Fundamental Design Issues for the Future Internet

Scott Shenker

The Current Internet

- Best Effort Service
- No admission control (no one is refused the right to use the network)
- Dominated by traditional uses
  - remote terminal (telnet)
  - file transfer (ftp, http)
  - electronic mail
- Elastic applications
  - Tolerate packet delays and losses gracefully
  - Congestion control: decrease transmission rate in presence of congestion

Demands of Real-Time Applications

- Video and voice
- Typically less elastic (tolerant of delay variations)
- Do not perform adequately on current Internet
  - Delay variations too extreme
  - Too many dropped packets
- Do not back off in presence of congestion
  - May interfere with more traditional applications

Change Implementations or Architecture?

- **Option 1: Change router implementation**
- Want fairness between congestion-avoidant and congestion-ignorant applications
- Fair Queueing packet scheduling
- Routers ensure every user has “fair share” of bandwidth
- Problem: service delivered to high-bandwidth applications may still be inadequate

Change Implementation or Architecture?

- **Option 2: Change application implementations**
- Make voice, video applications more adaptive to variations in packet delays
- e.g., scalable compression, nv, vat
- Advantages:
- No changes required to any network interfaces
- So, make changes incrementally

- **Disadvantages of implementation changes**
  - Network delivers same class of service to all
  - No assurances of service quality

Changing Architecture

- **Option 3: Changing Internet Architecture**
- Extend Internet serivce model
- Include a variety of service classes
- May include admission control
- What are the issues?.....

Goals of Network Design

- Network performance: Degree to which user application's service requirements are satisfied
- U = utility functions
- Describes the mapping between service and application (how application performance depends on delivered service)
- V = efficacy or total utilization
- V = sum of U_i's

- **Matching service to application increases overall utility**
- Extra cost to provide mechanisms for service classes

- **Alternative to multiple service classes: separate networks for different applications, each with its own service class**
- e.g., separate telephony, video and data networks (as we have now)
- Inefficient: combined network allows statistical sharing of resources

- **But in a mixed network, note:**
- Performance-sensitive clients gain from better service model
- Less-performance-sensitive clients LOSE
  - Their performance is worse

Deciding Which Service To Give

- **Implicitly Supplied:**
  - Network chooses service class
  - Application doesn't specify service requirements
  - Requires no changes to network interface
  - No guarantees made that must be fulfilled
  - **Practical disadvantages:**
  - Fixed set of applciation classes
  - Routers have to know about applications
  - Can't service correctly those they don't know
  - Embedding application info into network layer violates design tenets of Internet
  - IP successful because separate from application and physical implementation

Deciding Which Service To Give

- **Applications explicitly request desired service**
  - Network offers set of service classes
  - Applications pick one
  - **Disadvantages:**
  - Incentives
  - Lack of flexibility of service model

Incentives

- Users explicitly ask for service level
- **Why not ask for highest bandwidth?**
- **Must be some incentive to ask for less (COST!)**
- Threatens whole culture of Internet, where bandwidth is considered "free"
- Usage-based pricing will change design and culture
- Mechanisms for authentication, accounting, billing
- **Regardless of whether we add service classes, will need to address incentives**
- Junk mail, automatic agents: must have some incentive not to use too much bandwidth
- Usage pricing for high-quality services only?

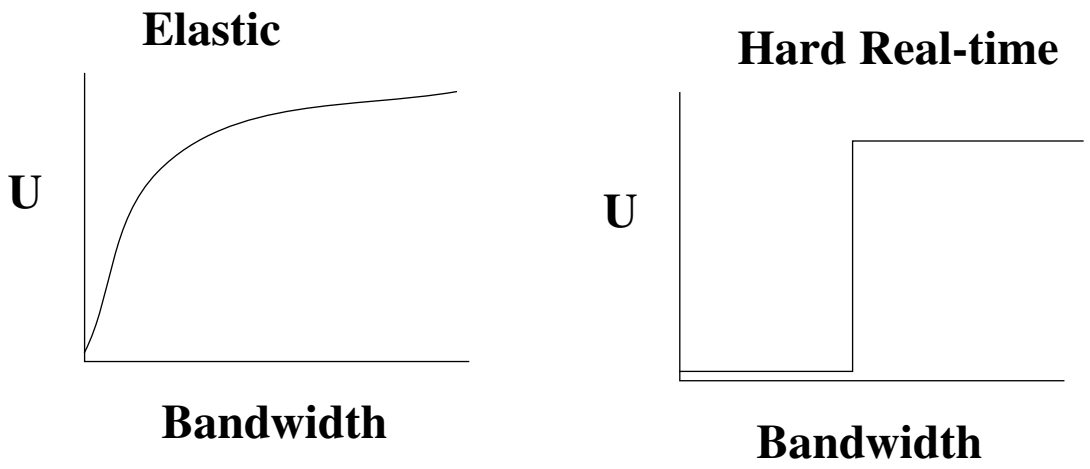
Lack of Flexibility of Service Model

- Network service offerings must be known to appl.
- **Requires that service model be stable (but extensible) over long period**
  - **Or applications break**
- Advantage: useful abstraction of network service
- Disadvantage: if initial service offerings are not well-designed, will be very painful to change them
- Uniformity required in explicit approach: all routers must understand the same levels of service
- Traditionally, Internet could include any subnet that supported IP
- Allow subnets to support subset of service model

Do We Need Admission Control?

- **Best-effort service with priority levels: make no guarantees about service**
  - Allow all traffic to be admitted
- **Bounded delay services: provide quantitative characterizations**
  - Require explicit resource reservation and admission control
- **Admission control: network turns away additional flows when admitting them would lead to violating existing guarantees**
- Overloading: dropping a flow increases overall utility

Utility of Applications



Utility of Applications

- **Traditional applications: elastic, tolerant of delays**
  - Use congestion control to adjust to network traffic
  - No role for admission control
- **Hard real-time requirements: not tolerant**
  - Don't care if packets arrive early
  - Application performs badly if packets arrive later than specified delay bound
  - Function better with admission control
- **Most audio, video can be tolerant of occassional delay-bound violations and dropped packets**
  - Intrinsic bandwidth requirements due to data generation rate (e.g., commpression rate)
- **Rate-adaptive: adjust in response to congestion**

Questions

- For traditional elastic applications,
  - Efficacy is maximized by accepting all flows
- Real-time applications (hard real-time or delay-adaptive):
  - Efficacy is maximized when some flows are turned away

- **Admission control?**
- **Or overprovision network so that overloading rarely occurs?**

Overprovisioning

- Is it cost-effective to overprovision the entire network?
- **Shortl-term: No hope of overprovisioning everywhere:**
  - Only pockets of high-bandwidth users will want (be willing to pay for) high bandwidth
  - Limited to fast LANs
- **Long-term: mature networking industry**
  - US phone network successfully overprovisioned: seldom get call blocking
  - Future internet load will be more variable than phone network

Because of high variance, overprovisioning not cost-effective for Internet

- **Normal usage: less than 1 Mbit/sec**
  - (telephony, low-quality video, etc.)
  - Cost-effective to overprovision for normal usage
- **Leading-edge usage: extremely high bandwidth**
  - Overprovisioning would require large percentage increase in capacity of links
  - Low average utilization
  - Very expensive--cost borne by leading-edge users
  - Prediction: unwiling to pay for overprovisioning
  - Will tolerate call blockage to avoid paying

Summary

- Argues for extending service model: increases network efficacy
- Service should be explicitly requested by application, not supplied implicitly by the network
- Argument for admission control for high-bandwidth flows
- Against overprovisioning: not cost-effective