

CS640: Introduction to Computer Networks

Aditya Akella

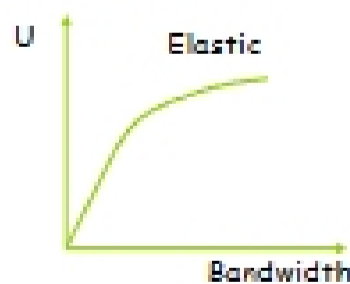
Lecture 20 -
QoS

Why a New Service Model?

- Best effort clearly insufficient
 - Some applications need more assurances from the network
- What is the basic objective of network design?
 - Maximize total bandwidth? Minimize latency?
 - **Maximize user satisfaction** - the total utility given to users
- What does utility vs. bandwidth look like?
 - Must be non-decreasing function
 - Shape depends on application

2

Utility curve - Elastic traffic



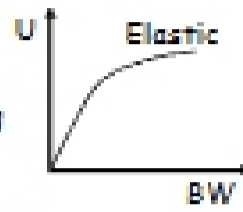
Does equal allocation of bandwidth maximize total utility?

2

Admission Control

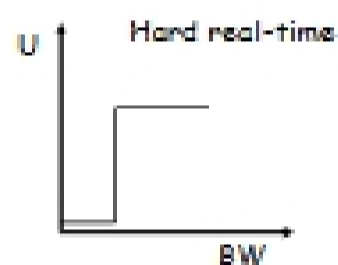
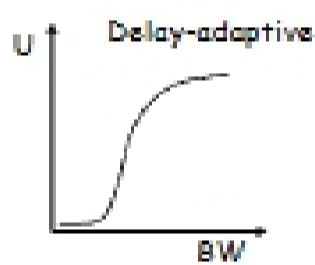
- If $U(\text{bandwidth})$ is concave
→ elastic applications

- Incremental utility is decreasing with increasing bandwidth
- Is always advantageous to have more flows with lower bandwidth
 - No need of admission control and explicit QoS mechanisms



4

Utility Curves - Inelastic traffic



Does equal allocation of bandwidth maximize total utility?

5

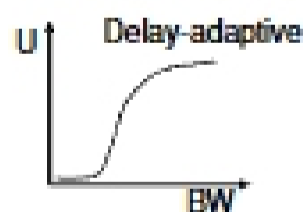
QoS and Admission Control

- If U is convex → inelastic applications

- $U(\text{number of flows})$ is no longer monotonically increasing

- Need admission control and special QoS mechanisms

- Admission control → deciding when the addition of new people would result in reduction of utility



6

QoS Instantiation #1: Integrated Services

Key components:

1. Type of commitment
What does the network promise?
2. Packet scheduling
How does the network meet promises?
3. Service interface
How does the application describe what it wants?

7

Type of Commitments

- **Guaranteed service**
 - For **hard real-time** applications
 - Fixed guarantee, network meets commitment as long as rates clients send at match traffic agreement
- **Predicted service**
 - For tolerant (e.g. delay-adaptive) applications
 - Two components
 - If conditions do not change, commit to current service
 - If conditions change, take steps to deliver consistent performance (help apps minimize playback delay). Ensure that such apps continue to see a lightly loaded network.
- **Datagram/best effort service**

8

Scheduling for Guaranteed Traffic

- Use token bucket filter to characterize traffic
 - Described by rate r and bucket depth b
 - FlowSpec or flow specification
- Use **Weighted Fair-Queueing** at the routers
- Parekh's bound for worst case queuing delay = b/r

9
