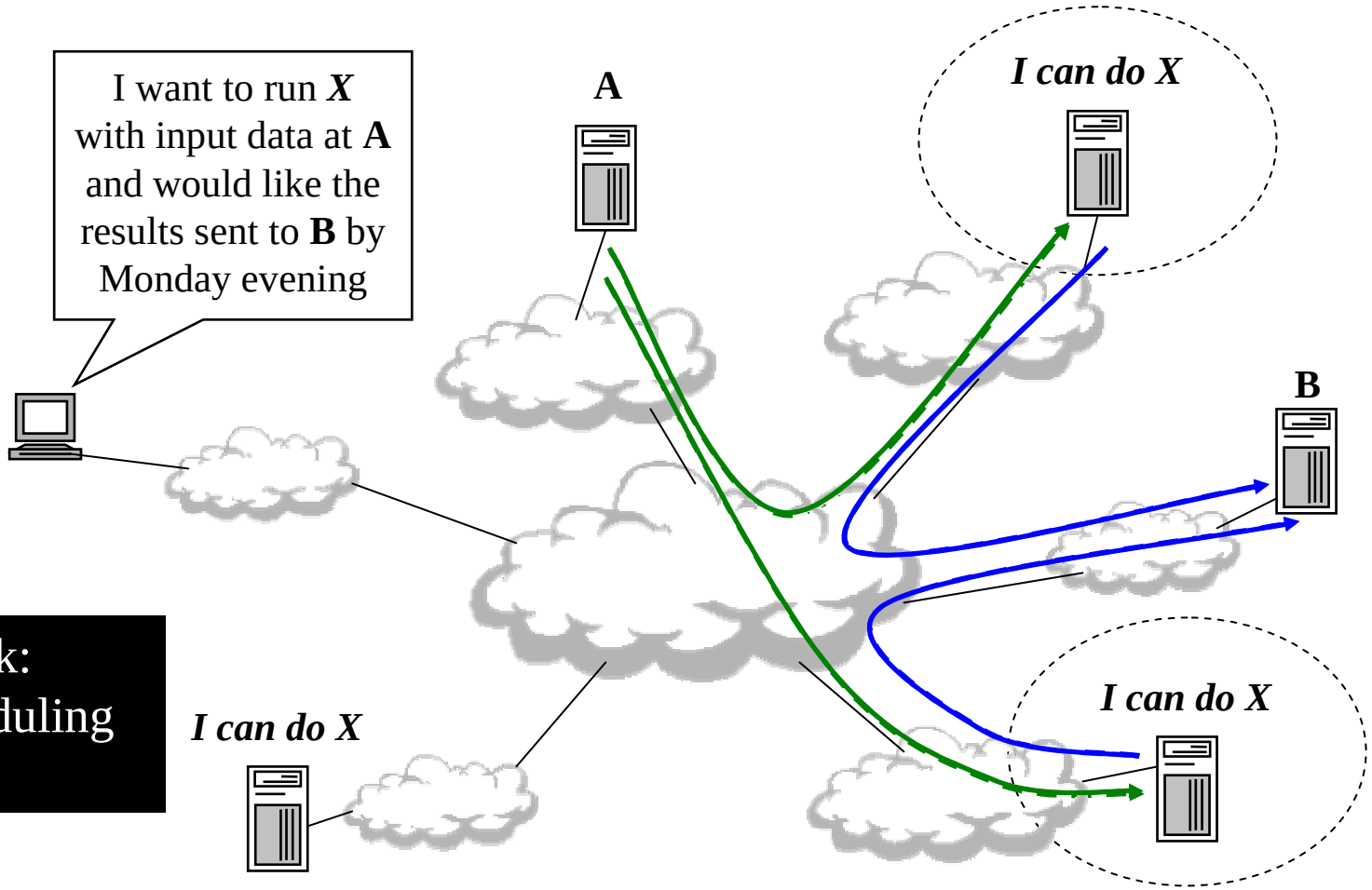
Network QoS support for data intensive distributed applications

Saleem N. Bhatti
Dept. of Computer Science &
Networked Systems Research Group (NETSYS)

http://www.cs.ucl.ac.uk/staff/S.Bhatti/
http://www-mice.cs.ucl.ac.uk/multimedia/
http://www.netsys.ucl.ac.uk/

Overall scenario (example)

harmonised CPU + network scheduling



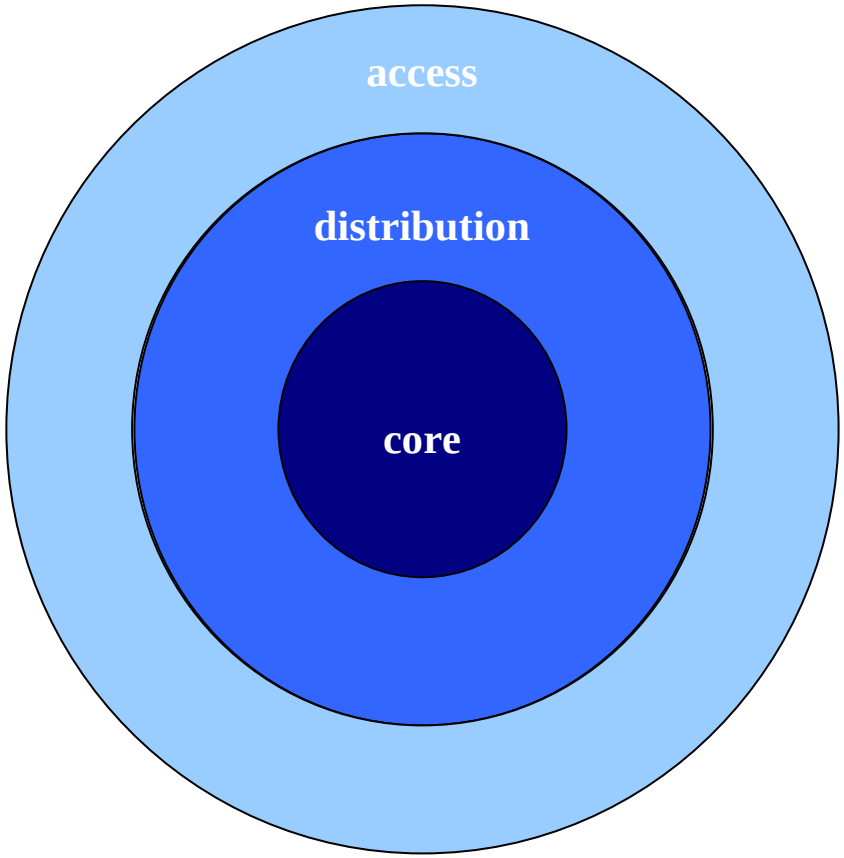
GRS work: network scheduling only

Big data … big problems!

- Particularly relevant to Grid/e-Science
- User in Glasgow wants to access the HGP data
- HGP database:
  - 0.3PB (growing at ~1TB/week)
- SuperJANET4 (SJ4):
  - 10Gb/s backbone (still <2.5Gb/s access in places)
- Extreme case – transfer all of the HGP data
- So, **iff** user gets **all** the SJ4 backbone capacity:
  - transfer of HGP data still takes ~55½ hours!
  - no one else can use the network at all during this time
- **Can't do it!** ☹

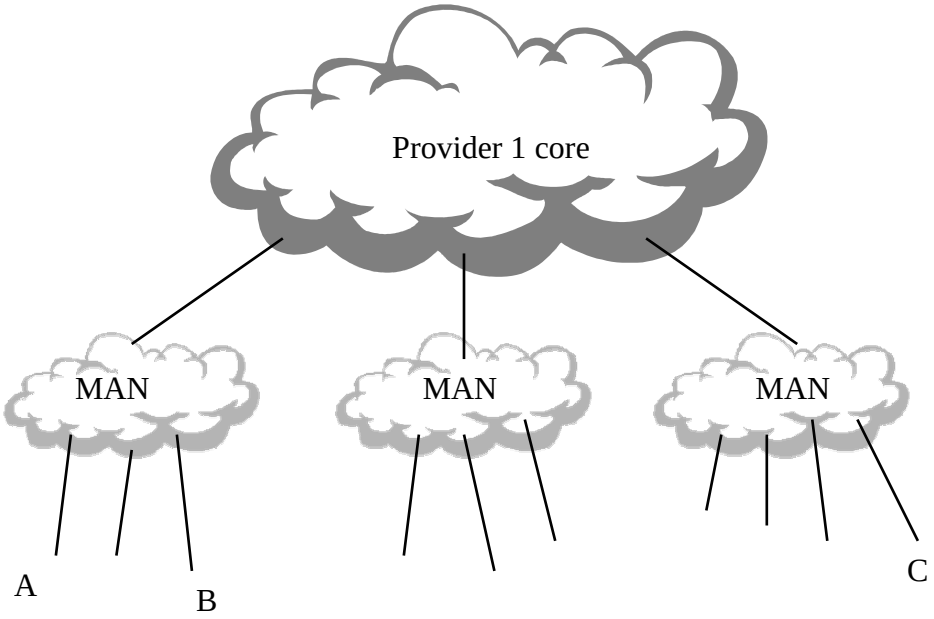
Problem: network hierarchy

- Access network:
  - low multiplexing
  - low volume of traffic
- Distribution network:
  - interconnectivity at local level
  - low multiplexing
  - medium volume of traffic
- Core network – backbone:
  - high volume of traffic
  - high multiplexing
- **Different administrative domains**



Problem: administrative domains

- Network QoS reservations require *state* to be set-up, stored, maintained
- State information:
  - what?
  - where?
  - when?
  - how much?
- General problems:
  - signalling
  - scaling
  - (accounting + charging)

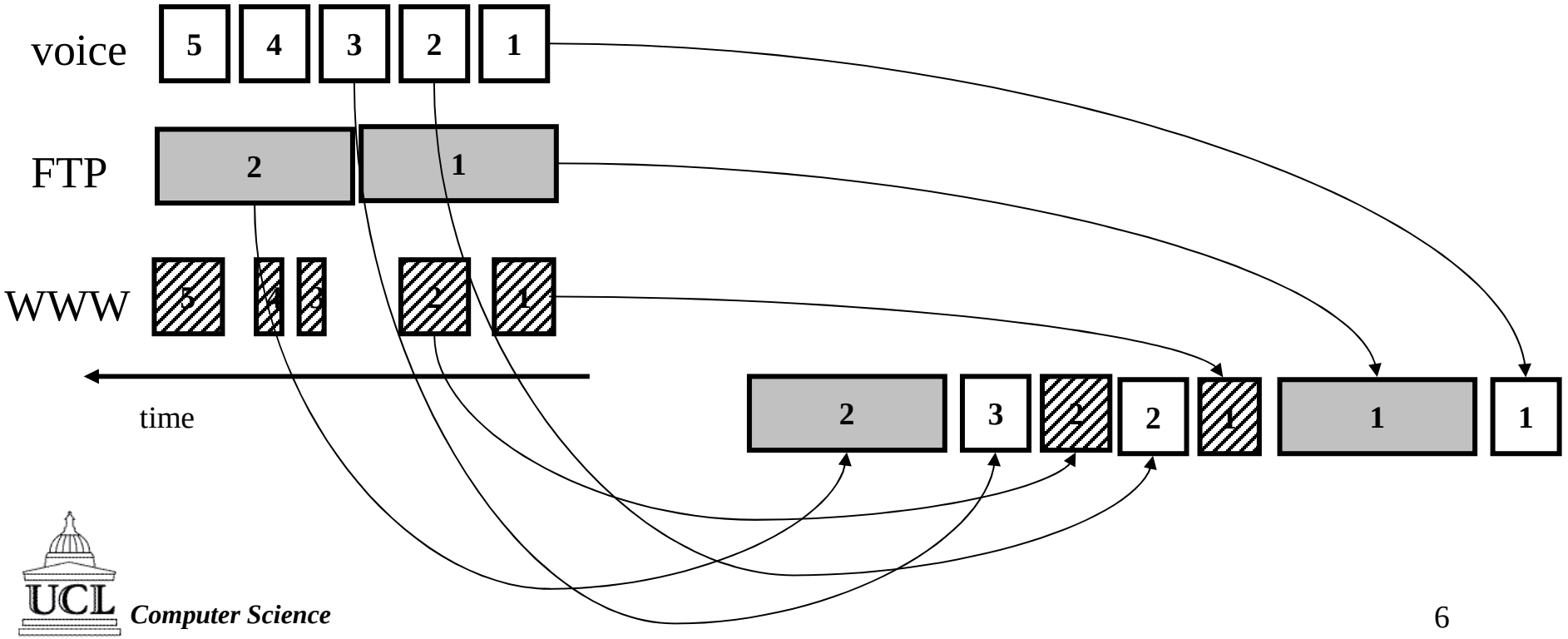


A ⇔ B : localised scope

A ⇔ C: non-localised scope

Problem: mixing traffic

- Example – voice, FTP and WWW traffic through a router:
  - 3 input lines: serviced FCFS at a router
  - 1 output line (1 output buffer)



Problem: modelling traffic

- **Poisson Model used for computational convenience, not for accuracy!**
- V. Paxson, S. Floyd, "*Wide-area Traffic: The Failure of Poisson Modelling*", IEEE/ACM Transactions on Networking, pp.226-244, June 1995.
  http://www.aciri.org/floyd/papers/WAN-poisson.ps.Z
- W. Leland, M. Taqqu, W. Willinger, D. Wilson, "*On the Self-Similar Nature of Ethernet Traffic (Extended Version*)", IEEE/ACM Transactions on Networking, 2(1), pp. 1-15, February 1994.
  http://math.bu.edu/people/murad/pub/source-printed-version-posted.ps
- Mark Crovella, Azer Bestavros, "*Self-similarity in world wide web traffic: Evidence and possible causes*. IEEE/ACM Transactions on Networking, 5(6):835-846, December 1997.
  http://www.cs.bu.edu/fac/best/res/papers/ton97.ps
- V. Paxson, S. Floyd, "*Why We Don't Know How to Simulate the Internet*", Proc. 1997 Winter Simulation Conference, December 1997.
  http://www.aciri.org/floyd/papers/wsc97.ps

Problem: network traffic profiles

So what can we do about it?

- **Build a new and better network (of course)!**
  - very high capacity (Gb/s → Tb/s → Pb/s → Eb/s)
  - users can have access from their desktop
  - provide (QoS-)controlled access
- Two broad problems to consider:
  - **control**: *how do we mix different types of traffic and still control the traffic flows in the network sensibly?*
  - **capacity**: *what happens when you run a very high capacity network with very high capacity access links?*
- This talk highlights the QoS ***Research*** issues:
  - there are also ***Operational*** issues! (but that's SEP ☺)

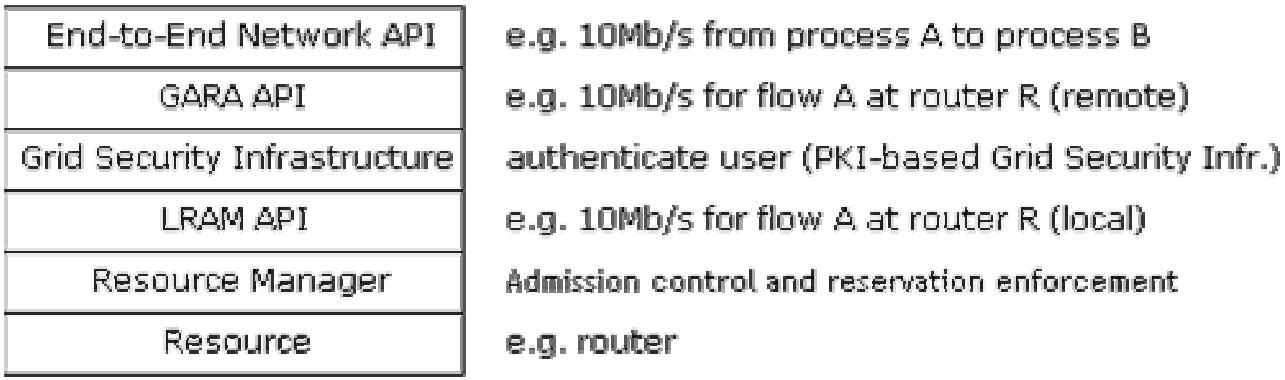
What do we need?

- Application-level access mechanisms:
  - APIs and signalling protocols
  - access mechanisms
- Network mechanisms:
  - QoS architecture + IP-level QoS mechanisms
  - network management and accounting
- High-speed transmission and network components:
  - optical and hybrid-optical systems
- Administration:
  - access infrastructure and mechanisms
  - security and access control
  - charging and billing
  - SLAs, audit trails, etc.

QoS reservations in a Grid environment

- IETF work:
  - INTSERV/RSVP – poor scaling
  - DIFFSERV – deployment and resource management

- GARA – http://www-fp.mcs.anl.gov/qos/
  - based on Globus



Problems with GARA

- General purpose:
  - reservations of CPU cycles, disc, etc.
  - some problems with network-specific usage
- Multiple domains:
  - security and access control
  - managing reservation state
  - centralised management – scaling
  - API (EENR API in progress)
  - time synchronisation
  - tied to Globus (though in principle it is “portable”)

GRS – distributed reservations

- http://www.cs.ucl.ac.uk/research/grs/
- Uses peer-to-peer signalling:
  - exploits IETF standard protocols
  - (aim to build OGSA interface)
- Tries to maintain state at edges of network
- Highly decentralised
- Localised scope for security and access control
- Allows notifications for:
  - QoS reservation set-up tear down
  - QoS changes and violations (user or network)

General problems

- Multiple domain operation
- Changing research usage:
  - “big science” Grid users: HEP, bio-informatics, etc.
- Changing and (mostly) unpredictable traffic patterns (access and core)
- Changing networking landscape:
  - access speeds vs. core speeds
  - over-provisioning may not cut it in the future
- **Complex system behaviour:**
  - learning curve – technical and operational
  - next generation HE networks (SJ5, SJ6 …)