CS 414 – Multimedia Systems Design Lecture 16 – Multimedia Transport Subsystem (Part 3)

Klara Nahrstedt

Spring 2009

Administrative

- **MP2:** deadline Monday, March 2, demos 5-7pm (sign up in class on Monday)
- **Multimedia related talk**: Thursday, February 26, 4pm 1109 SC, “VoIP over 802.11” by Henning Schulzrinne
- **HW1:** posted on Friday, February 27 and due Friday, March 6
- **Midterm review session**: Friday, March 6, in class
- **Midterm**: **Monday March 9, in class**
- **Class canceled** on Friday, March 13 due to EOH

Outline

- Transmission Phase
  - Traffic Shaping
    - Isochronous Traffic Shaping
    - Shaping Bursty Traffic
  - Rate Control
  - Error Control

QoS Enforcement – Traffic Shaping

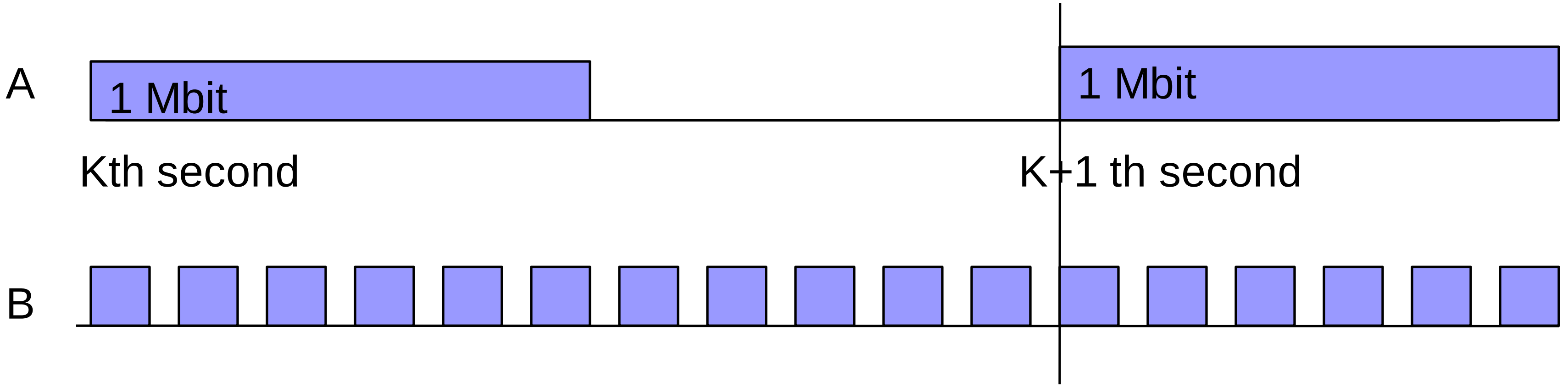
- In Packet Network, admission control, reservation is not sufficient to provide QoS guarantees
- Need **traffic shaping** at the entry to network and within network
- Traffic shaping
  - Decides how packets will be sent into the network , hence regulates traffic
  - Decides whether to accept a flow's data
  - Polices flows

Purpose of Traffic Shaping

- **Traffic shape**
  - A way of a flow to describe its traffic to the network
- Based on traffic shape, network manager (s) can determine if flow should be admitted into the network
- Given traffic shape, network manager(s) can periodically monitor flow's traffic

Example

- If we want to transmit data of 100 Mbps,
  - Traffic Shape A: Do we take 1 packet size of size 100 Mbit and send it once a second, or
  - Traffic Shape B: Do we take 1 packet of size 1 Kbit and send it every 10 microseconds?

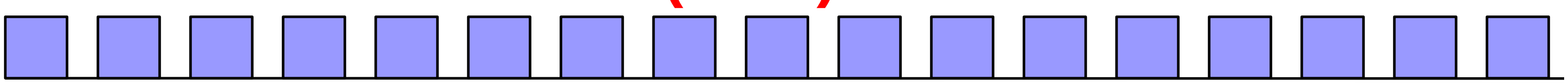


Flow's Traffic Shape Parameters (Network QoS)

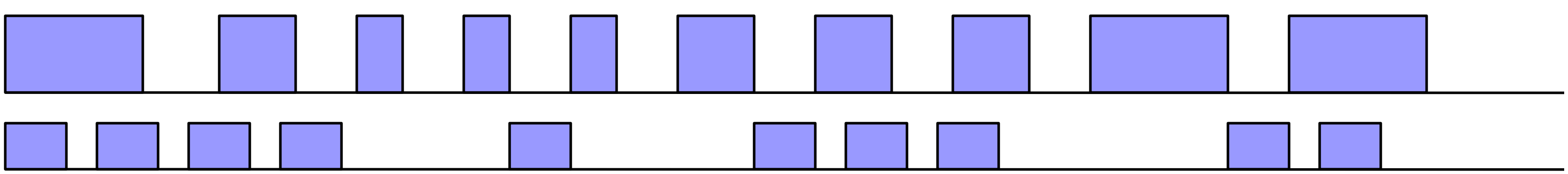
- Traffic Envelope
  - Peak rate
  - Average rate
  - Burst length
  - Burst duration
- Service Envelope
  - Maximum tolerable delay
  - Desired delay jitter
  - others

Source Classification

- Classification of sources :
  - Data – bursty, weakly periodic, strongly regular
  - Audio – continuous, strong periodic, strong regular
  - Video – continuous, bursty due to compression, strong periodic, weakly regular
- Classification of sources into two classes:
  - **Constant Bit Rate (CBR)** – audio



  - **Variable Bit rate (VBR)** – video, data

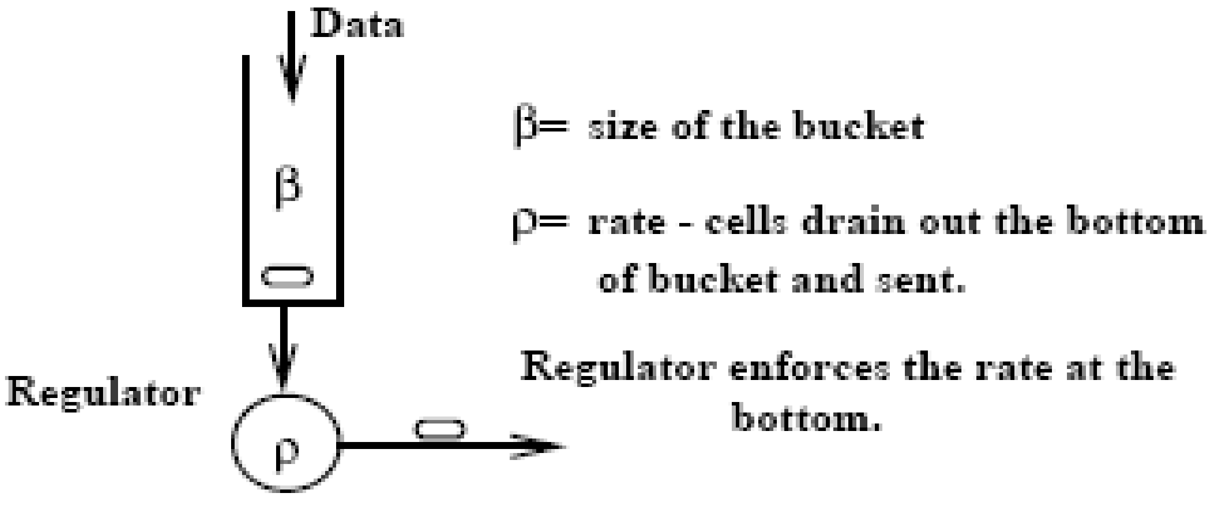


Bandwidth Allocation

- CBR traffic (shape defined by peak rate)
  - CBR source needs peak rate allocation of bandwidth for congestion-free transmission
- VBR traffic (shape defined by average and peak rate)
  - average rate can be small fraction of peak rate
    - underutilization of resources can occur if pessimistic allocation (peak rate allocation) is applied
    - Losses can occur if optimistic allocation (average rate allocation) is applied

Isochronous Traffic Shaping (Simple Leaky Bucket Traffic Shaper)

- Developed by Jon Turner, 1986 (Washington University, St. Louis)



Each flow has its own leaky bucket.

Example

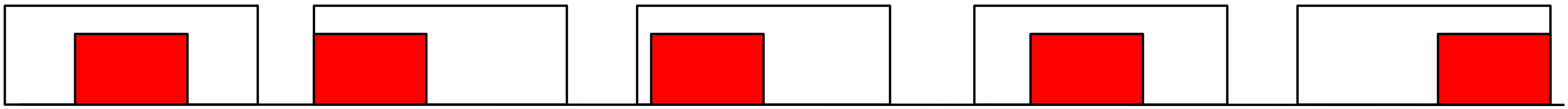
- Consider for audio flow, size of the bucket
  - β = 16 Kbytes
  - Packet size = 1 Kbytes (one can accumulate burst up to 16 packets in the bucket)
  - Regulator's rate ρ = 8 packets per second, or 8KBps or 64Kbps
- Consider video flow, size of bucket
  - β = 400 Kbytes
  - Packet size = 40 Kbytes (burst of 10 packets)
  - Regulator's rate ρ = 5 packets per second, 200 KBps, 1600Kbps

Isochronous Traffic Shaping (r,T)-smooth Traffic Shaper

- Developed by Golestani, 1990
- Part of stop-and-go queuing/scheduling algorithm
- Traffic divided into T-bits frames, where T is fixed
- r-bits packet size per flow is considered, where r varies on a per flow basis

(r,T) Traffic Shaper

T-bits frames, sent every T-bit times



r-bits packets

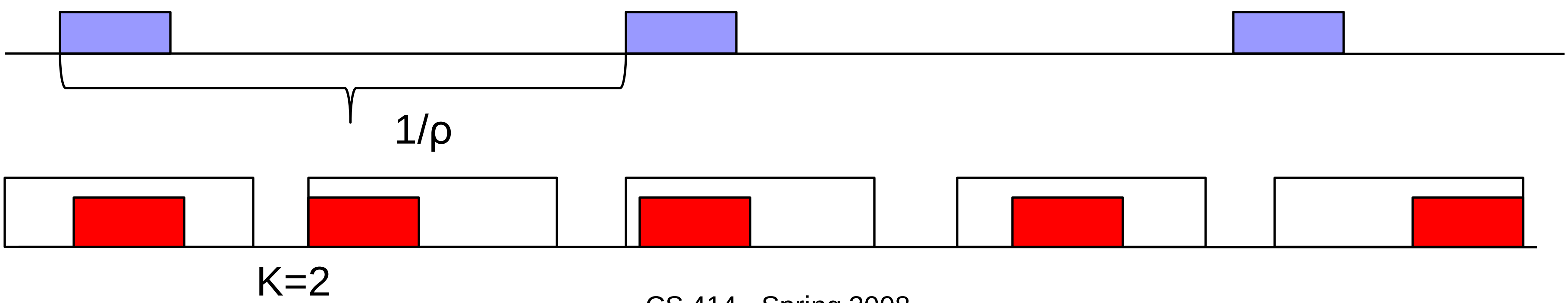
Time line

$r \leq T$

- Flow is permitted to inject no more than r bits of data into the network frame in any T bit times
- if the sender wants to send more than one packet of r-bits, it must wait for next T-bit frame.
- A flow that obeys this rule has (r,T)-smooth traffic shape.

Comparison

- It is relaxed from the simple leaky bucket traffic shaper because
  - Rather than sending one packet of size c every 1/ρ time units, (in simple leaky bucket)
  - The flow can *send c*k* bits every *1/ρ* time units , where *k* is T-bits times within the period *1/ρ*



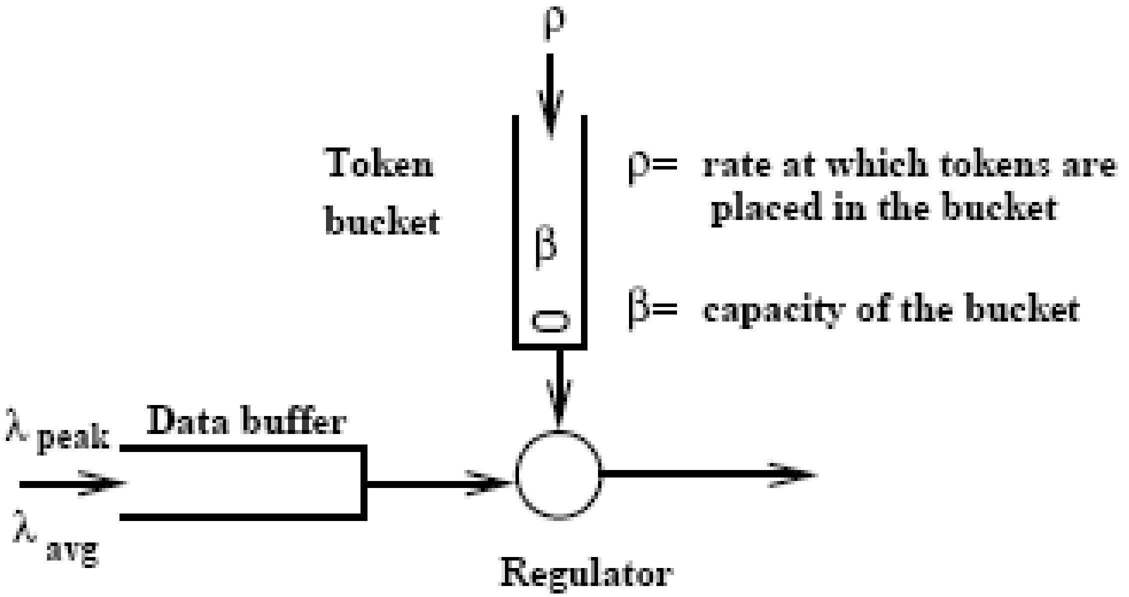
Limitations of Isochronous Traffic Shaping

- In case of (r,T)-smooth traffic shaping, one cannot send packet larger than r bits long, i.e., unless T is very long, the maximum packet size may be very small.
- The range of behaviors is limited to fixed rate flows
- Variable flows must request data rate equal to peak rate which is wasteful

Isochronous Traffic Shaping with Priorities

- Idea: if a flow exceeds its rate, excess packets are given lower priority
- If network is heavily loaded, packets will be preferentially dropped
- Decision place to assign priority
  - At the sender
    - Application marks its own packets since application knows best which packets are less important
  - In the network (policing)
    - Network marks overflow packets with lower priority

Shaping Bursty Traffic Patterns (Token Bucket)



$\lambda_{peak} > \rho > \lambda_{avg} \Rightarrow$

stability and bandwidth utilization

Token Bucket

- The effect of TB is different than Leaky Bucket (LB)
- Consider sending packet of size *b* tokens *(b<β)*:
  - Token bucket is full – packet is sent and b tokens are removed from bucket
  - Token bucket is empty – packet must wait until b tokens drip into bucket, at which time it is sent
  - Bucket is partially full – let's consider B tokens in bucket;
    - if b ≤ B then packet is sent immediately,
    - Else wait for remaining b-B tokens before being sent.

Comparison between TB and LB

Token Bucket	Simple Leaky Bucket
TB permits burstiness, but bounds it	LB forces bursty traffic to smooth out
Burstiness is bounded as follows: - Flow never sends more than $\beta+\tau*\rho$ tokens worth of data in interval τ and - Long-term transmission rate will not exceed ρ	Flow never sends faster than ρ worth of packets per second
TB does not have discard or priority policy	LB has priority policy
TB more flexible	LB is rigid
TB is easy to implement -Each flow needs counter to count tokens, - each flow needs timer to determine when to add new tokens to the counter	LB is easy to implement

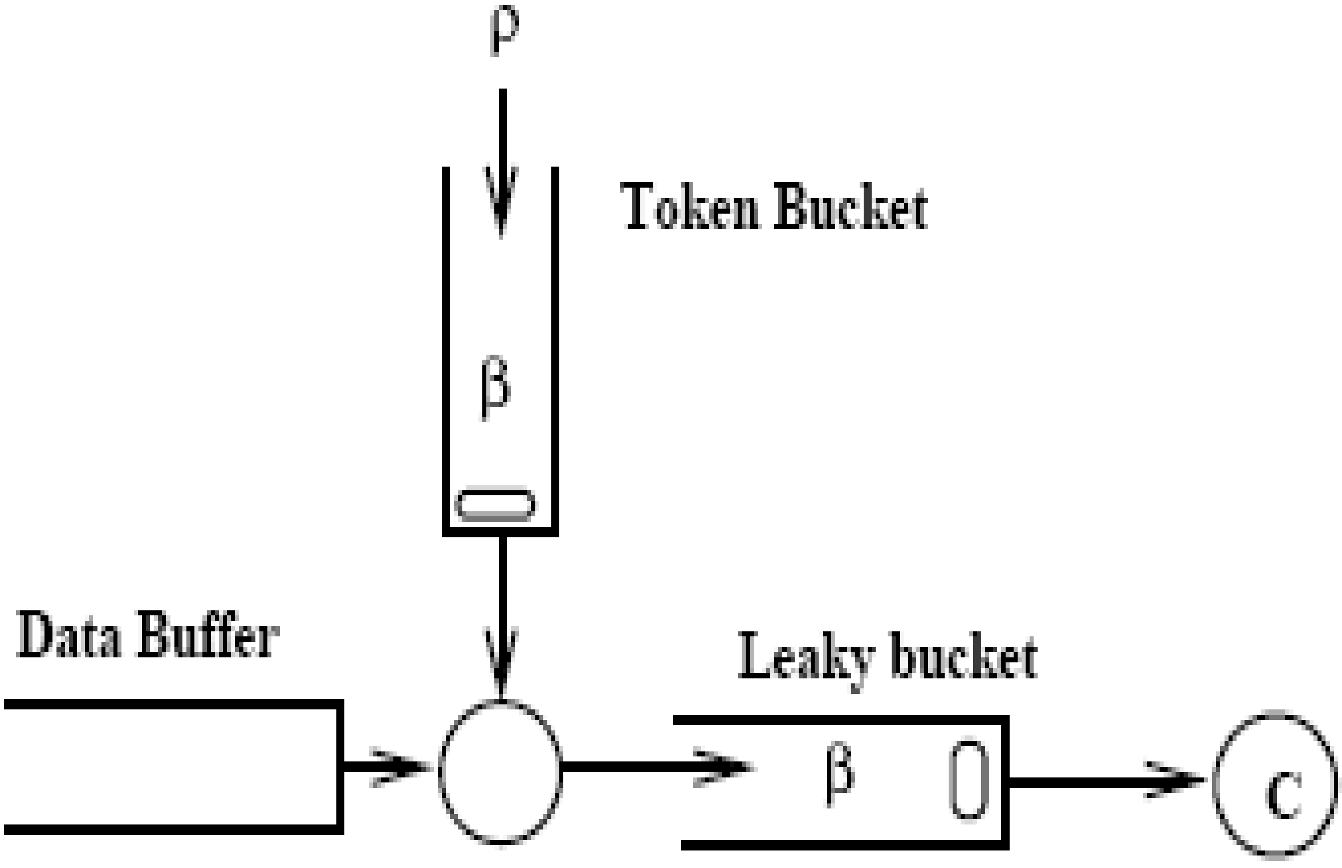
Token Bucket Limitation

- Difficulty with policing
  - At any time the flow is allowed to exceed rate by number of tokens
- Reasoning
  - At any period of time, flow is allowed to exceed rate ρ by β tokens
  - If network tries to police flows by simply measuring their traffic over intervals of length τ, flow can cheat by sending $\beta+\tau*\rho$ tokens of data in every interval.
  - Flow can send data equal to $2\beta+2\tau*\rho$ tokens in the interval 2τ and it is supposed to send at most $\beta+2\tau*\rho$ tokens worth of data

Token Bucket with Leaky Bucket Rate Control

- TB allows for long bursts and if the bursts are of high-priority traffic, they may interfere with other high-priority traffic
- Need to limit how long a token bucket sender can monopolize network

Composite Shaper



Composite Shaper

- Combination of token bucket with leaky bucket
- Good policing
  - But remains hard, although confirming that the flow's data rate does not exceed C is easy
- More complexity for implementation
  - Each flow requires two counters and two timers (one timer and one counter for each bucket)

Performance Guarantees

- Every traffic management needs QUEUE MANAGEMENT (QM)
- Statistical versus Deterministic Guarantees
- Conservation of Work
  - QM schemes differentiate if they are work conserving or not
  - **Work conserving system** – sends packet once the server has completed service (examples – FIFO, LIFO)
  - **Non-work conserving scheme** – server waits random amount of time before serving the next packet in queue, even if packets are waiting in the queue

Conclusion

- Traffic Shaping happens at the entry to the network
- It is a very important function to regular and police traffic at the edges to avoid huge bursts coming into the network