QoS in the Future Internet

Sara Oueslati, James Roberts

France Telecom R&D

Workshop on "Visions of Future Generation Networks"
Würzburg, July 31 – August 1, 2006

QoS is still not deployed in the Internet...

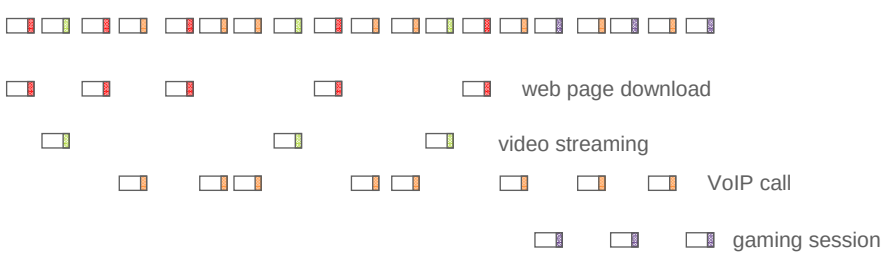
- ➔ so far, little economic incentive
  - users can't tell who is responsible for QoS degradation
  - and, anyway, they can't do much about it!
  - no reward for the invested cost in QoS
- ➔ but also practical difficulties (e.g. Diffserv)
  - marking, policing and inter-domain issues
  - imprecise guarantees ⇒ requires heavy monitoring
  - and are we sure the available mechanisms are sufficient?

... though the need is growing

- ➔ new services, new dangers
  - new streaming services (IPTV, fixed and mobile VoIP, …)
    - requiring low packet latency and loss
  - higher access rates (video servers, enterprise access, FTTH)
    - potentially more aggressive user flows (high rate packet bursts)
- ➔ a new approach to QoS is required
  - over-provisioning is not sustainable
  - reservation doesn't work for variable rate flows
  - Diffserv aggregates are not manageable
- ➔ taking proper account of IP traffic characteristics and resource sharing performance
  - understanding the 3-way relation: QoS ↔ demand ↔ capacity

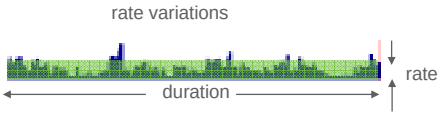
IP traffic as a superposition of flows

- ➔ a complex packet arrival process…
- ➔ …but packets belong to flows
  - identified by header attributes and packet timing
- ➔ facilitating traffic description and performance evaluation

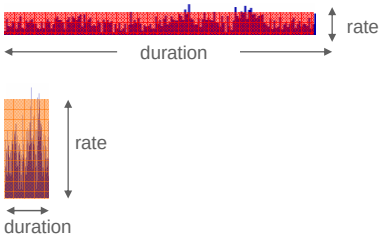


Streaming & elastic flows

➔ streaming flows
- audio, video applications
- intrinsic duration and rate
- QoS=low packet delay and loss

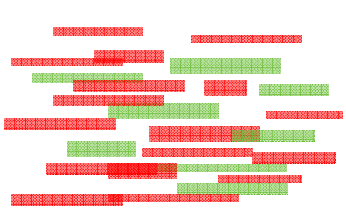


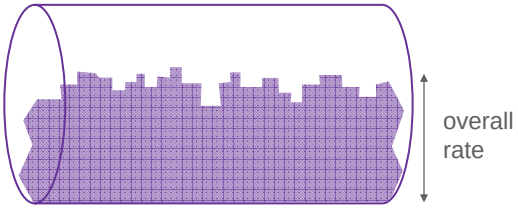
➔ elastic flows
- file transfers
- of a given size (bytes)
- QoS=low response time



Three traffic regimes

1 – the "transparent" regime



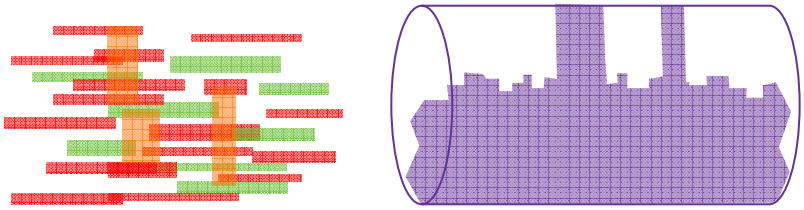


- low rate flows
- an over-provisioned link
- excellent quality for all

⇒ FIFO is sufficient

Three traffic regimes

2 – the "elastic" regime

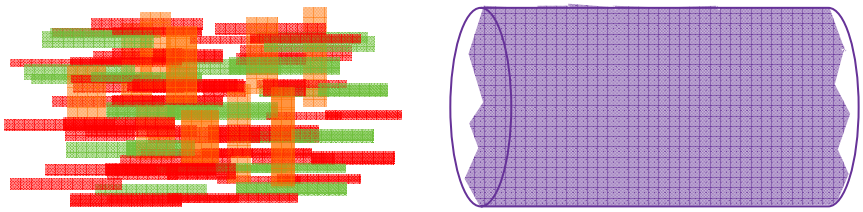


- a few high rate flows
- an over-provisioned link
- temporary degradations

⇒ need for differentiation

Three traffic regimes

3 - the "overload" regime



- too many flows !
- a saturated link
- poor quality for all

⇒ need for overload control

Traffic management in the "elastic" regime

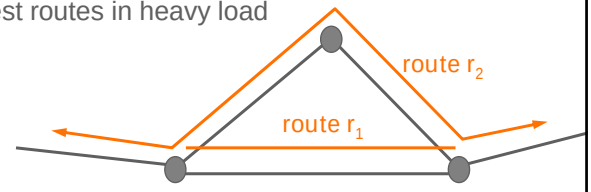
- ➔ flow-oblivious networking
  - enhanced congestion control + AQM (eg, a virtual queue)
    - efficiency and fairness for elastic flows
    - low latency and loss for streaming traffic
  - “a self-managed Internet” (cf. Kelly 2000)
  - but can we rely on user cooperation?
- ➔ flow-aware networking
  - enforce fairness in router queues by fair-queuing
    - independently of end-to-end protocol (TCP, HSTCP,...)
  - realizes implicit streaming/elastic differentiation
    - efficiency and fairness for elastic flows
    - low latency and loss for streaming traffic
  - proven feasibility and scalability of fair-queueing

Traffic management in the "overload" regime

- ➔ admission control: a flow-aware overload control
  - block new flows to protect ongoing flows
  - prefer an implicit approach (i.e., signalling free)
    - by selective packet discard
- ➔ admission control is a component of adaptive routing
  - supposing the network offers multiple paths...
  - ... and users can explore alternative paths
- ➔ overload control in a flow-oblivious network
  - adaptive traffic engineering (cf. Kandula et al, 2005)...
  - ... or multipath congestion control as a partial solution

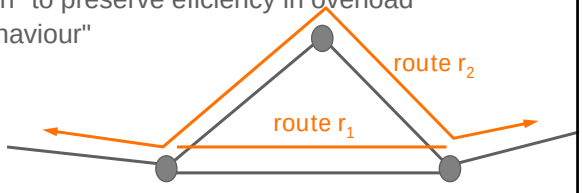
Adaptive routing: an unfortunate omission in IP networking!

- a given flow is currently offered one path only
  - even if the path is congested
  - slow convergence to new path in case of failure
- suppose the network offers several paths
  - how can we choose the best one?
  - how can we optimally use several at the same time?
- "ideal behaviour" of adaptive routing schemes
  - spread traffic over all routes in light load
  - concentrate on shortest routes in heavy load



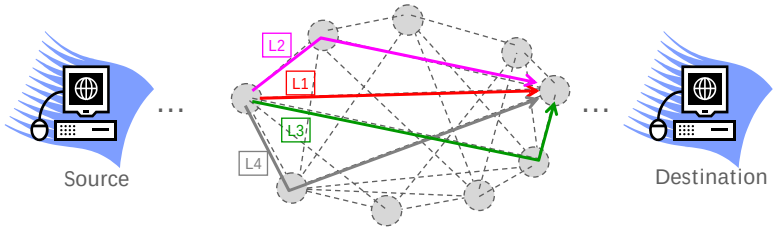
Enhanced performance using multi-path routing

- multi-path congestion control (cf. Massoulié & Key, 2005)
  - use several paths simultaneously
  - coordinated rate control realizes "ideal behaviour"
  - but again, performance relies on user cooperation
  - and no solution for generalized overload
- flow-aware adaptive routing (cf. Oueslati & Roberts, 2006)
  - use one or several paths
  - protect ongoing flows by admission control
  - using "trunk reservation" to preserve eficiency in overload
  - also realizes "ideal behaviour"



Discovering multiple paths using different flow labels

- routing protocols identify several eligible paths
  - eg, one and two (virtual) hop paths using MPLS
- load balance over paths using #{sce@, dest@, flow label}
  - user accesses different paths by changing flow labels
- applies to flow-oblivious and flow-aware networking



Conclusions: QoS is still an issue!

- a need for economic incentives...
- ... but first solve the practical issues
- account for traffic characteristics to realize performance objectives at flow level
- scheduling in the elastic regime
  - prefer flow-aware networking for robustness
- admission control in the overload regime
  - needs a flow-aware network
- adaptive routing: a long overdue Internet enhancement
- what else in a clean slate approach ???

References

- S Kandula, D Katabi, B Davie, A Charny, Walking the tightrope: responsive yet stable traffic engineering, Sigcomm 2005
- F. Kelly, Models for a self-managed Internet, Philosophical Transactions of the Royal Society A358 (2000) 2335-2348. Available at: http://www.statslab.cam.ac.uk/~frank/smi.html
- P. Key, L. Massoulié, D. Towsley, Combining Multipath Routing and Congestion-Control for Robustness, CISS 2006 Available at: http://conf.ee.princeton.edu/ciss/schedule/
- S. Oueslati, J. Roberts, Comparing Flow-Aware and Flow-Oblivious Adaptive Routing, CISS 2006, Princeton, USA. Available at: http://conf.ee.princeton.edu/ciss/schedule/