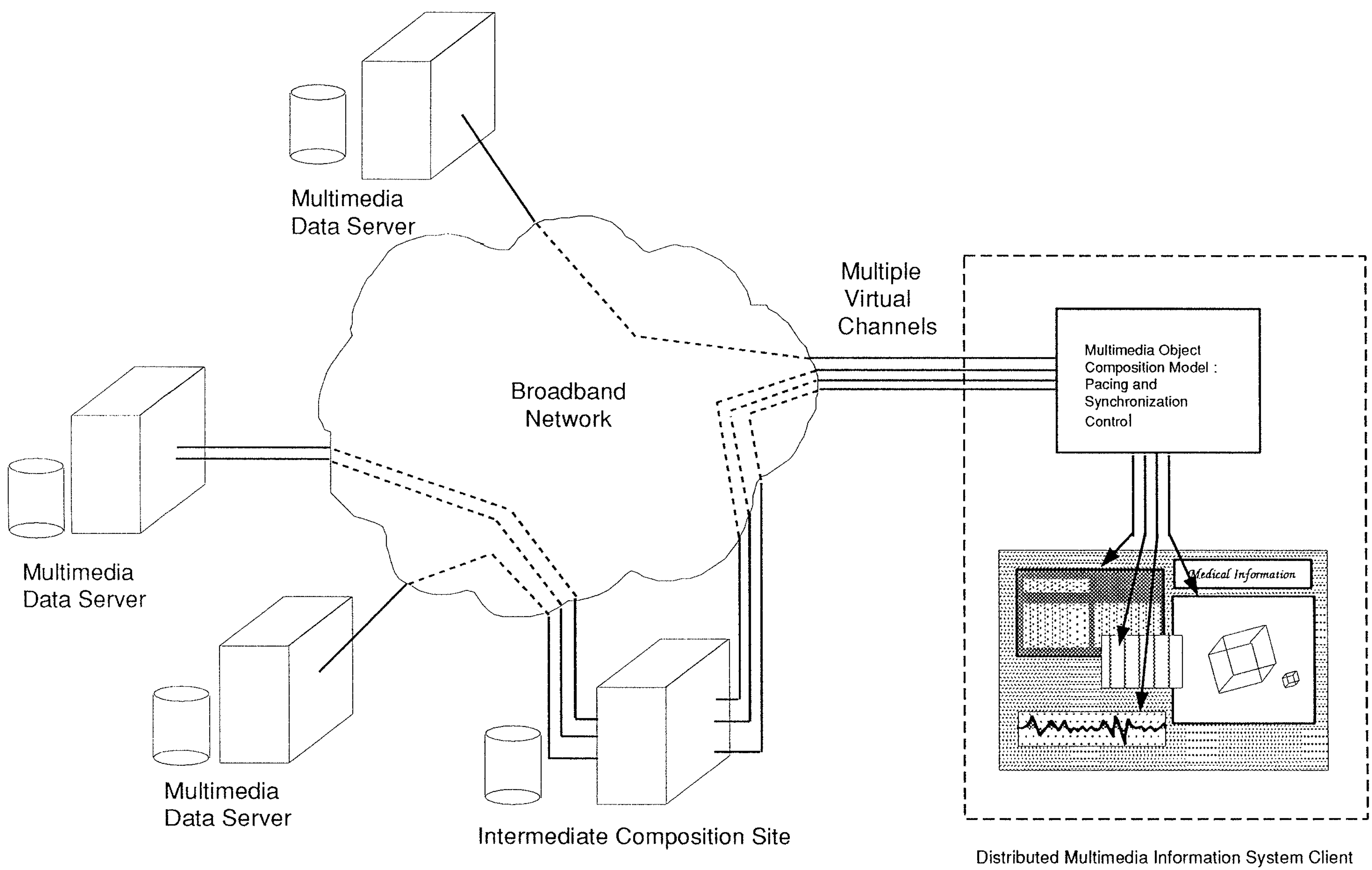


Figure : Trade-offs in resource planning, (a) Minimum resource surface (b) Schedulable region.

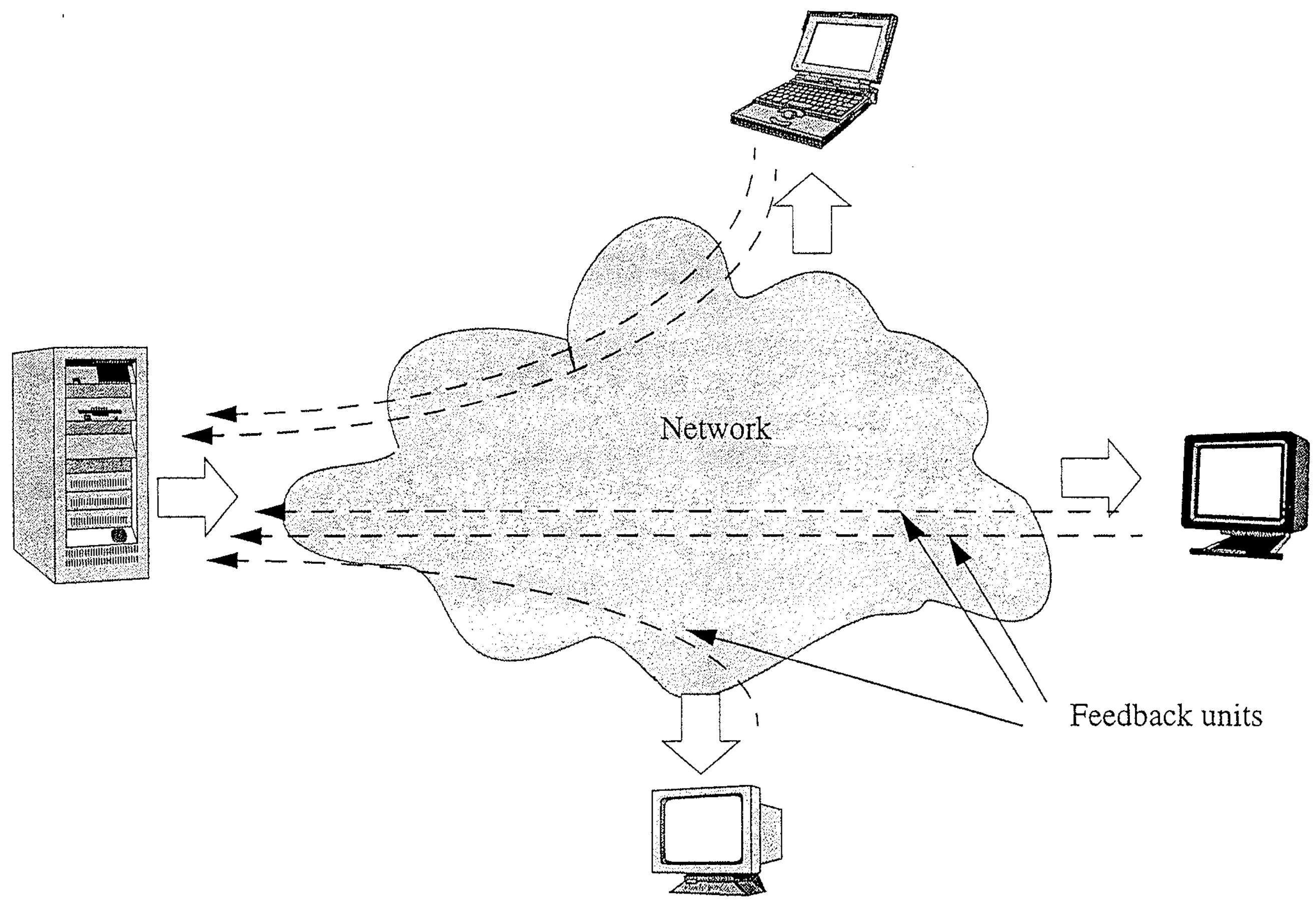
Static/Dynamic Resource Controlled Synchronization

Characteristics	Static	Dynamic
Resource Allocation Complexity	Low	High
Protocol/Signalling Complexity	Low	High
Resource Utilization	Low	High
Client-Server Participation	High	Low

Value-Added Network for Composing Distributed Multimedia Objects



Feedback Synchronization Technique



Synchronization using Feedback

- Each media player sends feedback units periodically to the server
- Server adjusts data rate of multimedia streams with respect to feedback units

Resynchronization techniques

- conservative
  - server reacts only when asynchrony exceeds some maximum tolerable limit
- aggressive
  - server reacts at the slightest change of asynchrony
- probabilistic
  - server reacts on the average. Average network delays are known a priori.

Pros and Cons

- Adaptive and flexible
- Increased load on network to carry feedback units
- Delayed response to asynchrony in WANs

Synchronization in MPEG

Audio & Video streams are multiplexed onto a single stream

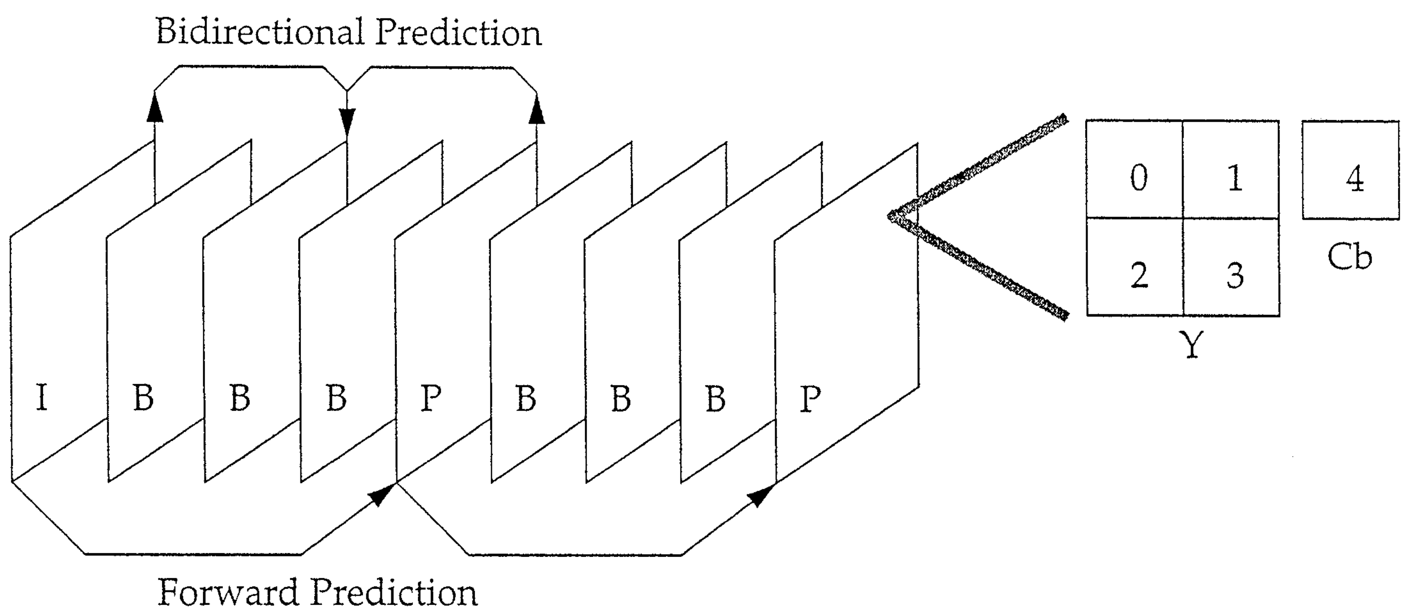
- pack layer
  - interleaved audio & video streams
  - common stream layer wrapped around a media specific compression layer
  - intra-stream synchronization is handled by this layer
- packet layer
  - independent compressed packetized multimedia streams
  - each packet is time stamped for playout
  - inter-stream synchronization layer

Features

- Synchronization information is embedded in the bitstream
- Additional synchronization mechanisms are required to overcomethe random behavior of the network

Synchronization in MPEG

Mpeg Video Representation



Group Synchronization in a Multimedia Environment

- Compensation for jitter delays
- Local clock drift