Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science

6.829 Fall 2001 **L13: Debating the Future Internet Service Model** October 29 2001

Overview. Best-effort vs. reservations. Admission control.
Readings: Shenker95.

1 Best-effort vs. multiple service classes

1.1 Motivation

- The Internet today offers a single class of "best-effort" service. No assurance of when or if packets will be delivered, no admission control.
- But the Internet isn't just ftp, email, and telnet any more.
- The above applications are *elastic* applications. Q: what does this term mean? (A: they can derive additional "utility" from every marginal increase in bandwidth starting from (close to) 0.)
- Starting to see more and more real-time streaming applications such as voice and video.
- Claim: these apps are not as elastic or adaptive. Delay variations and losses pose severe problems. And they don't *typically* react to congestion.

  This is a questionable assumption, but seems to be true to some extent for real-time Internet telephony.
- **Question: Should we change the underlying Internet architecture to better accomodate these applications?**
- I.e., should the fundamental best-effort Internet architecture be changed to accomodate multiple service classes? And should the architecture be changed to perform *admission control*?
- Goal of paper: To provide a concrete framework in which this debate can rage on.
- At the highest level, the "right" answers depend on the nature of future network applications, the cost of bandwidth, the cost of more complex mechanisms, etc. Hard to quantify! In particular, if all applications can become rate-adaptive, the conclusions of this paper will be questionable.
- Basic idea: multiple service classes and scheduling schemes in the network routers to allocate bandwidth.

1.2 What should network architects optimize?

- Throughput? Utilization? Packet drops? Delays?
- Claim: No! Really want to optimize "user satisfaction," which would in an efficient system be tied to revenue. More precisely, if s_i is the network service (in terms of throughput, drops, etc.) delivered to user i, there is a *utility function* $U_i(s_i)$ which measures the performance of the application and the degree of user satisfaction. The goal of network design is to maximize $V = \sum_i U(s_i)$.
- But what about other optimization criteria? For example, a service provider might choose to optimize revenue.
- Anyway, this formulation makes it apparent that it's worth rethinking the basic service model the architecture provides. Simple examples show that pure FIFO without favor toward particular flows doesn't always optimize V.
- But should we add multiple service classes, or just add more bandwidth?
- Note: while the goal is to optimize V, it won't necessarily optimize each of the $U_i's$. I.e., the nature of this optimization is to take bandwidth away from the elastic applications and satisfy the more sensitive, inelastic applications.

1.3 How is service for a flow chosen?

- Option #1: Implicitly supplied
  - Here, network automatically chooses service class based on packet/flow.
  - Advantage: No changes to end-hosts or applications.
  - Disadvantage: New applications cause problems, since routers don't know what to do about them.
  - In general, this approach embeds application information in the network layer, which has deficiencies.
  - Also suffers from stability problems, because of a changing service model, changing unknown to applications.
  - Uses multi-field classification capability to assign service to packet flow.
- Option #2: Explicitly requested
  - Here, applications explicitly ask for particular levels of service.
  - Problem: incentives. Why will some applications volunteer to ask for lower levels of service?
  - Possible solution: pricing.
  - Social implications of usage-based pricing. This will discourage the browsing mentality of users (which information providers find attractive).
  - Key point: even in a single best-effort class, the notion of incentives is important. Addresses the issue of "whether to send data."

  - Usage-based pricing at a higher granularity than an individual user might be more appropriate.
  - In the explicit model, the network service model is known to the application.

- Link-sharing as an implicit model. Works well when dealing with aggregates of flows.

1.4 Do we need admission control?

- Some services might need explicit resource reservation, so the network may need to turn away flows, which if admitted, will violate its current quantitative commitments.

- One approach to answering this question: If there are values of population size n and n' such that $V(n) > V(n')$ for $n < n'$, then $V()$ has a local maximum, and it may be better to turn flows away.

- Question amounts to monotonicity of $V = \sum_i U(s_i)$.

- What is the shape of $U(s_i)$ for various classes of applications?

  1. Elastic applications: concave. For this, $V()$ is monotonically increasing with n.
  2. Hard real-time apps: step function. Clearly admission control is important here (e.g., telephone networks).
  3. Delay-adaptive real-time apps: convex first then concave after inflection point.
  4. Rate-adaptive real-time apps: earlier inflection point than above case.

- If U is convex in some region, then $V()$ has a local maximum. This argues for admission control.

- Q: What does this argument miss? Dissatisfaction derived from denied flows! Also, it assumes that in general it is better to turn a flow away that terminate an existing flow. This may not be a good assumption in some cases.

- Overprovisioning as an alternative: how much you have to overprovision depends on variance of bandwidth use. Furthermore, while backbone bandwidths will continue to improve dramatically with time, it's unclear that the same trend is true for last-mile access links.

- Need to overprovision even if you have admission control!

- Shenker concludes that admission control is more cost-effective than the amount of overprovisioning you otherwise have to do.

- Bottom line: Paper sets wonderful framework for this debate, which is still very open. Next week, we will look at integrated and differentiated services.