The Integrated Services Architecture for the Internet

Kevin Jeffay
Department of Computer Science
University of North Carolina at Chapel Hill
jeffay@cs.unc.edu
November 23, 1999

http://www.cs.unc.edu/~jeffay/courses/comp249f99

Real-time data delivery on the Internet Today
What’s the problem?

- Do we need more bandwidth or just better management of the existing bandwidth?

![Diagram showing requirements over time]
Real-time data delivery on the Internet Today
Where do we go from here?

◆ Provide “best-effort” service by adapting media streams
  » Monitor & provide feedback on performance
  » Bias transmission and processing of media to ameliorate the effects of congestion

◆ Provide true quality-of-service through reservation of resources in the network
  » Requires coordination amongst all parties
    ❖ admission control
    ❖ policing
    ❖ ...

Integrated Services architecture for the Internet
Application requirements

◆ The Internet cannot support adaptive real-time multimedia communication today
  » Variable length queuing delays make low latency, continuous playout impossible
  » Packet loss exacerbates the playout problem

◆ This claim appears to be independent of the bandwidth requirements of the communication
An Integrated Services Internet is one that supports:

- **“Flows”**
  - *real-time communication* — contractual service guarantees
  - *best-effort communication* — today’s service model

- **Traffic management**
  - *controlled link sharing* — the ability of a service provider to allocate percentages of a link’s capacity to “classes” of traffic

**Axioms**

- **Resource reservation is required**
  - Network elements must maintain per-flow state information and use this information to ensure application performance contracts are met

- **Admission control is required**
  - To ensure performance contracts are met, network elements must ensure they do not over commit their resources

- **Applications must be policed**
  - To ensure performance contracts are met, network elements must ensure applications do not claim more resources than they contracted for
Integrated Services architecture for the Internet
Service models for flows

◆ Integrated services introduces the concept of a service model
  » A contract between a sender and the network for a particular quality of service

◆ Proposed service models (v1)
  » Guaranteed delay — An application receives a guarantee that all packets will be delivered within a fixed delay bound
  » Predictive — A statistically guaranteed delay bound
  » Best effort — Today’s service model

Integrated Services Service Models

Controlled load

◆ Predictive service has evolved into a controlled load service
  » A service that approximates the service a flow would receive under “unloaded conditions” in the network

◆ In a controlled load service, applications can assume:
  » A (very) high percentage of transmitted packets will be delivered
  » A high percentage of transmitted packets will experience a transit delay not significantly greater than the minimum transit delay experienced by any packet
To receive a service contract an application must specify the service it requires and the traffic it will generate

» Canonical flow specification — the token bucket

IETF traffic specification (TSpec)
- average rate
- token bucket depth
- peak rate
- maximum packet size
- minimum policed unit

Realizing Integrated Services
Reference implementation components

Classifier — Maps all packets into one or more classes that receive the same service

Packet Scheduler — Schedules packets for transmission so that performance contracts are enforced
Realizing Integrated Services
Reference implementation components

- Reservation setup protocol
  » Mechanism by which flow-specific state is created and maintained

- Admission control procedure
  » The decision procedure that is used to determine if a new flow can be accepted or not

Realizing Integrated Services
Reference implementation components

- End systems must support the same logical components
  » A real-time chain is only as strong as its weakest link
Realizing Integrated Services

A simple example

- Consider a set of videoconferencing applications that desire a guaranteed delay service
  » Each specifies a delay bound and a TSpec

- Every host and router in the network “shapes” the traffic using a *leaky bucket*

- Admission control
  » Does the sum of the $\rho_i$ exceed the capacity of the outbound link?

- Packet scheduling
  » FCFS suffices!
The Integrated Services Architecture

Architectural components

- Flow specifications
- Routing
- Resource reservation
- Admission control
- Packet scheduling