Traffic theory for the Internet and the future Internet

Orange Labs

Jim Roberts, Research & Development
29 August 2008, MAS Seminar

orange™

Internet traffic theory

- understanding the relationship between demand, capacity and performance
- sizing for performance objectives
  - what traffic characteristics are important?
- designing efficient traffic controls
  - to meet diverse QoS requirements

demand
– volume
– characteristics

capacity
– bandwidth
– how it is shared

performance
– response time
– latency

an example:
- Erlang's formula

$$B = \frac{A^N / N!}{\sum_{0 \le i \le N} A^i / i!}$$

B is blocking probability when *N* trunks are offered demand *A*

The Internet and the future Internet

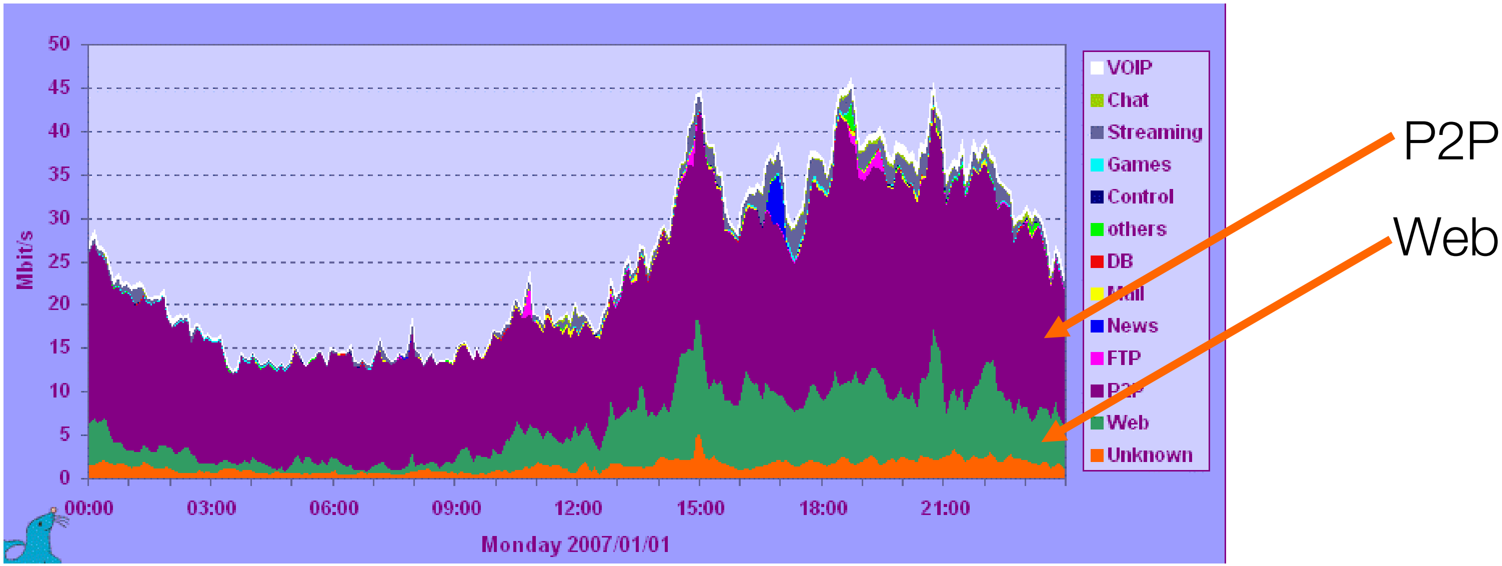
- the Internet, a victim of its success
  - all services are converging to IP, the Internet is indispensable
  - but IP was never designed for this and deficiencies are increasingly apparent: security, mobility, QoS,...
- some advocate a clean slate design?
  - GENI/FIND in the US, projects in Asia
  - FP7 programme on Network of the future: 4WARD, PSIRP, ...
- so, if we can start from scratch, how should the network be designed to meet QoS requirements?
  - accounting for the lessons of traffic theory
  - [and the realities of the Internet business environment,...]

outline

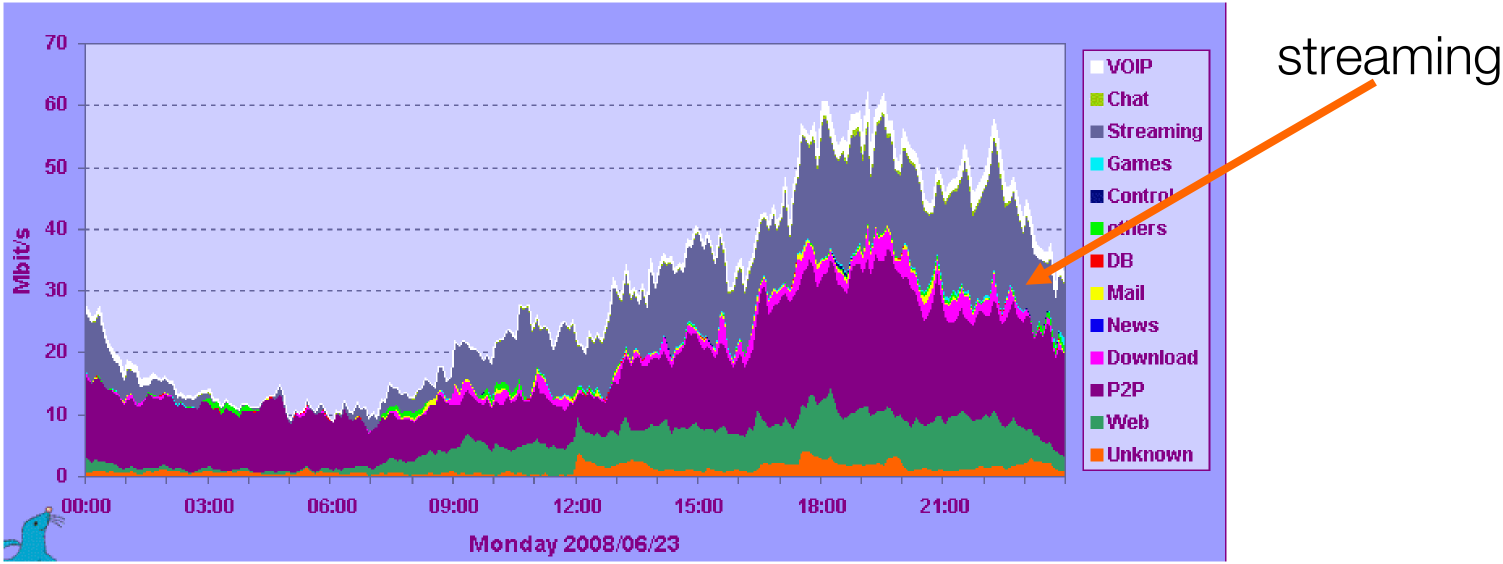
- nature of Internet traffic
- performance of statistical multiplexing
- performance of statistical bandwidth sharing
- service differentiation
- multi-path routing

Composition of Internet traffic

Jan 2007



June 2008

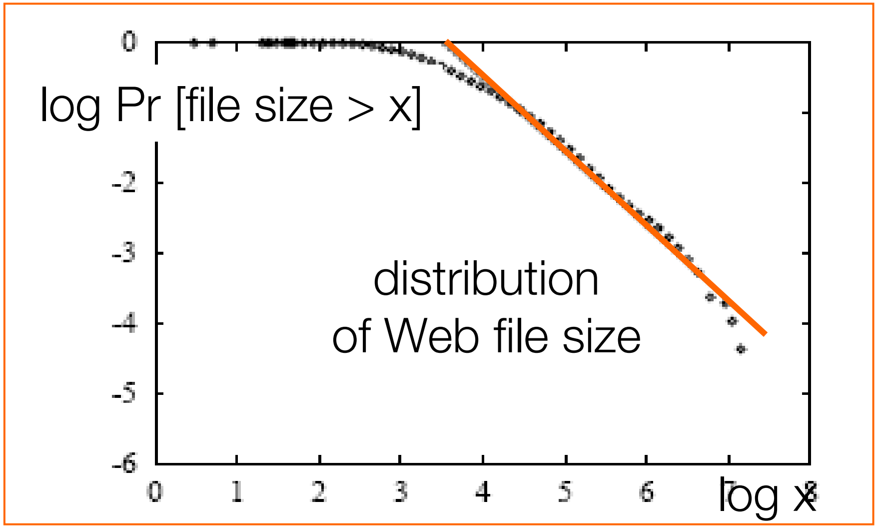


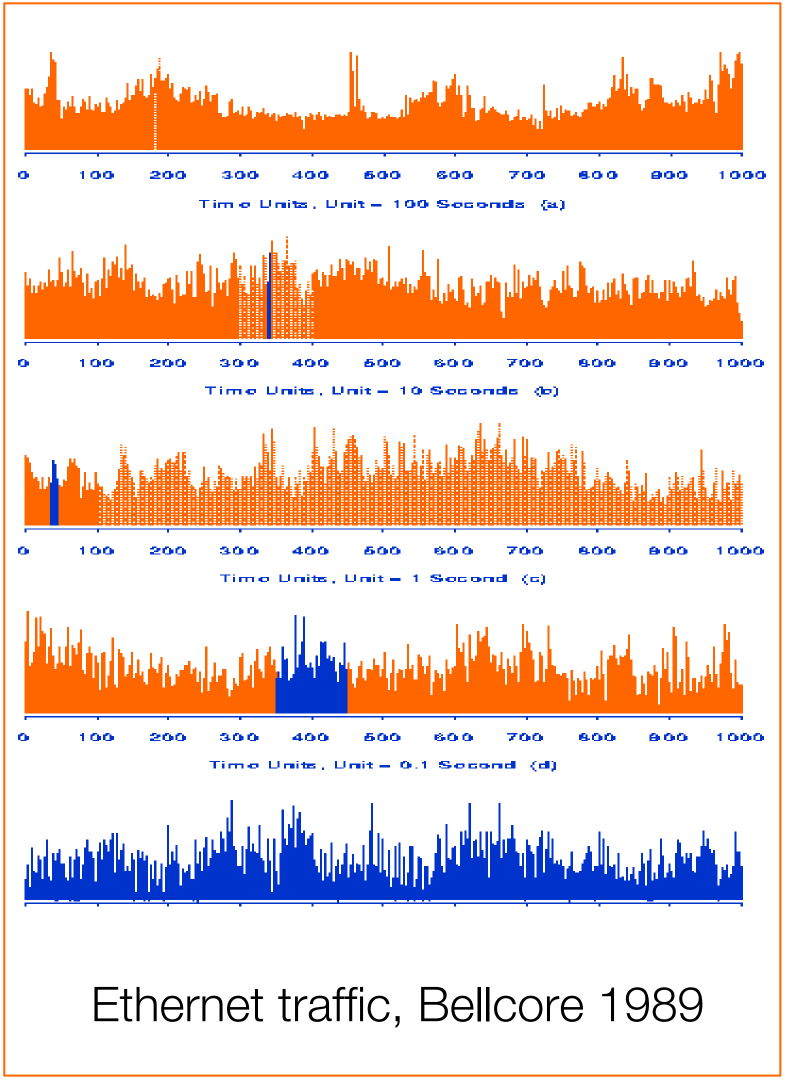
What traffic in the future Internet?

- more video? less P2P? ... new unimagined applications!
- but we can still distinguish two broad types of traffic:
  - open-loop controlled streaming traffic
    - audio and video, real time and playback
    - rate and duration are intrinsic characteristics
    - QoS $\Rightarrow$ negligible loss and delay
  - closed-loop controlled elastic traffic
    - digital documents (movies, Web pages, files, ...)
    - rate and duration are measures of performance
    - QoS $\Rightarrow$ adequate response time
- without forgetting adaptive rate coding, progressive download,...

Internet traffic is self-similar

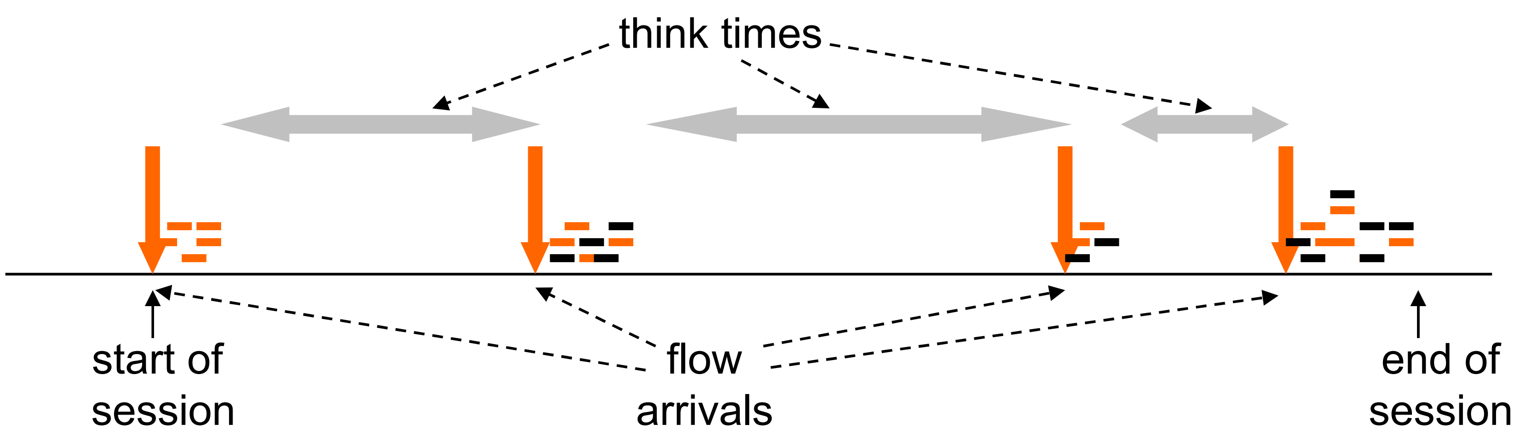
- it is well established that the packet arrival process is self-similar (and even multi-fractal)
- plausible explanations have been provided:
  - heavy-tailed flow size distribution
  - ... and TCP induced burstiness
- but *session* arrivals are Poisson





A session traffic model

- observed at some point in the network, eg, access, core link
- a session consists of a succession of flows separated by "think times"
  - flow characterics: size, peak rate, number of TCPs,...
  - think times begin at the end of each flow
  - sessions are mutually independent
- sessions occur as a homogeneous Poisson process
  - an Internet "invariant": [Floyd and Paxson, 2001]



outline

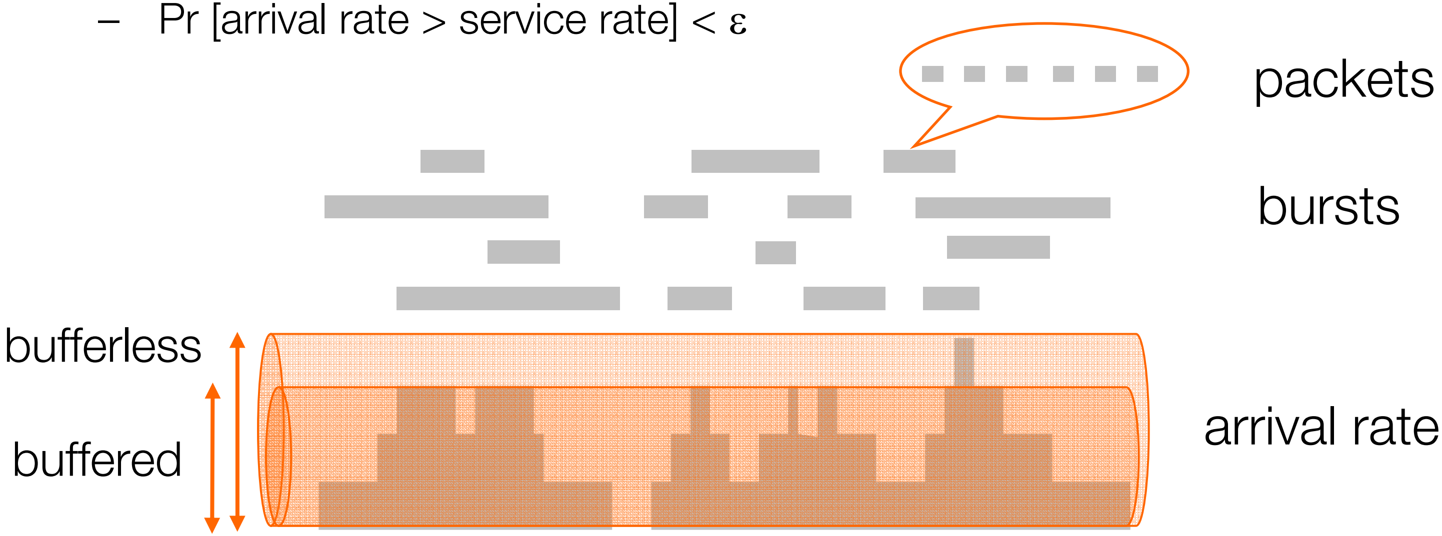
- nature of Internet traffic
- performance of statistical multiplexing
- performance of statistical bandwidth sharing
- service differentiation
- multi-path routing

Traffic theory for statistical multiplexing

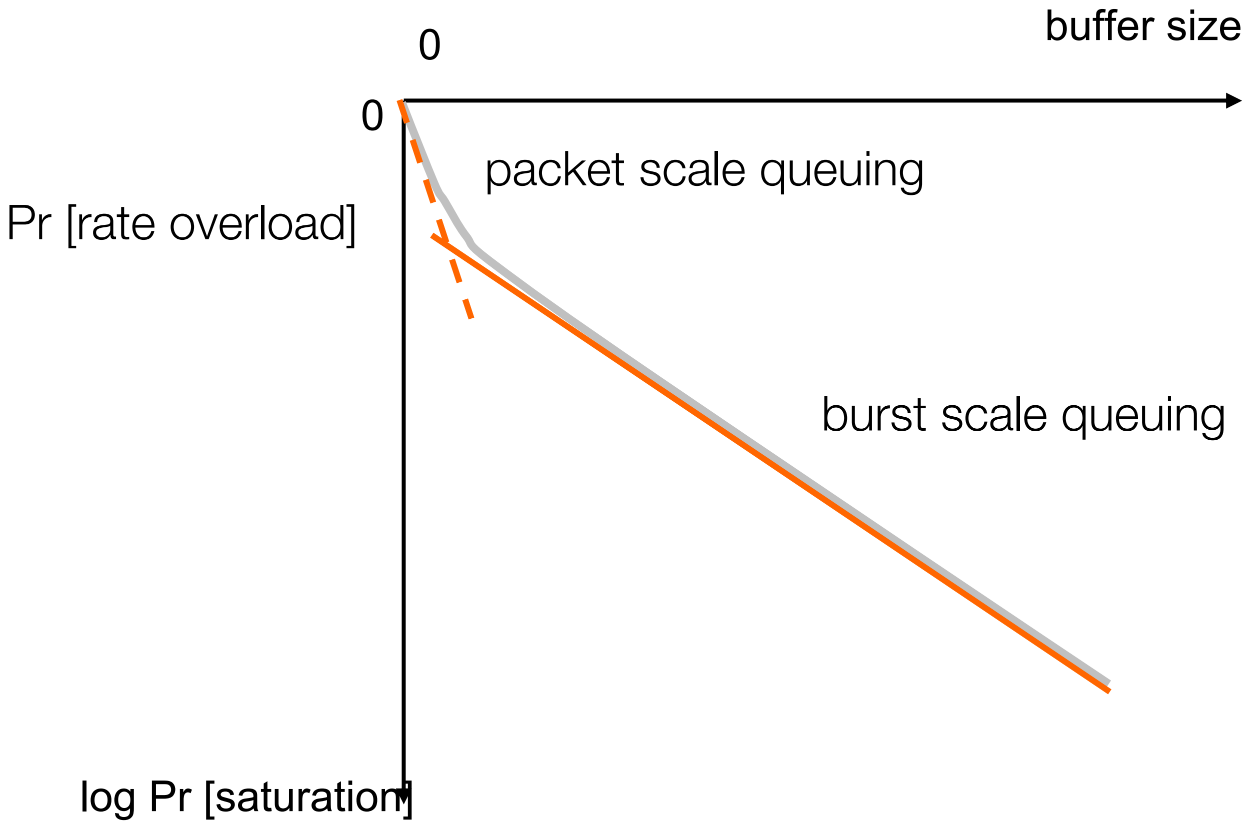
- assume intrinsic traffic characteristics
  - flows are not rate adaptable
  - typical of conversational and streaming audio/video
- seek to understand performance
  - demand – capacity – performance
  - for link and buffer sizing and designing traffic controls
  - at flow, burst and packet time scales

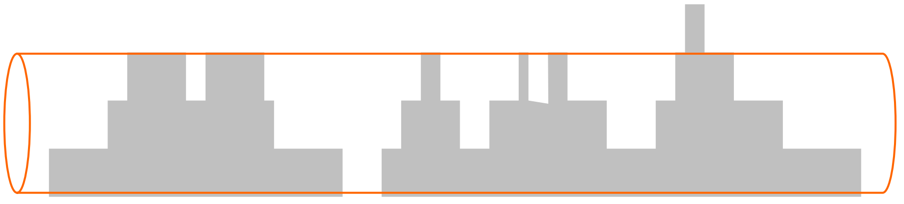
Buffered and bufferless multiplexing

- consider a superposition of on-off flows and distinguish buffered and bufferless multiplexing
  - performance models for sizing and admission control
- buffered multiplexing
  - Pr [delay > T] < ε'
- bufferless multiplexing
  - Pr [arrival rate > service rate] < ε

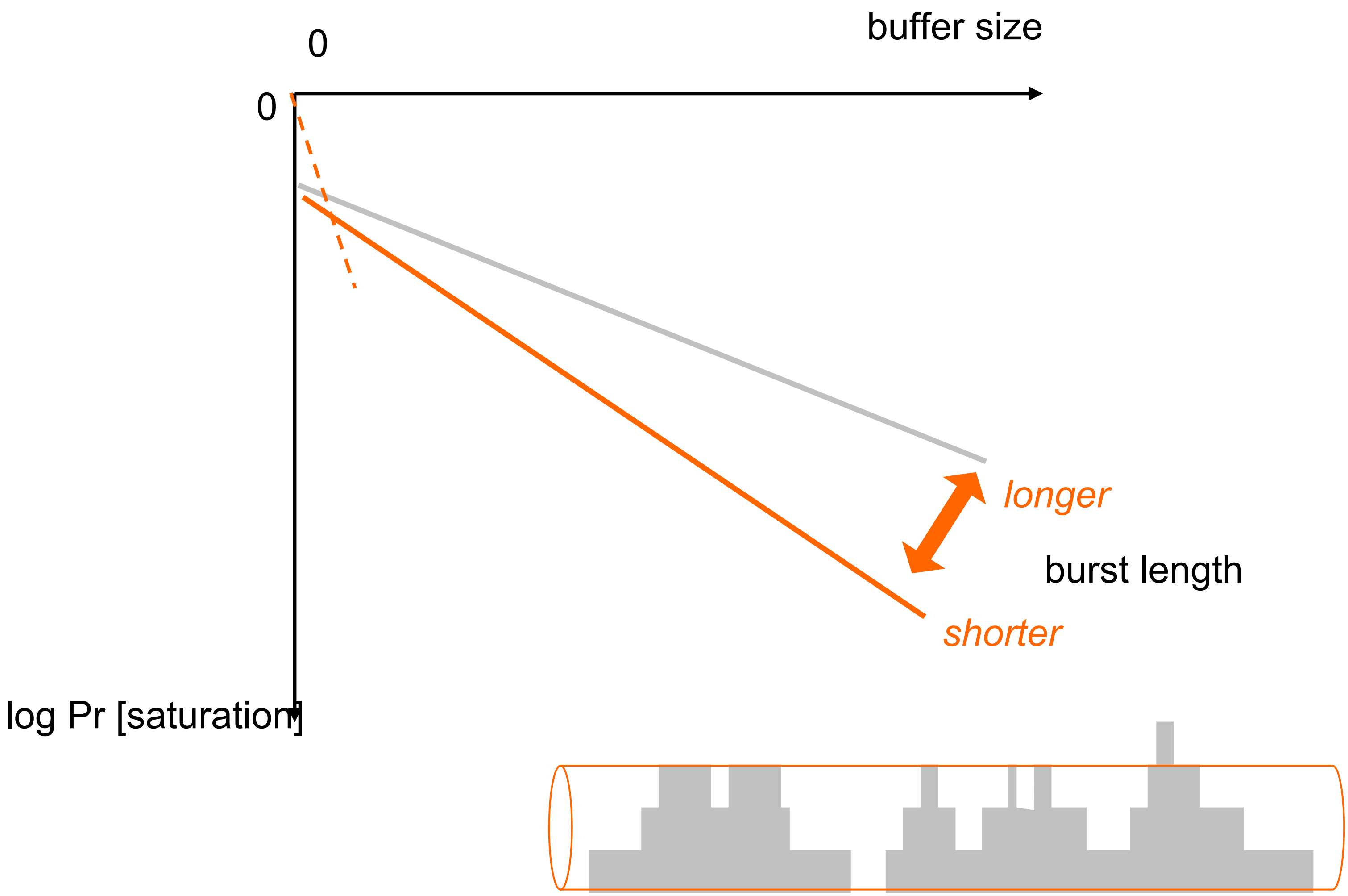


Statistical multiplexing performance: impact of traffic characteristics

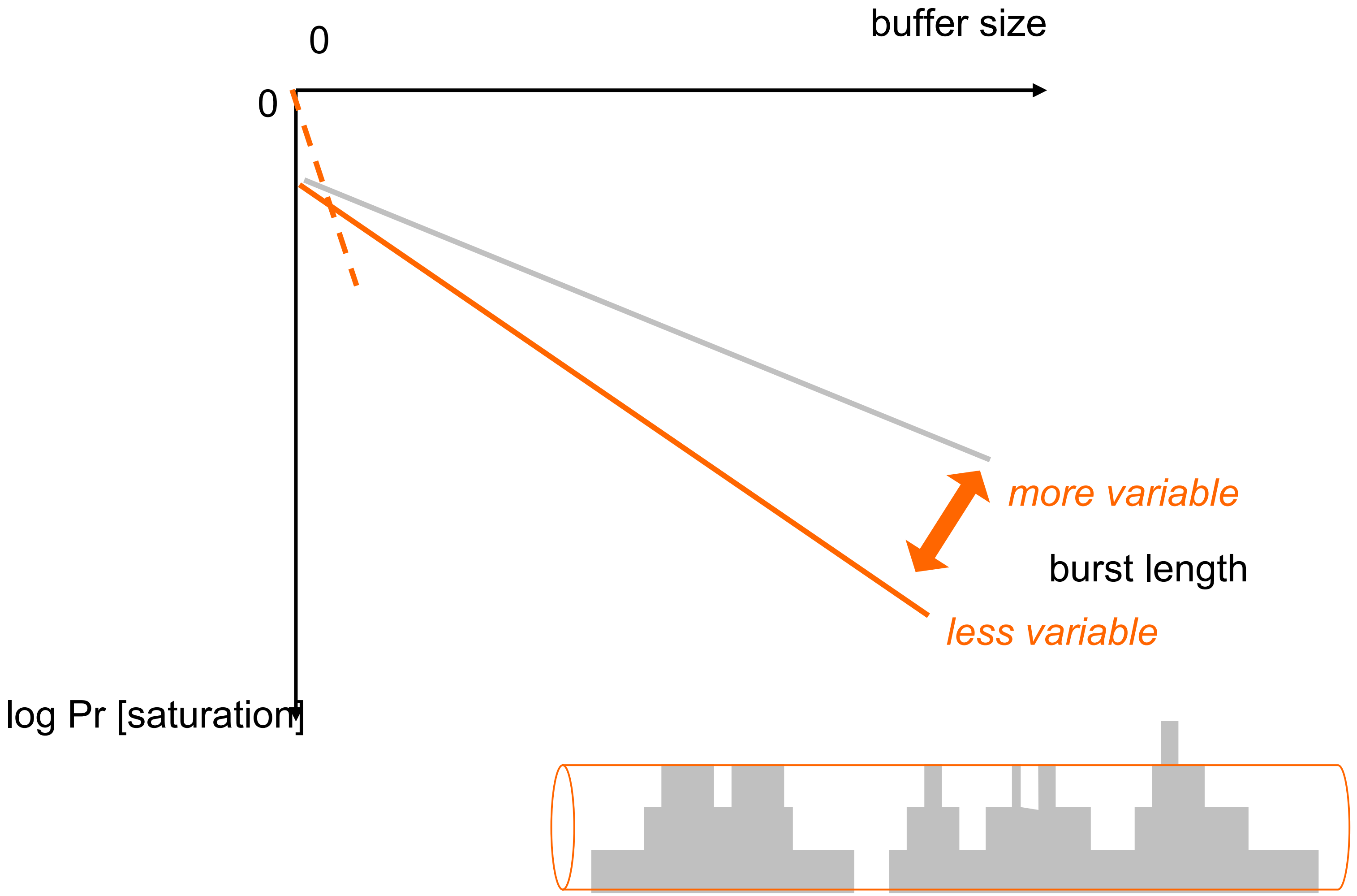




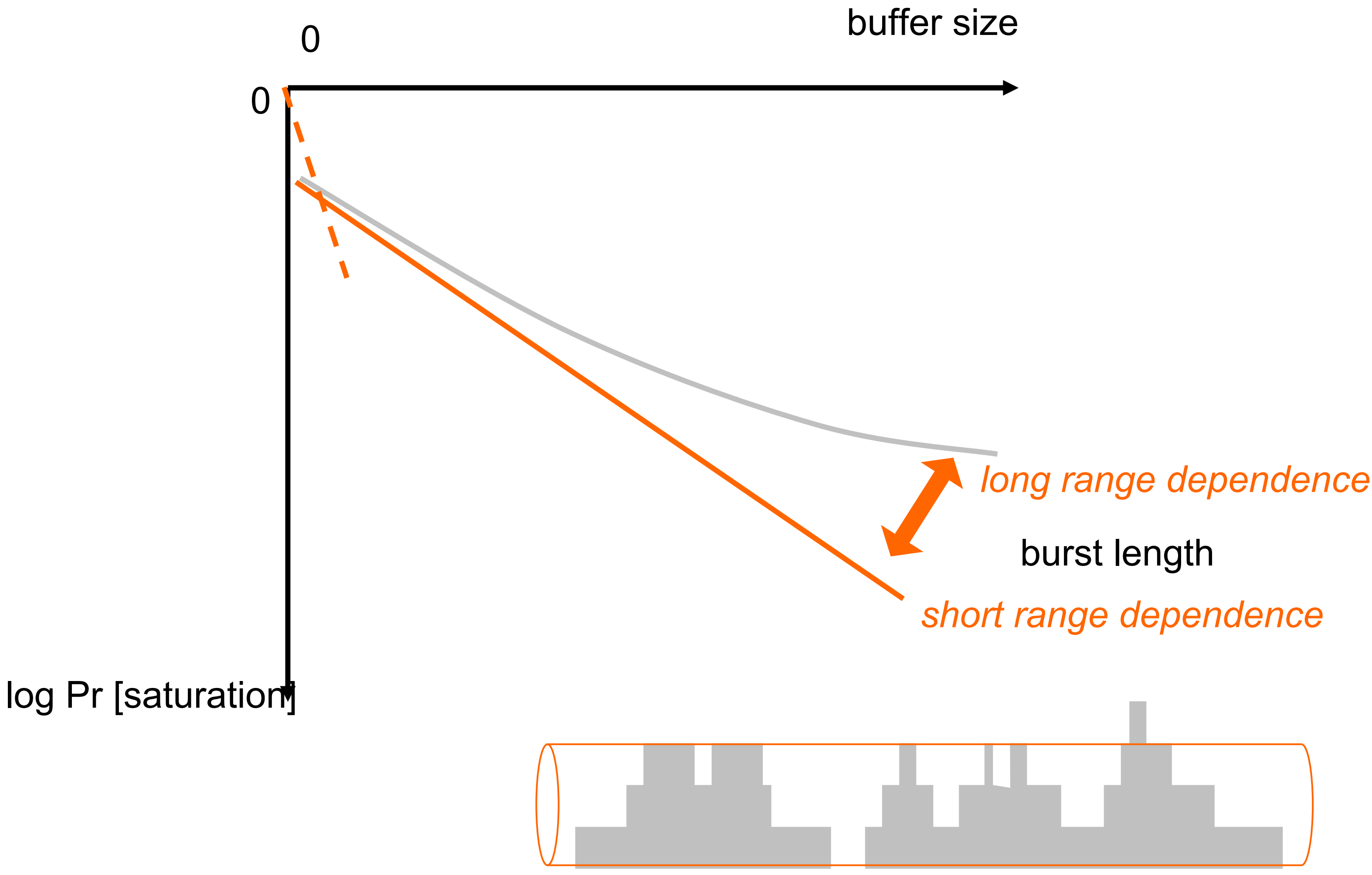
Statistical multiplexing performance: impact of traffic characteristics



Statistical multiplexing performance: impact of traffic characteristics



Statistical multiplexing performance: impact of traffic characteristics

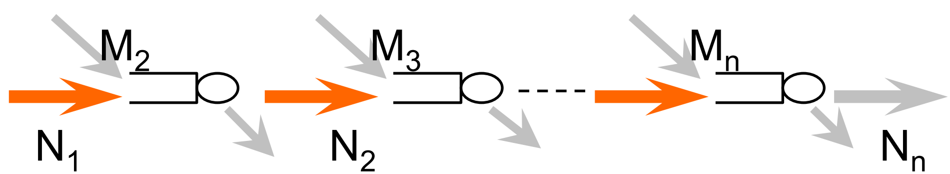


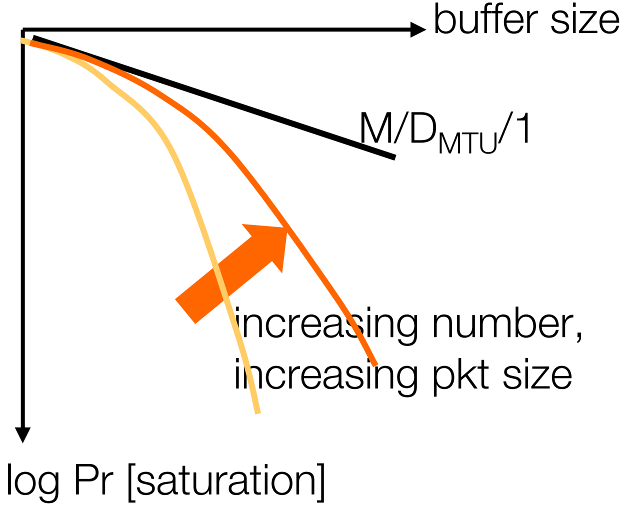
Prefer bufferless multiplexing for streaming traffic

- buffered multiplexing performance depends on detailed traffic characteristics
  - these characteristics are generally unknown and uncontrollable!
- bufferless multiplexing performance depends only on stationary rate distribution
- bufferless multiplexing can be efficient when flow rates are relatively small or streaming traffic is small proportion of whole

Bufferless multiplexing and packet scale queues

- a superposition of nominally constant rate bursts
  - $nD/D/1$, $\Sigma D_i/D/1$, $\Sigma D_i/D^{X_i}/1$ queues
  - delays upper bounded by $M/D_{MTU}/1$ (MTU is max packet size)
- but bursts acquire jitter in multiplexer queues
  - "negligible jitter conjecture": $M/D_{MTU}/1$ remains conservative,
  - partial justification but no proof!
  - except for a saturated tandem
- can use M/D/1 for sizing purposes





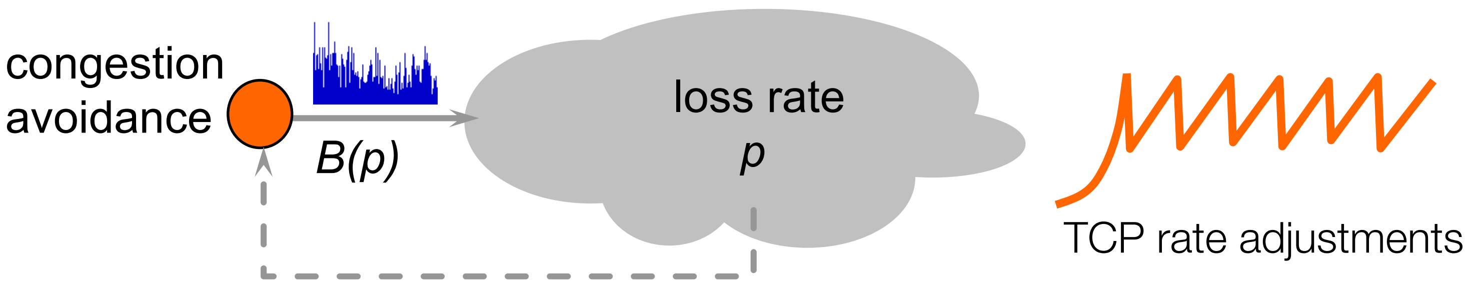
Admission control for streaming traffic: much work but still no perfect solution!

- accept a new flow only if QoS preserved
  - given flow traffic descriptor
  - and current link status
- no satisfactory solution for buffered statistical multiplexing
  - unknown and uncontrollable traffic characteristics
  - means unpredictable performance
- measurement-based control for bufferless statistical multiplexing
  - given flow peak rate and current measured rate (instantaneous rate, mean, variance,...)
  - remains problematic (but see Grossglauser & Tse, 2003)

outline

- nature of Internet traffic
- performance of statistical multiplexing
- performance of statistical bandwidth sharing
- service differentiation
- multi-path routing

Packet loss and bandwidth sharing

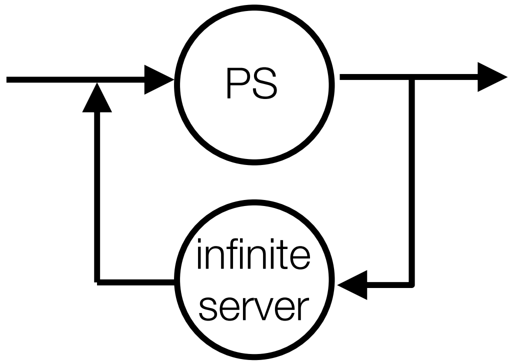


- a queue with a multi-fractal arrival process
  - but loss and bandwidth related by TCP congestion control ("additive increase, multiplicative decrease")
  - the "square root formula": $B(p) \approx \frac{k}{RTT\sqrt{p}}$
- loss is the result of bandwidth sharing
  - ⇒ study response times directly, not packet loss
- shares are inversely proportional to RTT
  - lower response times for short paths

Traffic theory for statistical bandwidth sharing

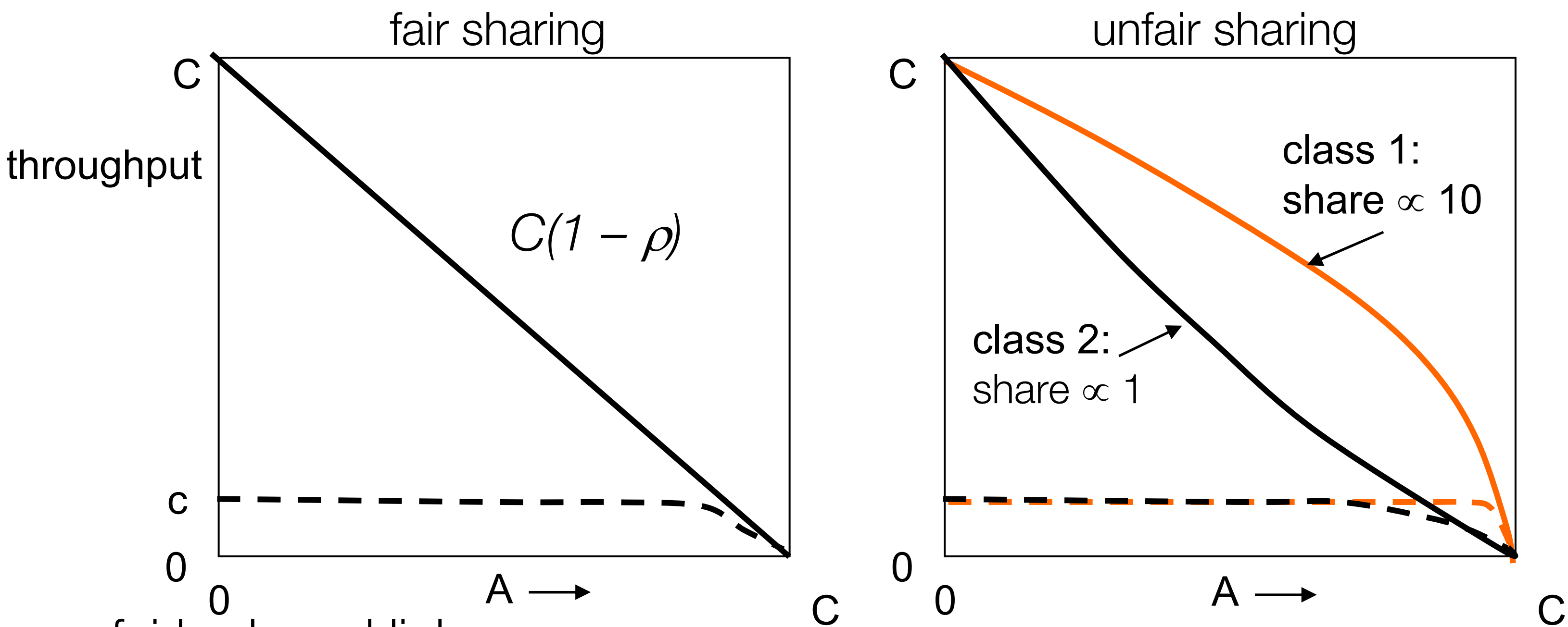
- elastic flows share link bandwidth
  - with some degree of fairness
  - through TCP congestion control
- flow performance is measured by the response time
  - that depends on its share of bandwidth
- traffic theory predicts response time for given capacity and traffic characteristics
  - an arrival process of finite sized flows
  - and a given sharing scheme

Processor sharing model of a single link



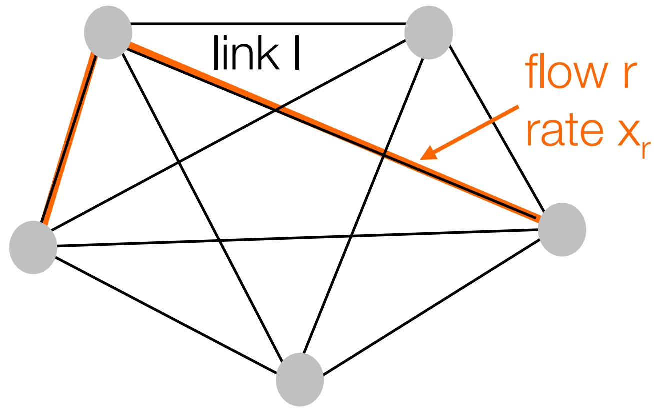
- flows arrive according to the session model
- share link bandwidth fairly (eg, no RTT bias) $\Rightarrow$ a simple stochastic network
- distribution of flow population on link: $\pi(x) = (1 - \rho)\, \rho^x$
- E [response time | size = s] = $s / C(1 - \rho)$
  - so "throughput" = $C(1 - \rho)$
- these results are insensitive:
  - to distributions of flow size, think time, number of flows per session
  - to correlations between successive sizes and times,...
- because service rates are balanced: $\phi_k(x) = \Phi(x-e_k)/\Phi(x)$ for each class k
  - cf. Whittle networks [Serfoso]

Throughput performance



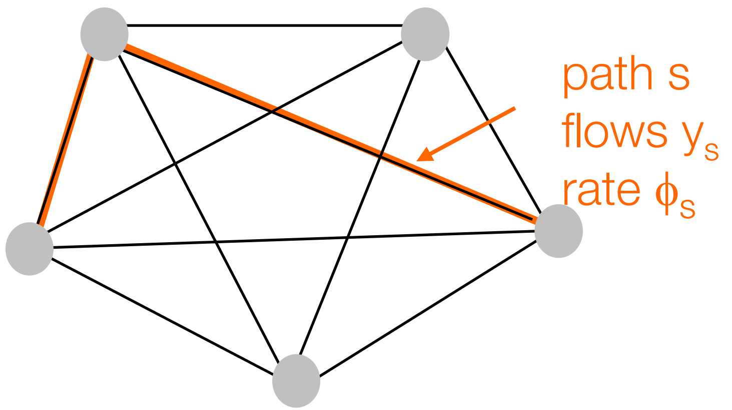
- fairly shared link
  - throughput depends on link capacity C and traffic A, only
  - insensitivity extends to common flow peak rate c
- biased sharing (eg, for different RTT)
  - unequal sharing is sensitive, but not much
  - unfairness significant only at high load

Bandwidth sharing in a network



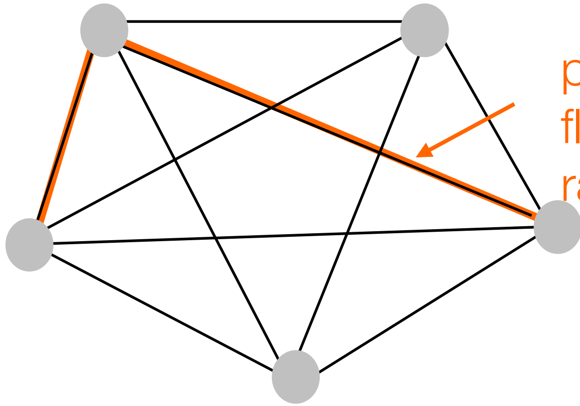
- sharing for maximum utility (Kelly, etc.):
  - choose x_r to maximize $\Sigma_r U_r(x_r)$ subject to $\Sigma_{l \in r} x_r \leq C_l$
  - eg, for "proportional fair" sharing: $U_r(x) = \log x$
- a distributed rate adjustment algorithm
  - eg, for proportional fair: $$\frac{dx_r}{dt} = \kappa_r \left(w - x_r \sum_{l \in r} p_l \left(\sum_{j:l \in j} x_j \right) \right)$$
  - where $p_l(y)$ is the "price" of link l when its load is y: eg, p=packet loss rate
  - a TCP-like algorithm: ie, additive increase, multiplicative decrease

Statistical bandwidth sharing in a network



- let number of flows on path s be y_s
  - assume same utility function and same peak rate c_s so they have equal shares
- utility maximization determines state dependent service rates $\phi_s(y)$
  - satisfying capacity constraints: $\phi_s(y) \le y_s c_s$, $\Sigma_{s \in l}\, \phi_s(y) \le C_l$
- in general, throughput performance evaluation is intractable
  - eg, for proportional fairness or max-min fairness

Statistical bandwidth sharing in a network



- define the alternative "balanced fair" allocation (cf. Bonald & Proutière)
  - $\phi_s(y) = \Phi(y-e_s) / \Phi(y)$
  - for Φ chosen such that the ϕ_s saturate at least one capacity constraint
- by construction, balanced fair bandwidth sharing has a tractable produc form state probability
  - $\pi(y) = \pi(0)\, \Phi(y) \prod A_s^{ys}$
  - where A_s is traffic offered to path s

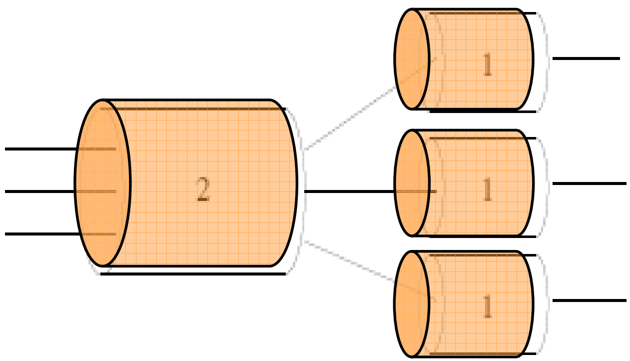
Properties of balanced fairness

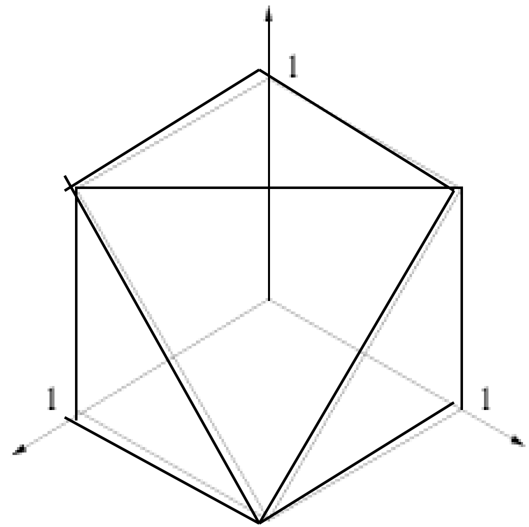
- performance is insensitive for Poisson session traffic model
- computable performance for some interesting cases
  - link sharing with heterogeneous peak rates
  - toy topologies: trees,...
- simple performance bounds for expected response time $R_k(s)$

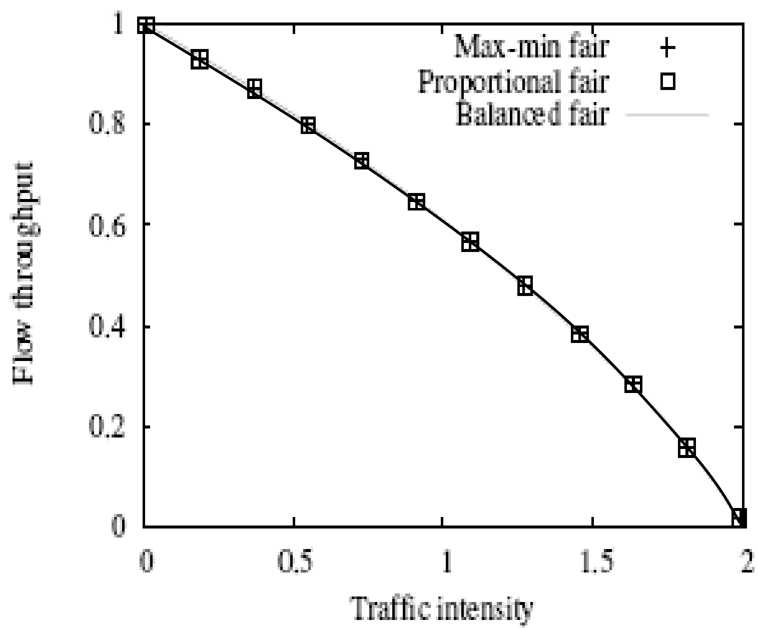
$$\max_{l \in r_k} \left\{ \frac{s}{c_k}, \frac{s}{C_l - A_l} \right\} \leq R_k(s) \leq \frac{s}{c_k} + \sum_{l \in r_k} \frac{s}{C_l - A_l}$$

- provable stability condition: $\rho_l < 1$ for all links
- performance roughly same as utility max allocations
  - eg, proportional fair, max-min fair

Comparison of balanced fairness and other kinds of fairness [BMPV06]







Comparison of balanced fairness and other kinds of fairness [BMPV06]

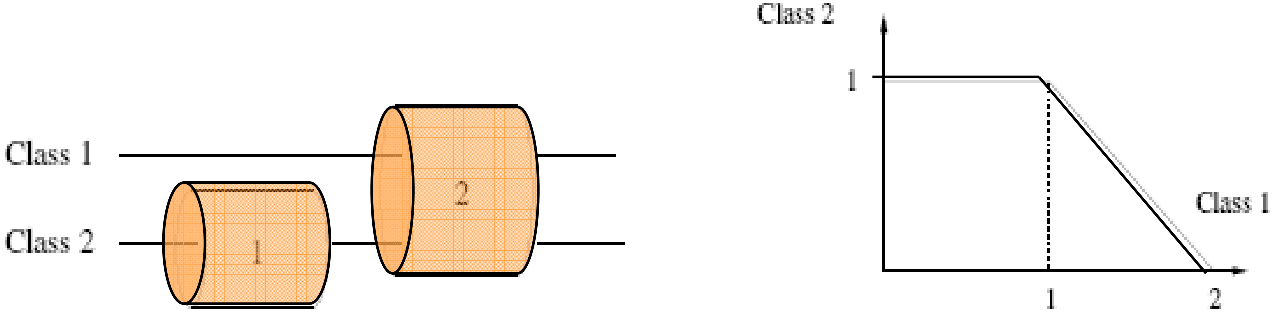
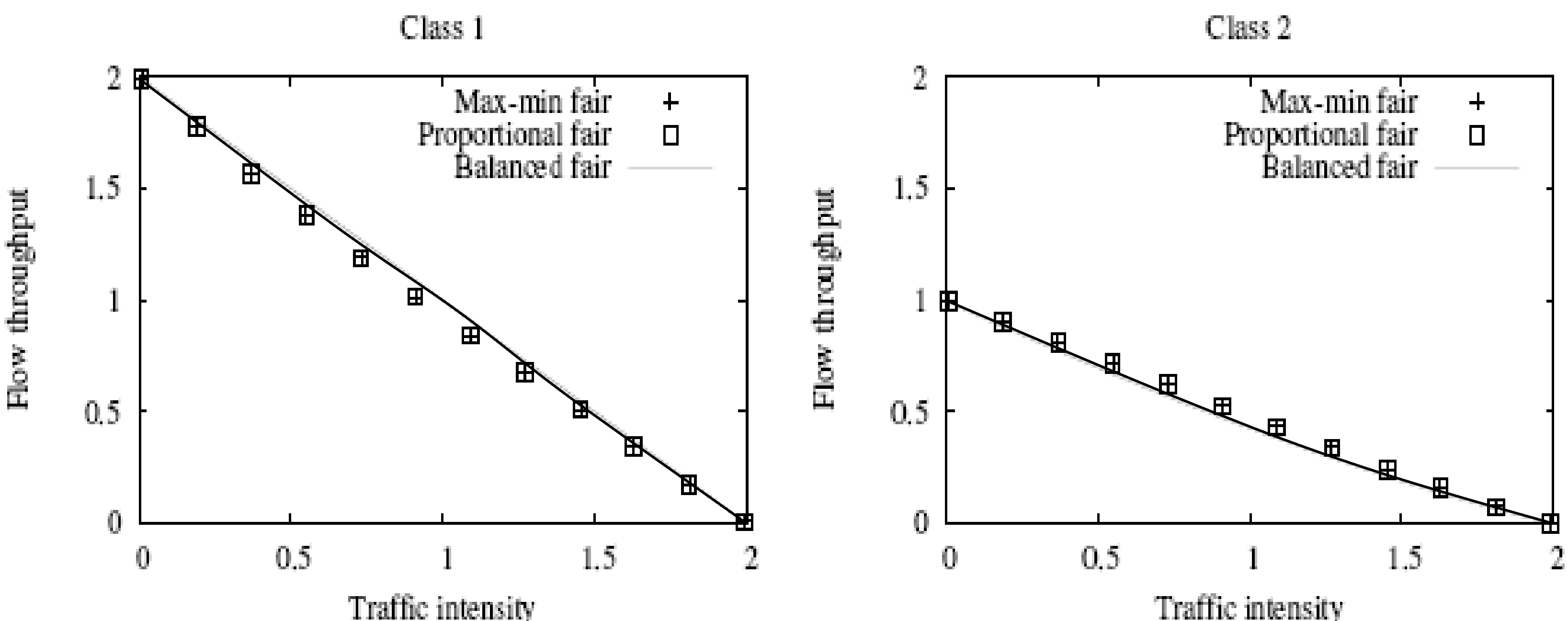


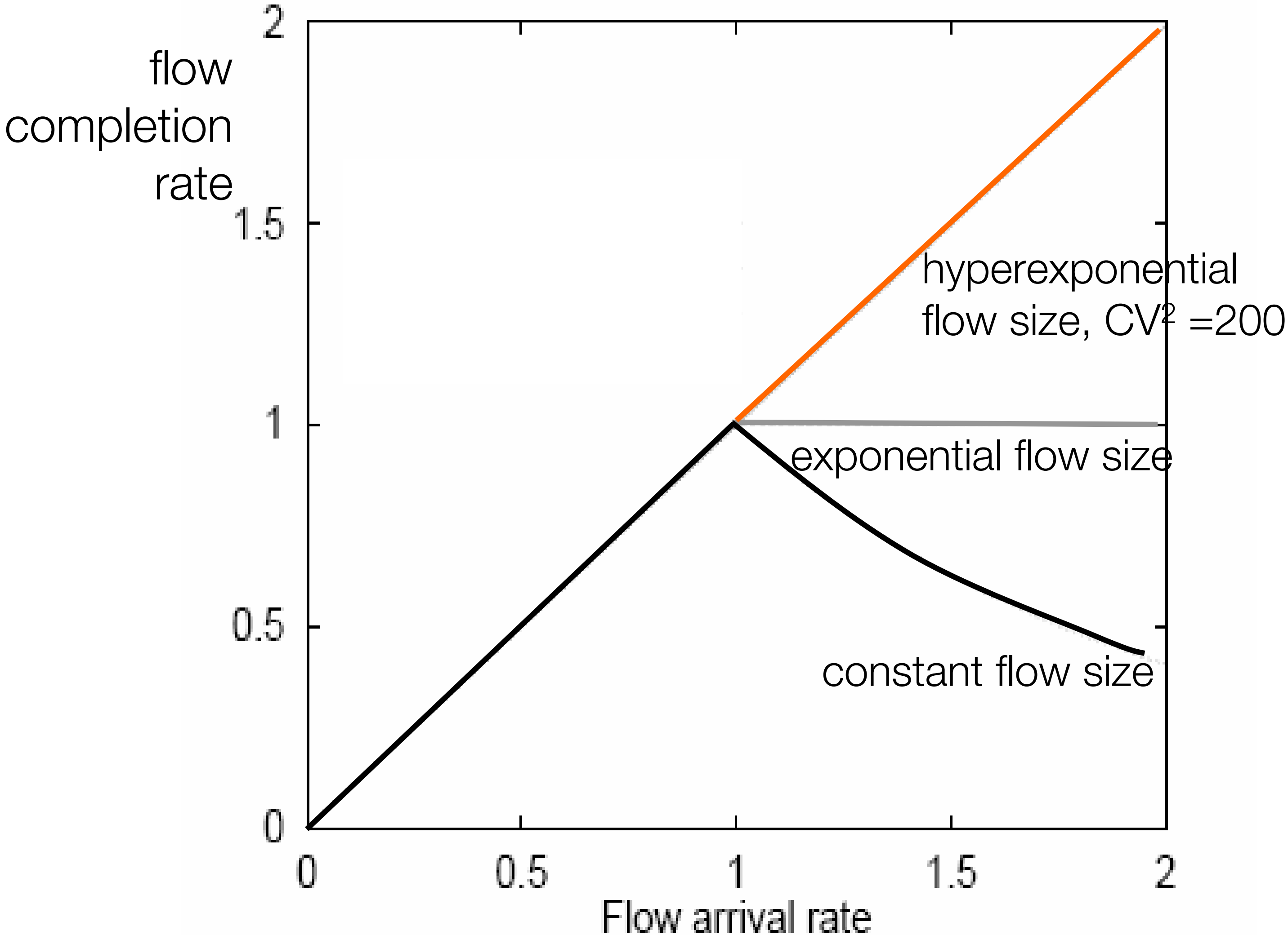
Figure 4: An asymmetric tree network and its capacity set.



Overload and admission control

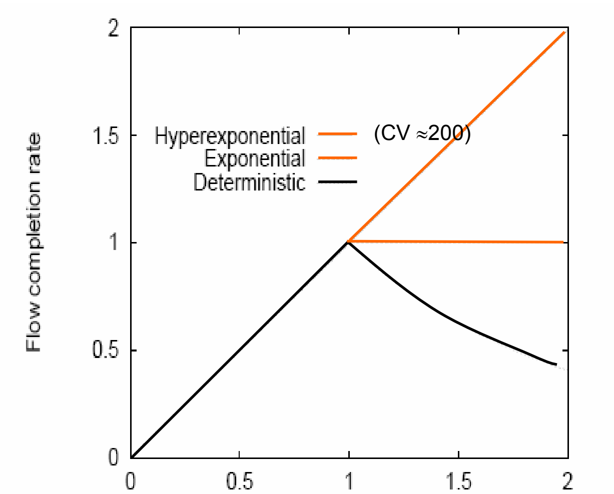
- when $\rho_l > 1$, PS model predicts instability, ie, $\Sigma\, y_s \rightarrow \infty$
- in practice, implies a need for admission control
  - eg, refuse new flows if $\Sigma\, y_s = 100$
- however, if flow size has a heavy-tailed distribution, population explosion may not occur within busy period

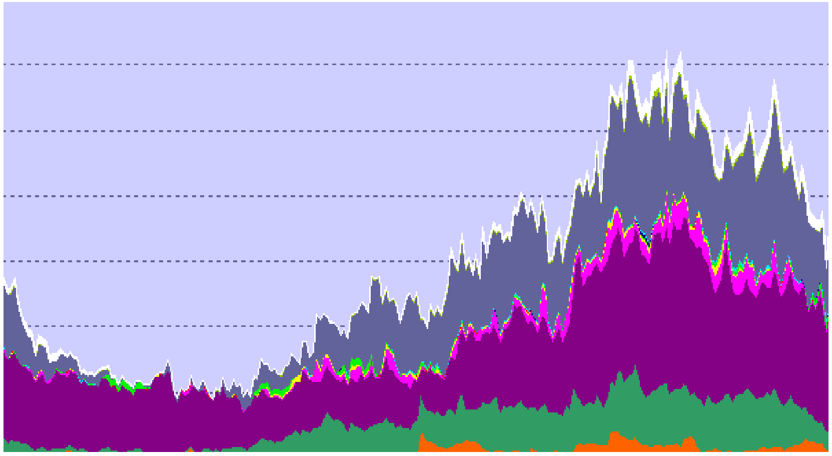
Completion rate of PS server (Jean-Marie & Robert, 1994)



Overload and admission control

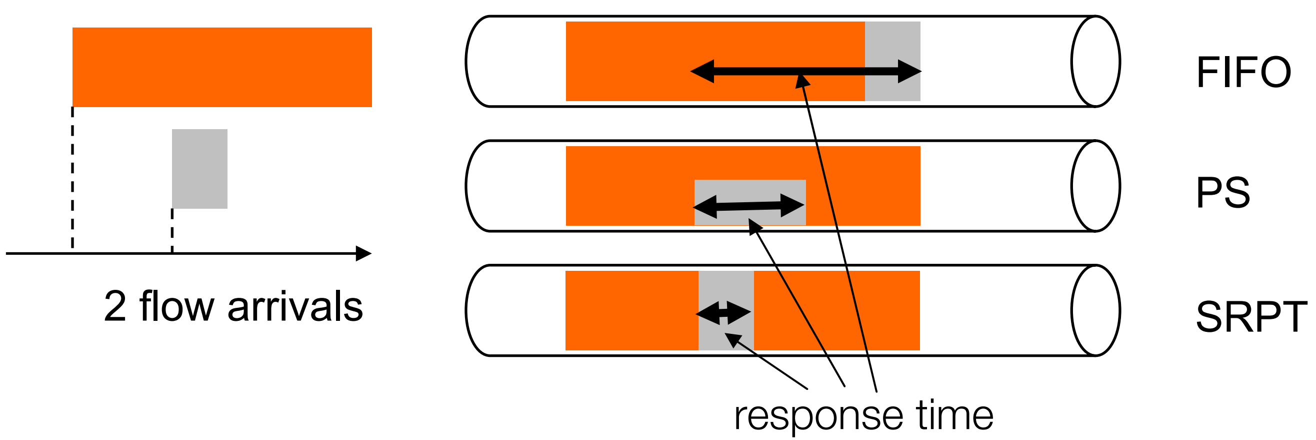
- when $\rho_l > 1$, PS model predicts instability, ie, $\Sigma\, y_s \to \infty$
- in practice, implies a need for admission control
  - eg, refuse new flows if $\Sigma\, y_s = 100$
- however, if flow size has a heavy-tailed distribution, population explosion may not occur within busy period
  - cf. results from Jean-Marie and Robert 1994
  - $\Sigma\, y_s$ may never reach 100 flows





Size-dependent sharing

- throughput performance can be improved by scheduling flows "unfairly", accounting for their size
  - eg, minimum expected response time by "shortest remaining processing time first" (SRPT) service
  - NB. utility maximization ignores this fact!
- performance improves for all flows when size distribution is heavy-tailed
- implementation in Internet
  - practical size-based schedulers exist: least attained service, multi-level PS
  - useful on access links, doubtful in core network



outline

- nature of Internet traffic
- performance of statistical multiplexing
- performance of statistical bandwidth sharing
- service differentiation
- multi-path routing

Integrating streaming and elastic traffic: performance

- class-based priority queuing
  - priority to streaming flows, rely on TCP for elastic flows
  - efficient bandwidth usage and (relatively) simple implementation
- performance analysis is difficult in general
  - "local instability" when residual capacity less than elastic traffic load
  - elastic throughput depends on mean and variance of instability periods
    - $\Rightarrow$ worse performance as streaming flows longer and more variable
  - instability impact less for high elastic flow size variability
    - $\Rightarrow$ better performance for more variable elastic flow size

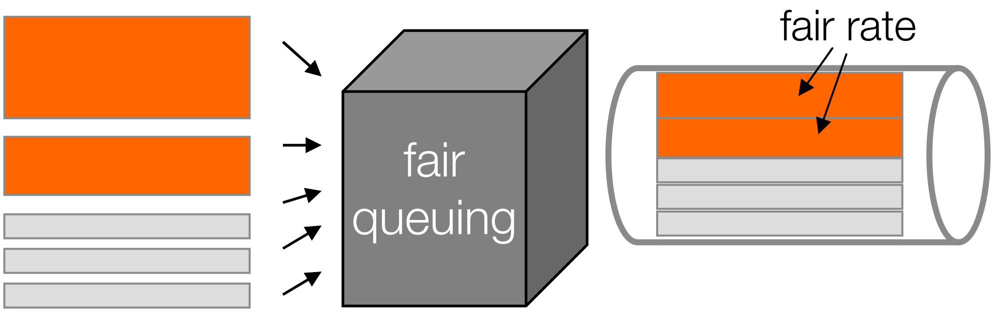
local instability when elastic demand > $C - \Lambda_s(t)$

Integrating streaming and elastic traffic: admission control

- admission control is applied to preserve performance in overload
  - ie, reject new flows when rate would be less than threshold θ
  - apply to streaming and elastic flows
- a quasi-stationary analysis is then accurate
  - ie, assume streaming flow duration is very large so that elastic traffic attains stationary regime between streaming state changes
  - the approximation is insensitive

Implicit service differentiation

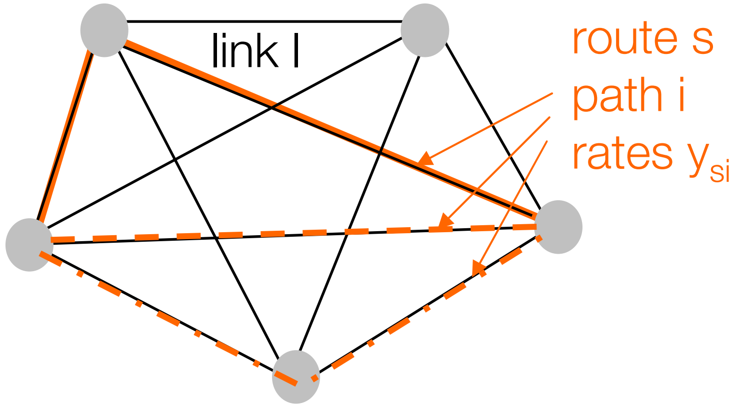
- class of service marking is problematic
  – charging, cheating, policing...
- per-flow fair queuing realizes implicit differentiation
  – imposes max-min sharing, for any congestion control
  – flows of rate < fair rate get low latency
- apply admission control to keep fair rate high enough in overload
- fair queuing is provably scalable
  – few bottlenecked flows, other flows rarely scheduled



outline

- nature of Internet traffic
- performance of statistical multiplexing
- performance of statistical bandwidth sharing
- service differentiation
- multi-path routing

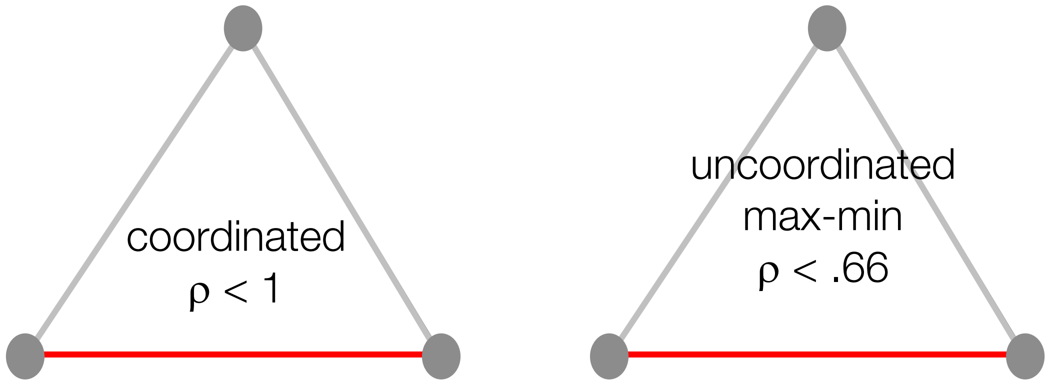
Multi-path routing (in the future Internet)



- for greater reliability, better performance
- a utility maximization formulation (cf. Kelly, etc)
  - maximize $\Sigma\, U(x_s)$ subject to $\Sigma\, y_{si} \le C_l$ for paths i used by route s
  - with $x_s = \Sigma_{i \in s}\, y_{si}$
- a distributed rate adjustment algorithm
  - eg, for proportional fair: $\frac{dy_{si}}{dt} = \kappa_r \left(w - x_s \sum_{l \in r} p_l \left(\sum_{tj:l \in tj} y_{tj} \right) \right)$
  - where p_l is the "price" of link l: eg, packet loss rate
  - note, multiplicative decrease is proportional to x_s
  - a coordinated congestion control protocol

Properties of coordinated congestion control multi-path routing

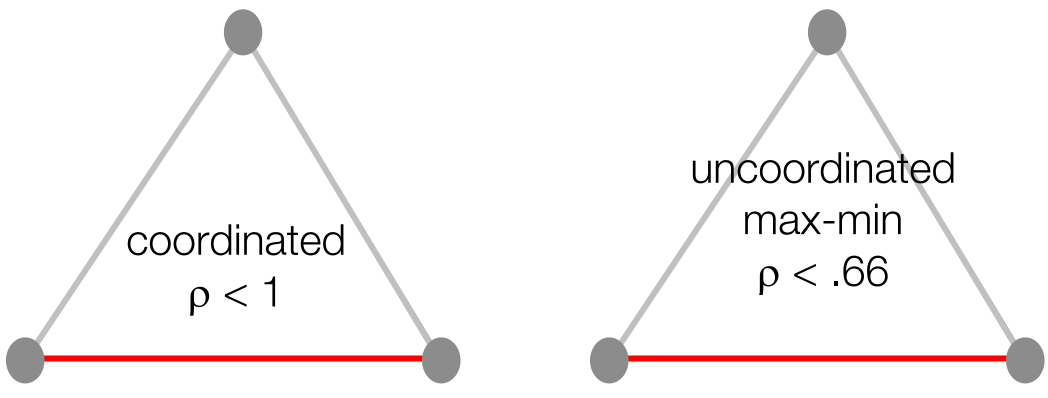
- traffic routed on minimum cost routes only
  - maximizes throughput in light traffic
  - short routes only in heavy traffic
- maximizes the traffic capacity (for any utility function)
  - a significant advantage in a toy 3-node network
- but optimality relies on accurate implementation of coordinated congestion control by end users



link capacity 1
traffic per node pair ρ
Poisson flows
exponential size

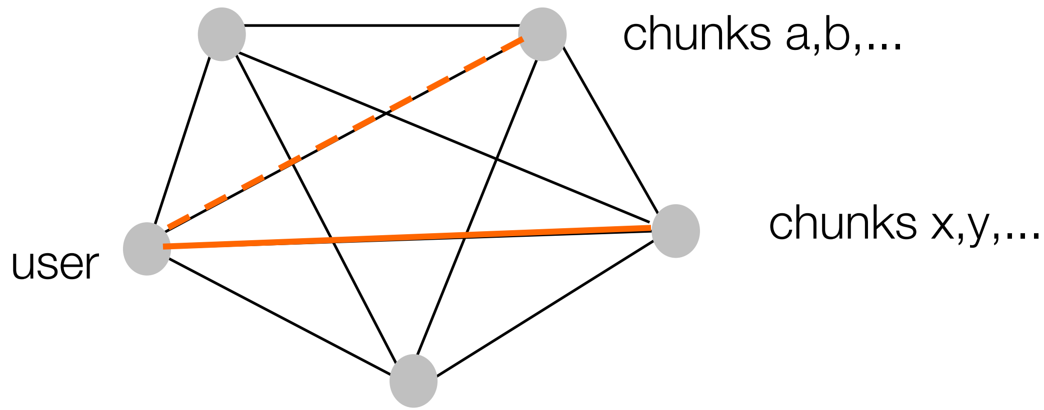
Flow-aware multipath routing

- to avoid relying on end users
- routers locally impose per-flow fair sharing
  - sharing is max-min fair between sub-flows
  - uncoordinated congestion control leading to reduced capacity
- but, admission control can be applied selectively to avoid long paths in heavy traffic (cf "trunk reservation" in phone network)
  - satisfactory performance for triangle network
  - what about performance in a large network?



link capacity 1
traffic per node pair ρ
Poisson flows
exponential size

Impact of overlays?



- overlays like BitTorrent swarms already perform multi-path routing
  – ie, users choose best connected peers
- limited motivation to provide multiple paths (to improve performance and reliability)
- coordinated congestion control is hardly feasible
- is this unfair? should we care?

Conclusions: QoS in the future Internet

- taking account of the lessons of traffic theory
  - bufferless multiplexing for streaming flows
  - approximate fair sharing for elastic traffic
  - for (roughly) insensitive performance
- two alternative promising resource sharing mechanisms
  - distributed congestion control for maximum utility... but avoid relying on altruistic end users, or
  - network imposed per-flow fair sharing... but avoid relying on user flow identification
- though neither may satisfy business requirements or actors in the future Internet value chain!

thank you