CS551
Integrated and Differentiated Services
Bill Cheng

http://merlot.usc.edu/cs551-f12
What’s Next: Integrated Services

- Integrated services
  - resource reservations (Internet: RSVP)
  - guaranteed or probabilistic bandwidth/delay

- Pros:
  - good match for real-time traffic (e.g., VOIP)
  - prefect for VPNs (ISPs can sell "virtual pipes")
  - make the most use out of your bandwidth

- Cons:
  - too much state for backbone routers
  - difficult policy issues between AS’s?
    ⇒ not widely deployed
What’s Next: Differentiated Services

Differentiated services
- assumes some overprovisioning
- very simple service model
  - best-effort and preferred (better-than-best-effort)
  - or in and out (best-effort and less-than-best-effort)

Pros:
- easy to implement and fast (no per-flow state)
- ISPs can charge extra for preferred

Cons:
- no guarantees
CS551
Fundamental Design Issues
[Shenker95a]

Bill Cheng

http://merlot.usc.edu/cs551-f12
Key Ideas

- Do we need to extend the Internet service model (currently best effort)?
  - Reservations, admission control, etc, or
  - Overprovision and keep best effort

- How do we even study this question?

- Simple model, very high level view
  - Asks fundamental questions
  - Helps guide the thinking for a very hard question
Model: Utility and Efficacy

Does the network make users happy?

Define $U(j)$ be the utility delivered to the $j$th user
- $U(j)$ maps the network’s performance to the user’s level of happiness
- For example, higher bandwidth delivered to a video application (up to a point) makes the user happier
- Similarly, lower delay delivered to application makes user happier

Goal of network is to maximize
- ... the sum of all $U(j)$s (the efficacy, denoted by $V$)
More Bandwidth or New Service Model?

- In a best-effort network, can increase bandwidth to increase efficacy.

- Or, for the same bandwidth, introduce new services matched to application needs, and increase efficacy that way.

- Key question: what’s the relative cost of adding bandwidth and adding new services?
  - Shenker: always better to add new services:
    - makes better use of available bandwidth
    - but cost of adding new services hard to estimate.
Other Considerations

Do separate networks for different applications provide higher efficacy?
- No. A single network can always use leftover bandwidth to increase efficacy.

Note: increasing efficacy does not mean increasing everyone’s utility.

Service models must map application requirements
- Otherwise, none of these arguments holds.
Implicit vs. Explicit Service Request

Should applications explicitly request service, or should the network determine service to deliver?

Implicit doable if number of services is small and well known and stable (e.g., port number)
- Need to embed application knowledge inside the network (BAD!)

Explicit supports larger variety of services but incentives needed so all do not request highest service
- Applications must know what they want!
- Pricing, accounting and billing: these are hard

Stable service model needed so apps know what to request
- Major coordination effort (imagine changing IP or Ethernet..)
Admission Control?

- Overload: a network is overloaded if by removing a flow would increase $V$ even though there are fewer flows.

- If $V(n)$ does not continue to increase as $n$ goes to infinity, then we either need admission control or over-provisioning.

- Best Effort never overloads (or does it?)
Utility Curve Shapes

- Elastic
- Delay-adaptive
- Hard real-time

If convex near origin, then need admission control.
Over-provisioning

- Works for "normal users" because need to overprovision by a small amount
- Over-provisioning for "leading edge" users is hard because these consume several orders of magnitude more than normal users
- Internet will be provisioned to rarely block normal users, but will block leading edge users frequently
Summary

- Internet should extend its service model
  - Service should be explicitly requested by applications
  - Service model should incorporate admission control
  - Abstract formulation of maximizing efficacy

- Digital convergency: *Integrated Services*
  - Data network
  - Telephone network
  - Cable network
  ⇒ under one IP