



CS 414 – Multimedia Systems Design
Lecture 13 –
Quality of Service (Part 2)

Klara Nahrstedt
Spring 2008



Administrative

- MP2 posted
- Discussion Section I on Wednesday, 2/13, 2008 (3405 SC)
- Discussion Section II on Monday, 2/18, 2008 (3405 SC)

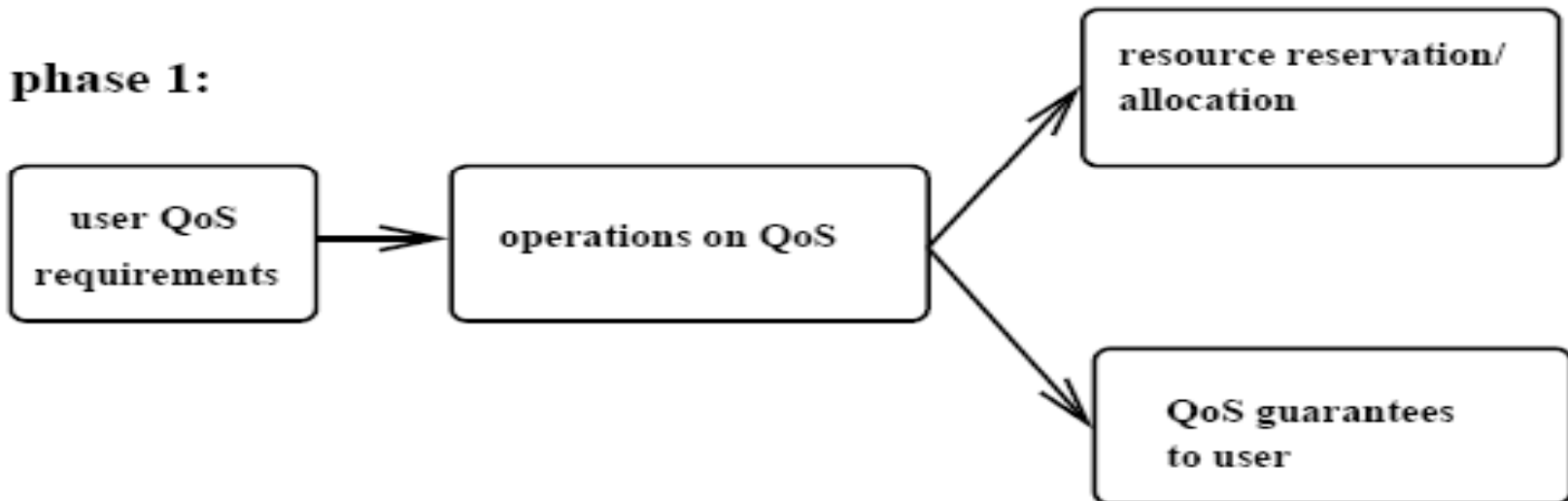


Outline

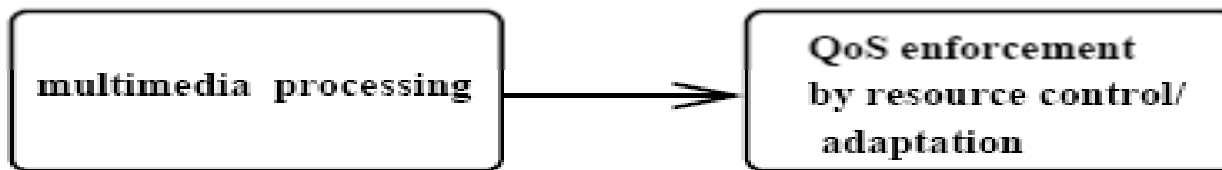
- Requirements on Multimedia
Communication and Operating System
- Real-time and multimedia
- Resource management
 - Resources
 - Quality of Service (QoS) Concept
 - Operations


Relation between QoS and Resources

phase 1:



phase 2:

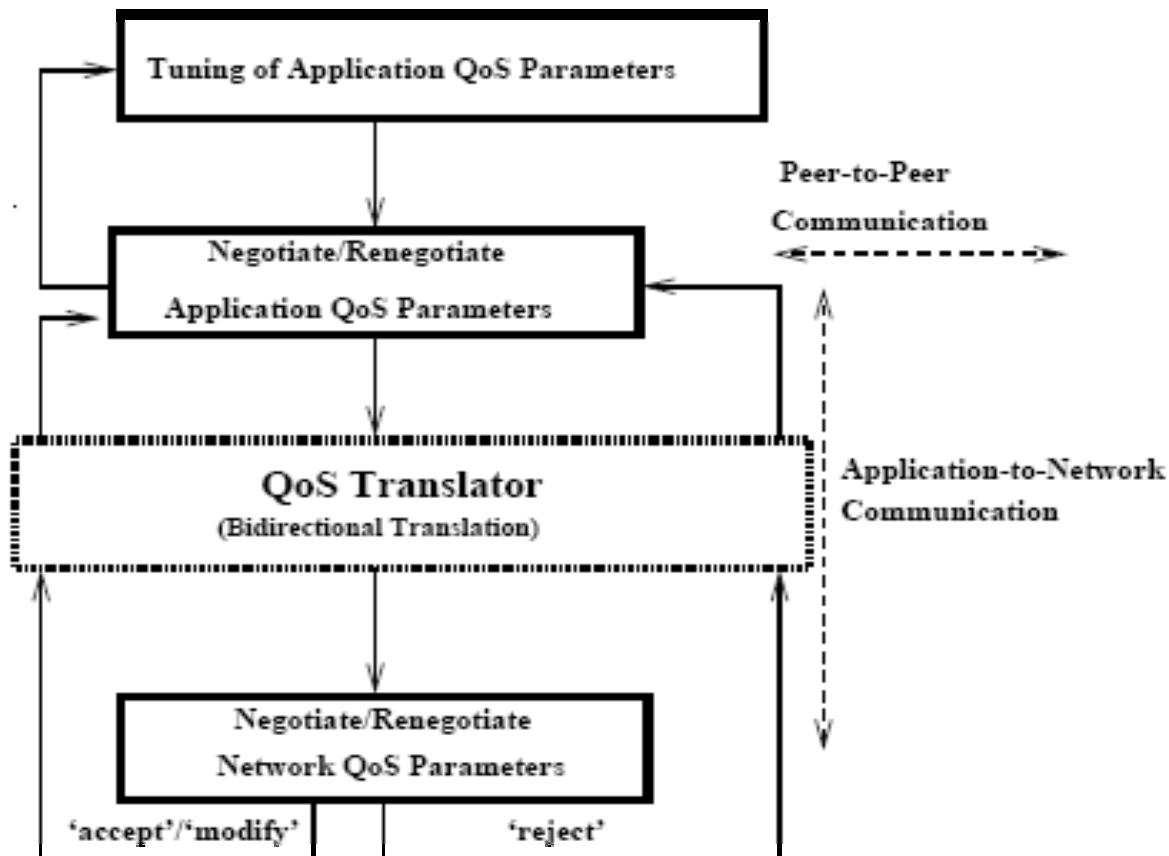




Phase 1: Establishment Phase (QoS Operations)

- QoS Translation at different Layers
 - User-Application
 - Application-OS/Transport Subsystem
- QoS Negotiation
 - Negotiation of QoS parameters among two peers/components
- QoS Routing along the end-to-end path

QoS Operations within Establishment Phase



User/Application
QoS Translation

Overlay P2P
QoS Negotiation

Application/Transport
QoS Translation

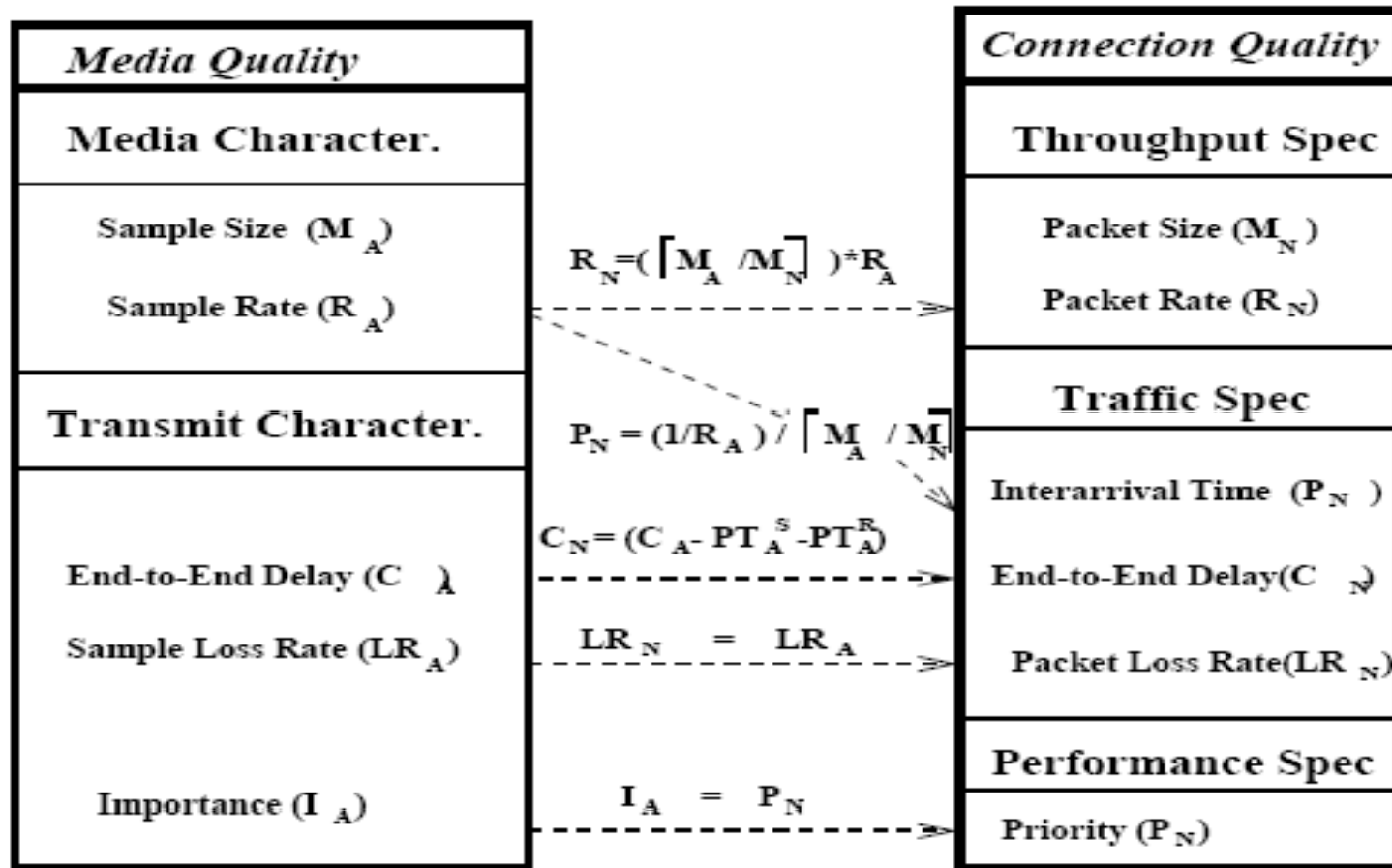
QoS Negotiation/
QoS Routing in
Transport Subsystem



Operations on QoS in Phase 1 (QoS Translations)

- Layered Translation of QoS parameters
(must be bidirectional)
 - Human (user QoS) – application QoS
 - Application QoS – system QoS
 - System QoS – network QoS
- Media Scaling
 - Transparent scaling
 - Non-transparent scaling

Layered Translation (Example)





Media Scaling (Examples)

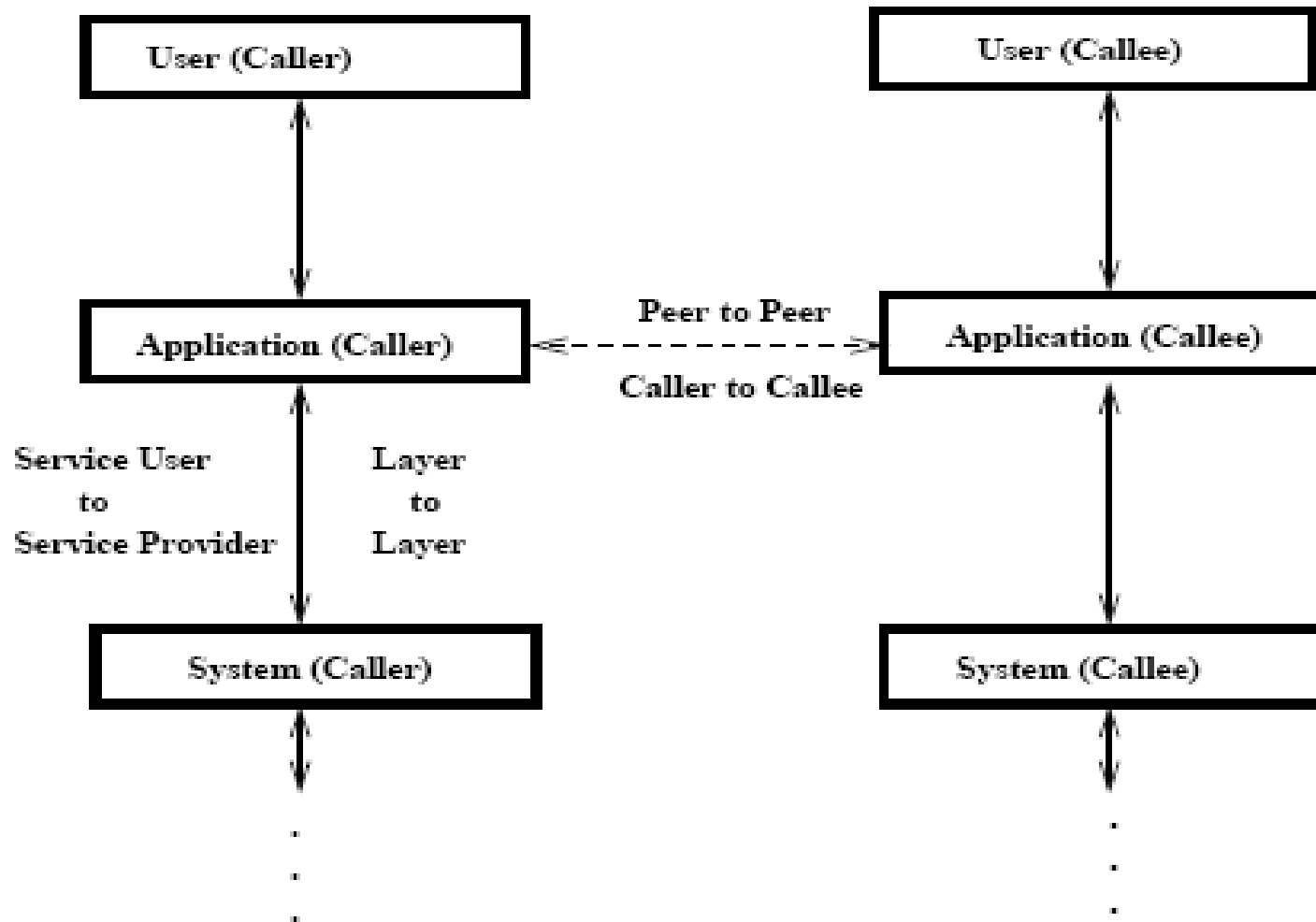
■ Audio

- Transparent scaling difficult (one hears the quantization noise)
- Non-transparent scaling should be used

■ Video

- Temporal scaling
- Spatial scaling
- Color space scaling (reduction of number of entries in color space)

QoS Negotiation

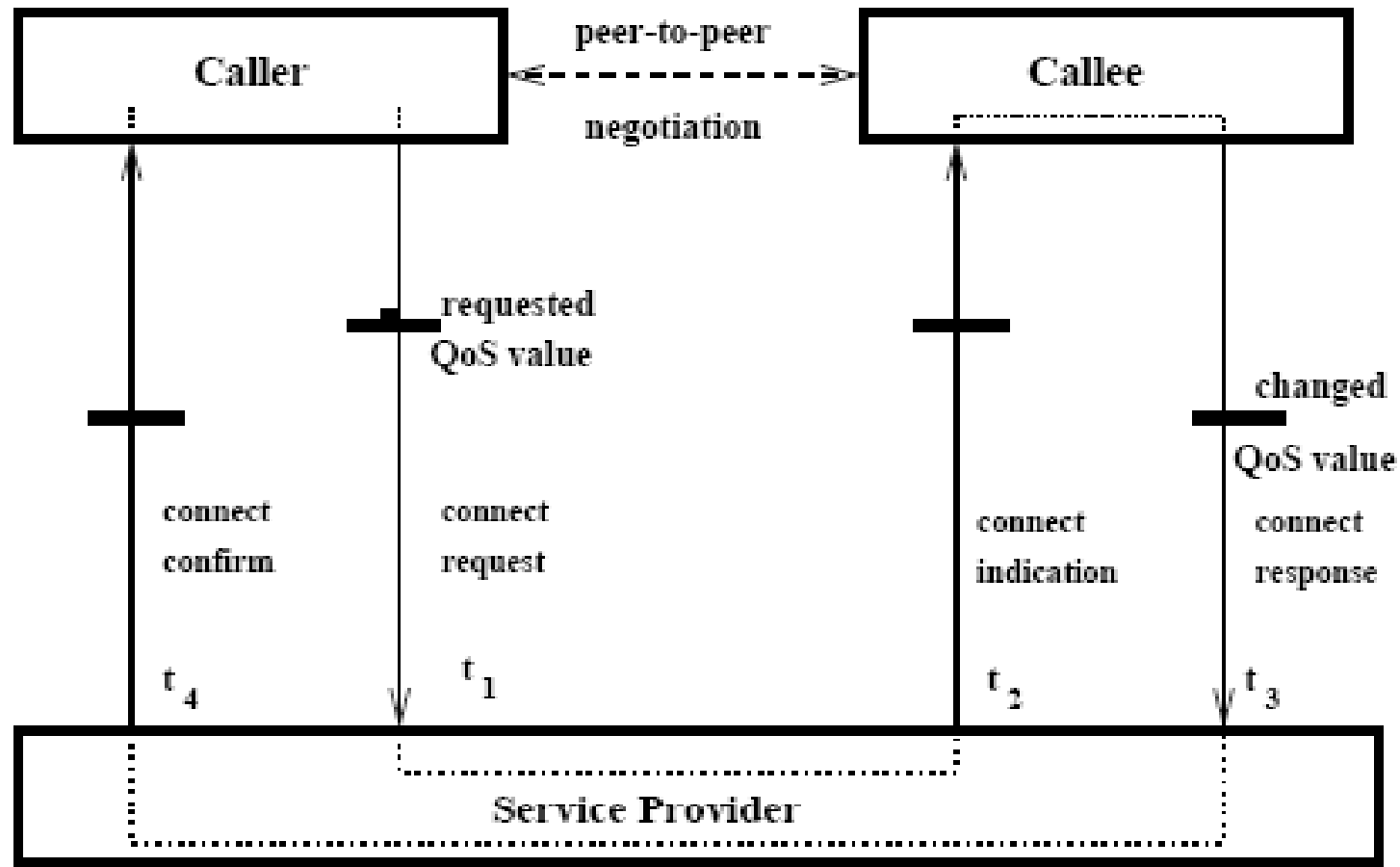




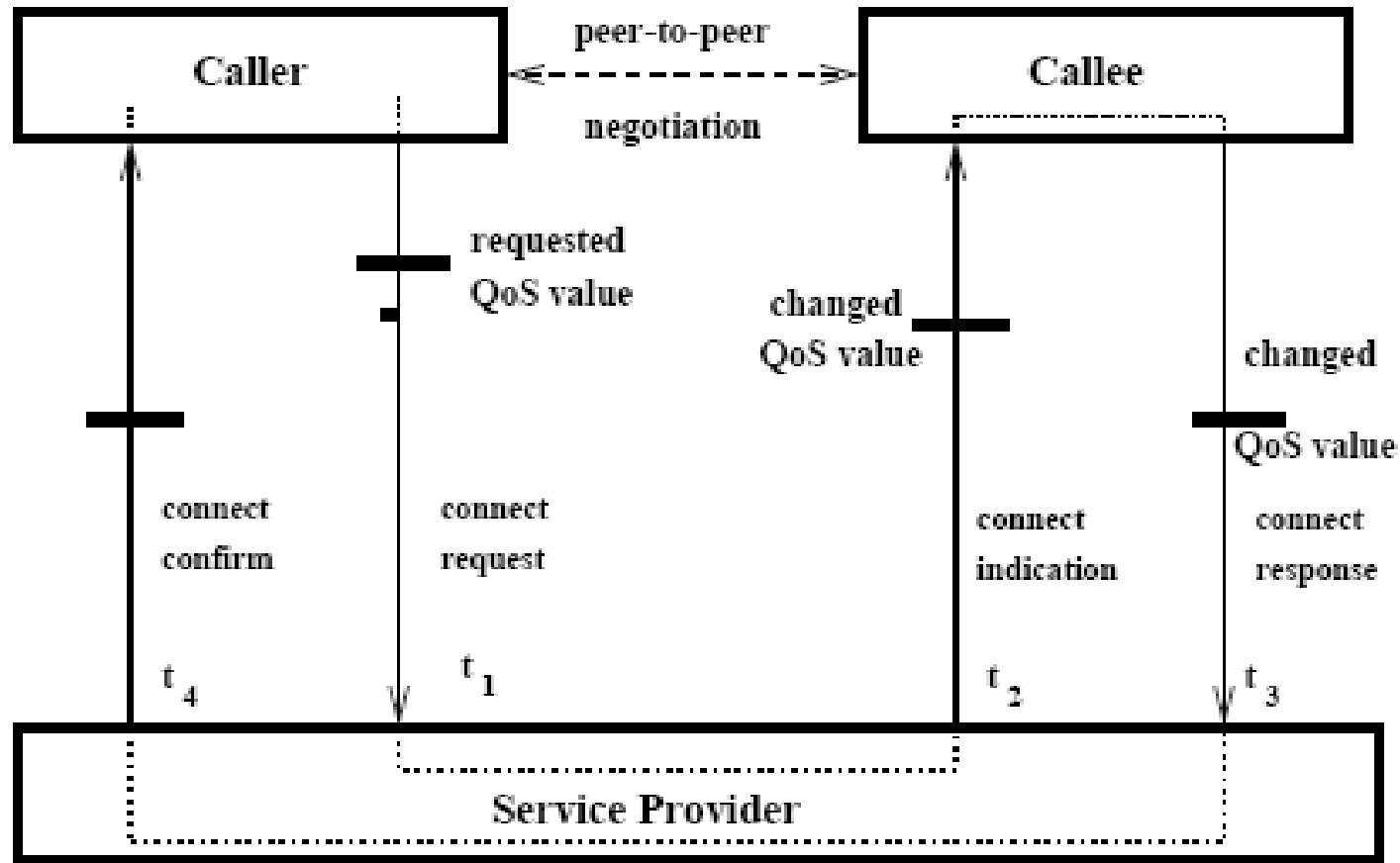
Different Types of Negotiation Protocols

- Bilateral Peer-to-Peer Negotiation
 - Negotiation of QoS parameters between equal peers in the same layer
- Triangular Negotiation
 - Negotiation of QoS parameters between layers
- Triangular Negotiation with Bounded Value

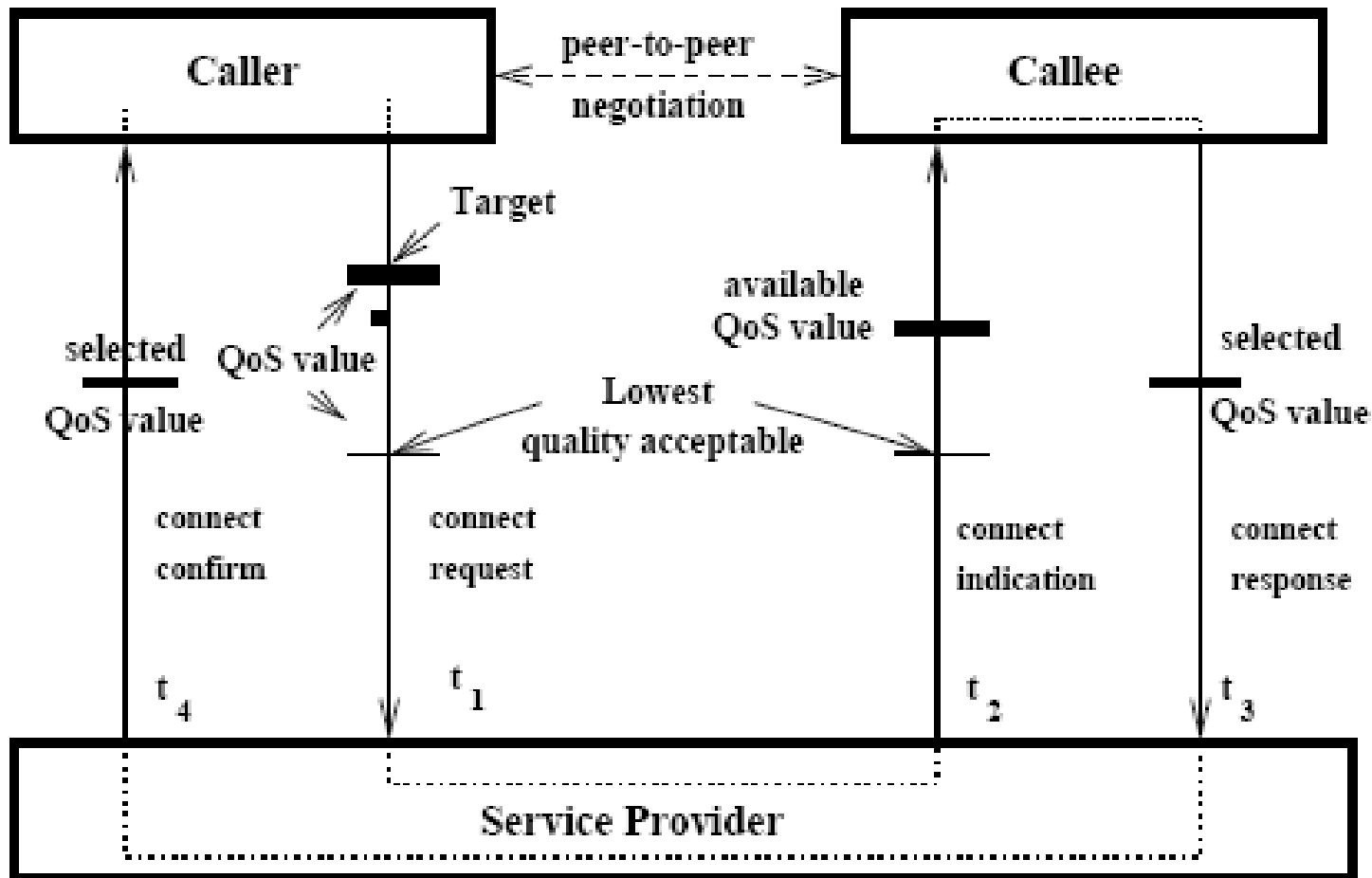
Bilateral QoS Negotiation



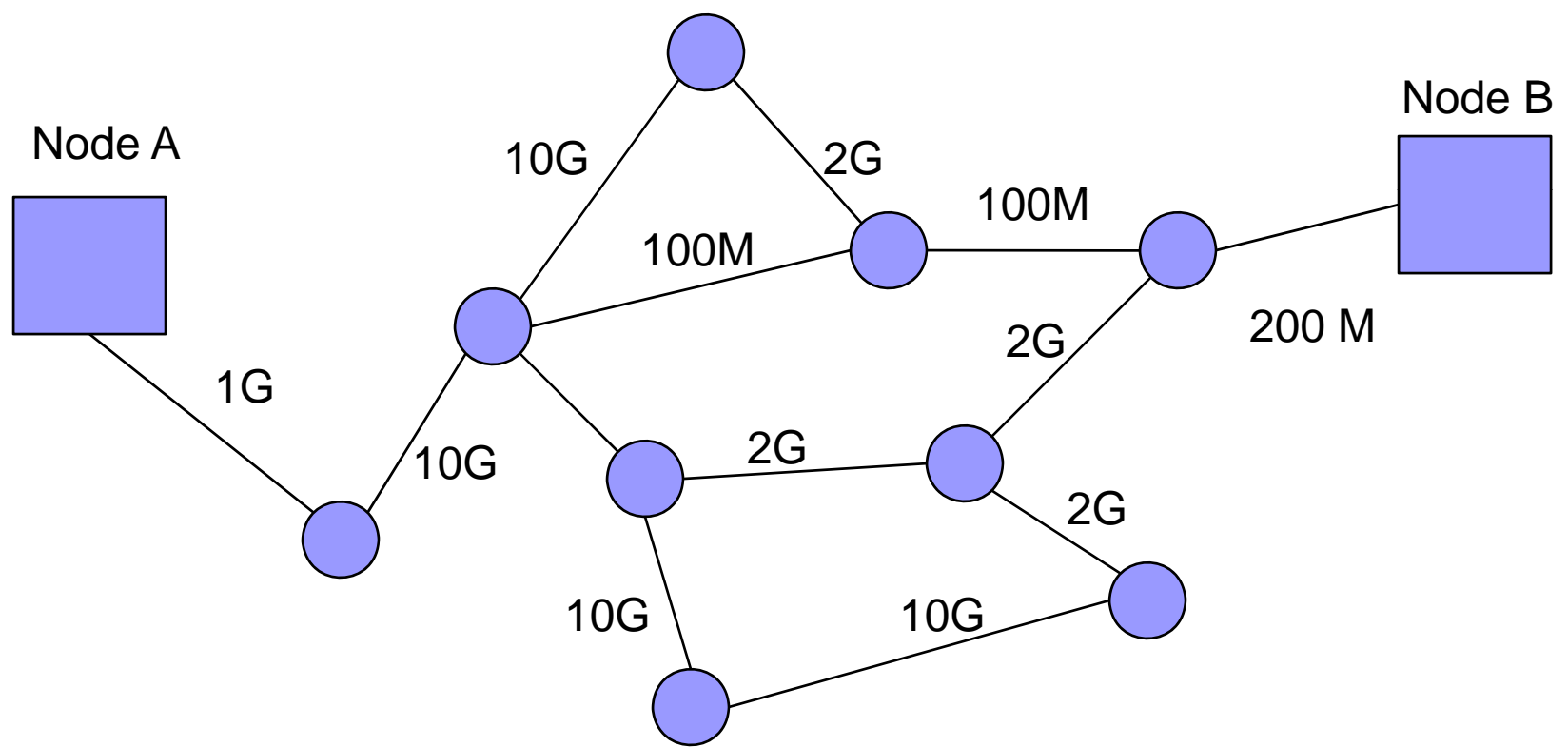
Triangular QoS Negotiation



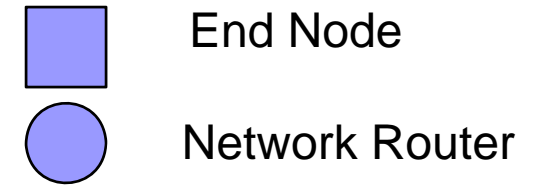
Triangular Negotiation with Bounded Value



QoS Routing



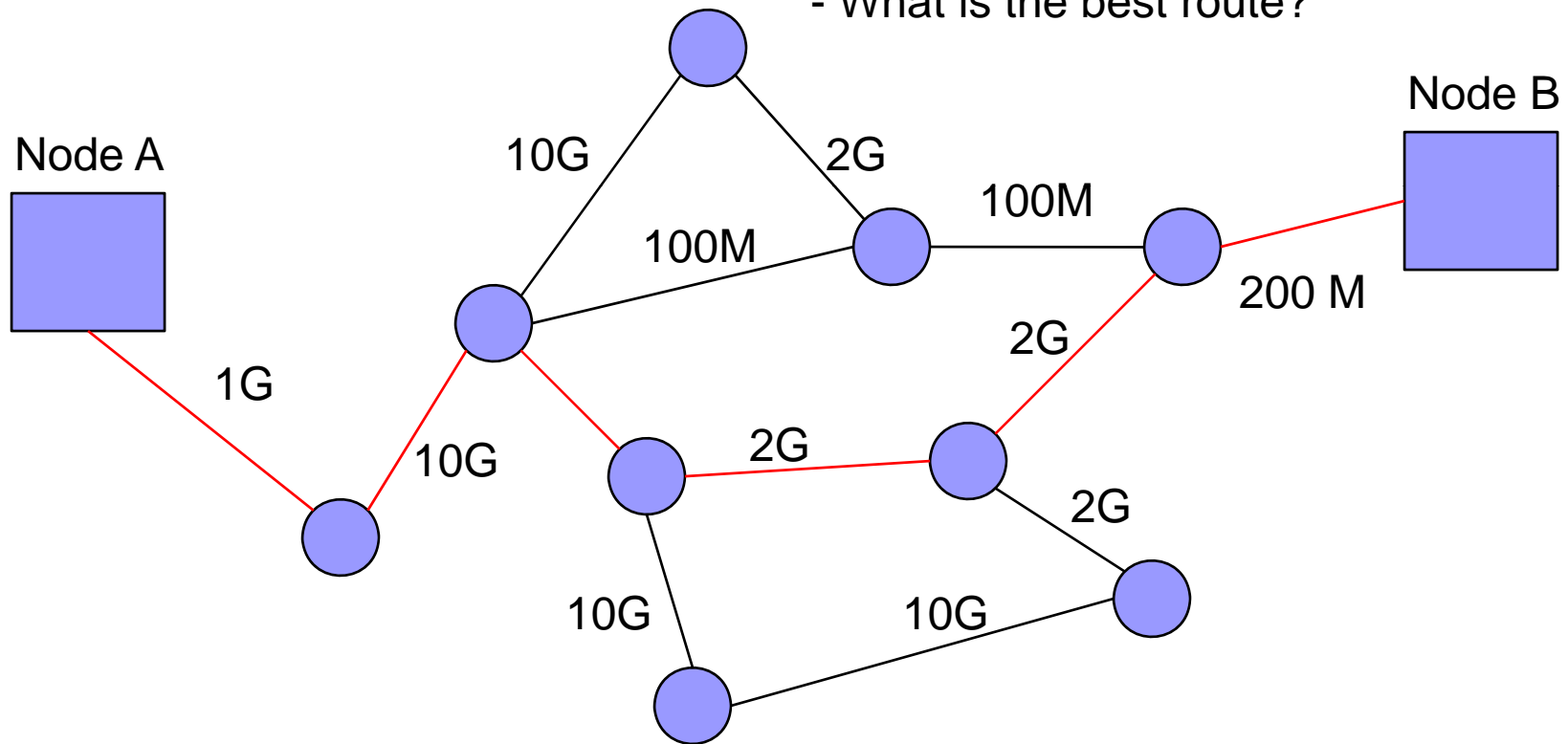
1G = 1Gbps
10G = 10 Gbps
100 M = 100 Mbps



QoS Routing

If QoS Request on a connection from Node A to B is 150 Mbps, the QoS Routing question is

- Does a route from A to B exist that satisfies the QoS requirement?
- What is the best route?




1G = 1Gbps
10G = 10 Gbps
100 M = 100 Mbps



QoS Routing

- Performed during establishment phase mostly, but also during transmission phase to adapt a route if needed
- Need to discover route (path) that meets QoS requirements such as throughput, end-to-end delay, loss rate
 - End-to-end Throughput is a min-based metric
 - End-to-end Delay is additive metric



Unicast QoS Routing

- Problem Formulation:
 - Given a source node A, destination B, a set of QoS constraints C, and possibly an optimization goal, we aim to find the best feasible path from A to B which satisfies C.
- **Bandwidth-optimization problem:** to find a path that has the largest bandwidth on the bottleneck link (widest path)
- **Bandwidth-constrained problem:** to find a path whose bottleneck bandwidth is above a required threshold value
- **Delay-optimized problem:** to find a path whose total delay is minimized
- **Delay-constrained problem:** to find a path whose delay is bounded by a required value.



QoS Routing Strategies

■ Source Routing

- Each node maintains global state and feasible path is locally computed at the source node

■ Distributed Routing

- Control messages exchanged among nodes and the state information kept at each node is collectively used for the path search

■ Hierarchical Routing

- Nodes are clustered into groups creating multi-level hierarchy
- One can use source routing within a cluster and distributed routing among clusters



Multimedia Resource Management

- Resource managers with operations and resource management protocols
 - Various operations must be performed by resource managers in order to provide QoS
- Establishment Phase
 - Operations are executed where schedulable units utilizing shared resources must be admitted, reserved and allocated according to QoS requirements
- Enforcement Phase
 - Operations are executed where reservations and allocations must be enforced, and adapted if needed



Establishment Phase Operations

- QoS to Resource Mapping
 - Need translation profiles
- Resource Admission
 - Need admission tests to check availability of shared resources
- Resource Reservation
 - Need reservation mechanisms along the end-to-end path to keep information about reservations
- Resource Allocation



Continuous Media Resource Model

- One possible resource utilization model for multimedia data – Linear Bounded Arrival Process Model (LBAP)
- LBAP models message arrival process:
 - M – maximum message size (in bytes)
 - R – maximum message rate in messages per second
 - B – maximum burstiness (accumulation of messages)



LBAP Resource Model

- If we have (M, R, B) , we can predict utilization of resources:
 - Maximum number N of messages arriving at the resource: $N = B + R \times \text{TimeInterval}$
 - Maximal Average Rate R' (in bytes per second): $R' = M \times R$
 - Maximal Buffer Size (BS in bytes): $BS = M \times (B+1)$



Example of LBAP

- Consider $M = 1176$ Bytes per message, $R = 75$ messages per second, $B = 10$ messages
- During a time interval of 1 second, the maximum number of messages arriving at a resource must not exceed $N = 10 \text{ messages} + (75 \text{ messages/second} * 1 \text{ second}) = 85$ messages
- Maximum average data rate in bytes per second is $R' = 1176 \text{ bytes} * 10 \text{ messages/second} = 88200$ bytes/second
- Maximum buffer size in bytes in $BS = 1176 \text{ bytes} * (10 \text{ messages} + 1) = 12936$ bytes



Admission Tests

- **Task schedulability tests** for CPUs
 - This is done for delay guarantees
- **Packet schedulability tests** for sharing host interfaces, switches
 - This is done for delay and jitter guarantees
- **Spatial tests** for buffer allocation
 - This is done for delay and reliability guarantees
- **Link bandwidth tests**
 - This is done for throughput guarantees



Resource Reservation and Allocation

- Two types of reservations
 - Pessimistic approach - Worst case reservation of resources
 - Optimistic approach - Average case reservation of resources
- To implement resource reservation we need:
 - Resource table
 - to capture information about managed table (e.g., process management PID table)
 - Reservation table
 - to capture reservation information
 - Reservation function
 - to map QoS to resources and operate over reservation table



Resource Reservation

- Two types of reservation styles:
 - Sender-initiated reservation
 - Receiver-initiated reservation



Conclusion – Current State

- Lack of mechanisms to support QoS guarantees
 - Need research in distributed control, monitoring, adaptation and maintenance of QoS mechanisms
- Lack of overall frameworks
 - Need QoS frameworks for heterogeneous environments (diverse networks, diverse devices, diverse OS)