Managed Bandwidth Next Generation

F. Saka

UCL NETSYS (NETwork SYStems centre of excellence)

http://www.mb-ng.net

Outline

- Motivation and project achievements
- Areas of study and findings
- Future activities
- Conclusion

Industrial and research partners

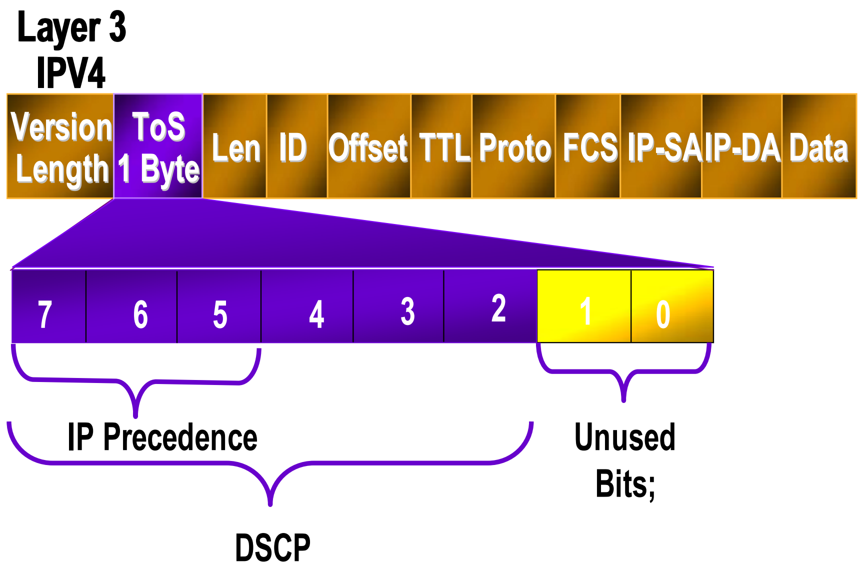
- **MB-NG (http://www.mb-ng.net)**
  - Academic partners: UCL, Manchester, Southampton, RAL, Cambridge, Lancaster.
  - Industrial partners: Cisco, Spirent and Acterna, UKERNA & ULCC.
- **Significant support and interest!**

Project Achievements

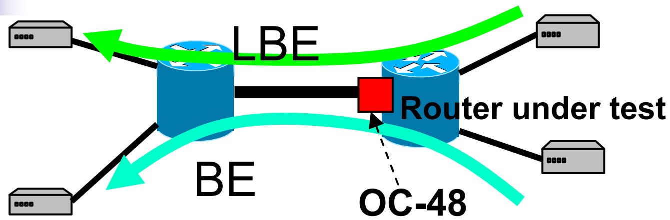
- Implemented Practical QoS Test-bed in the UK.
  - Multi-domain
  - MPLS in the core
  - Diffserv enabled
- Allowing Applications to use QoS enabled Internet via a Middleware
- High-bandwidth data transfers
- End-to-End network services to Europe and US

QoS

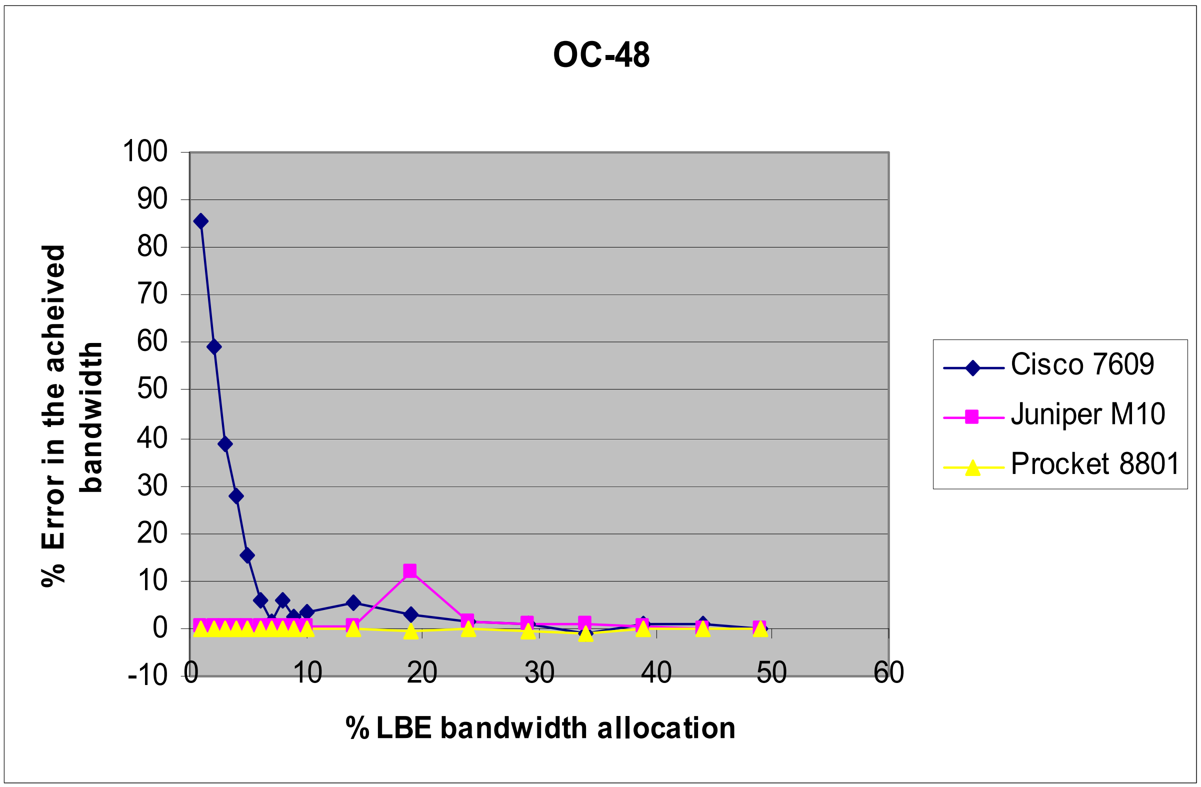
- End-to-end using the differentiated services model (Diffserv)
- Achieved multiple classes (EF, BE and LBE).
- Multiple congestion points (Edge and core domain, different router models)



Router performance (UDP-CBR)

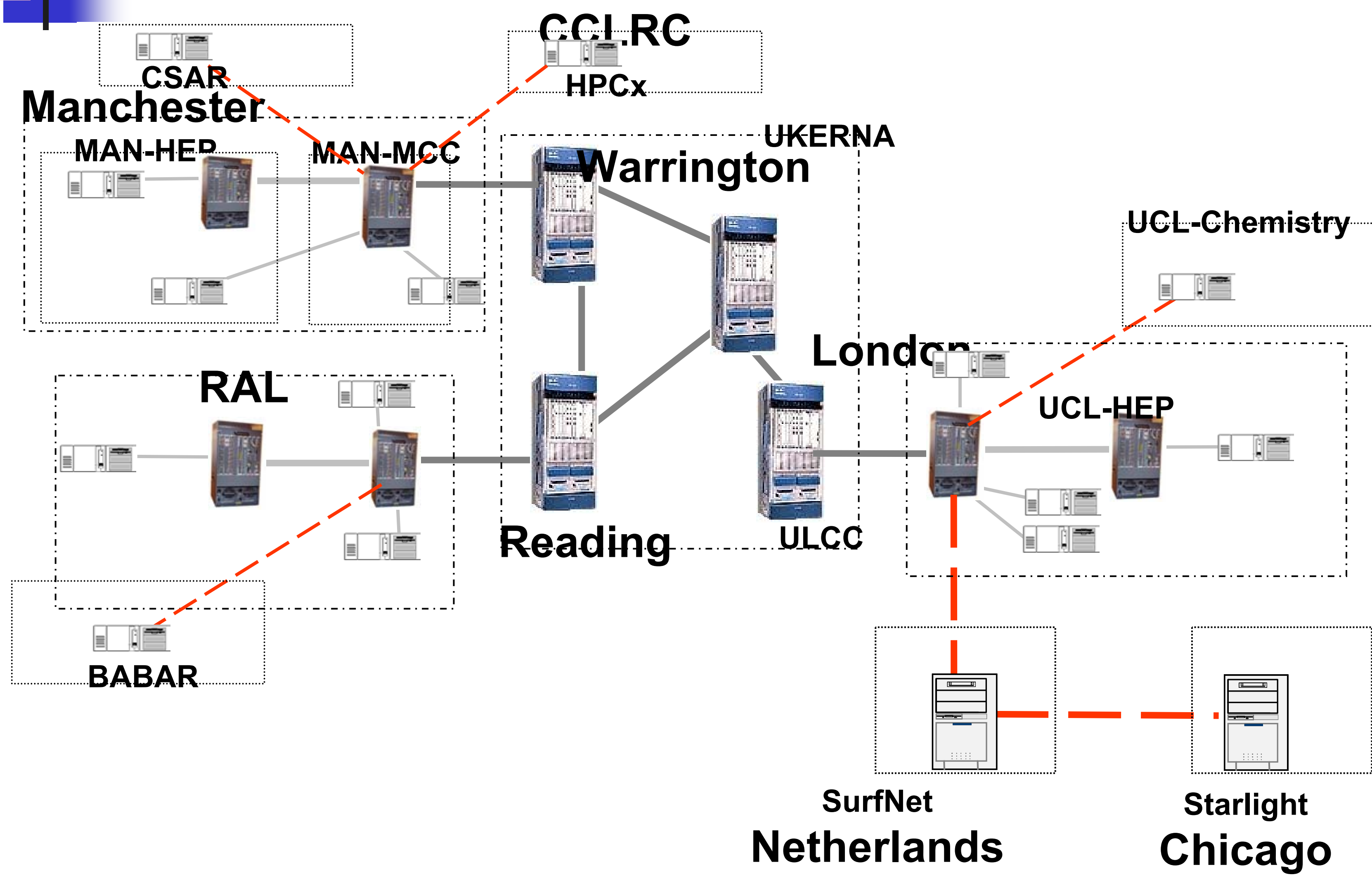


- 2 Classes: BE and LBE
- Error in BE class always below 2.5% for all routers



- All routers perform well.
- Configuration depends on routers, line card and IOS version.
- Scheduling algorithms:
  - Cisco 7609: Weighted Fair Queuing (WFQ)
  - Juniper M10: Weighted Round Robin (WRR)
  - Procket 8801: Deficit Weighted Round Robin (DWRR).
- Settings above 90-10 is not recommended.

The MB-NG network



MPLS

- Deployed in the core of the MB-NG network.
- MPLS with Traffic Engineering extensions helps in enabling efficient utilization of available networks resources.
- Carried extensive testing to check capabilities of Tunnels in respect of bandwidth reservation.

➡ Because RSVP works on the control plane only, QoS still need to be extensively deployed.

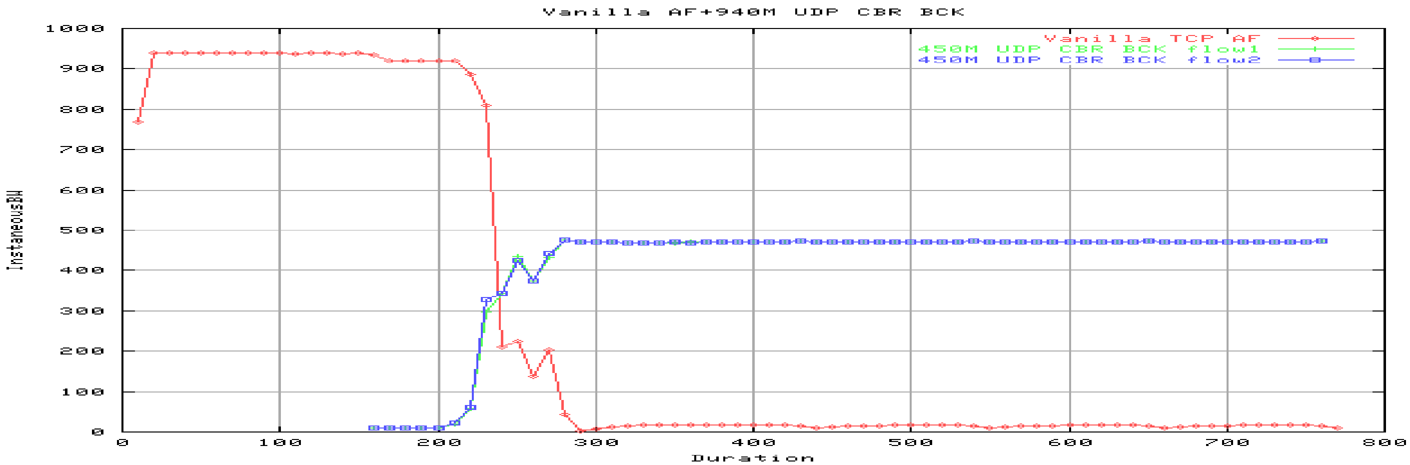
➡ Tunnels ease end-to-end traffic management but are not a complete solution to bandwidth allocation.

QoS and TCP 1

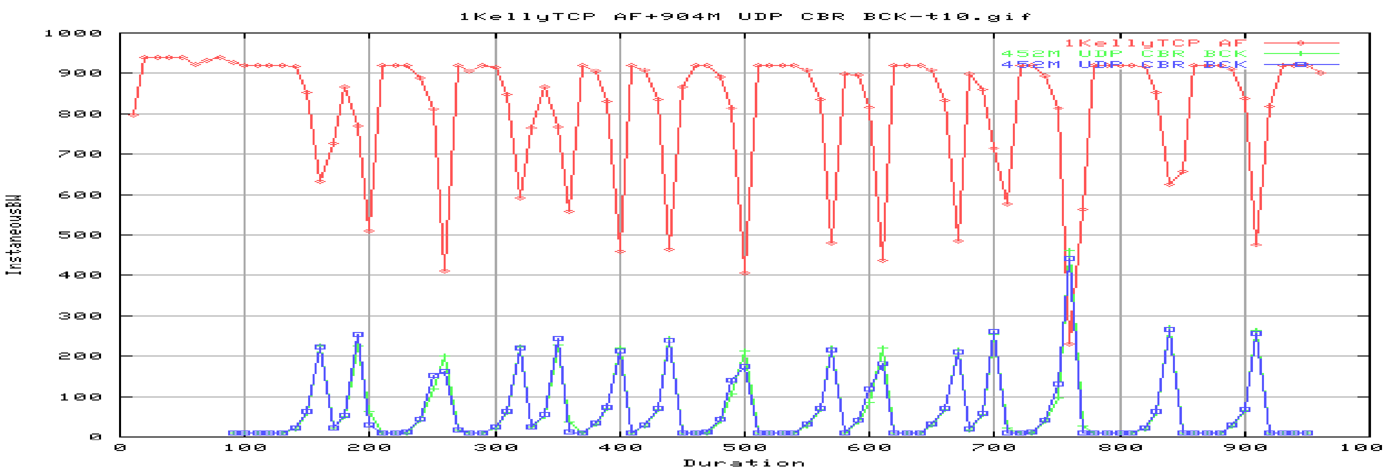
- QoS Works, But can TCP take advantage?
- The main issue is TCP in high-bandwidth delay product networks
  - To do with what is known as “Additive increase Multiplicative Decrease” (AIMD)
  - RTT in the UK of 6ms
  - RTT across Europe and to the US of 100 – 200ms

TCP on DataTAG (RTT=120ms). QoS enabled

Standard TCP



Scalable TCP



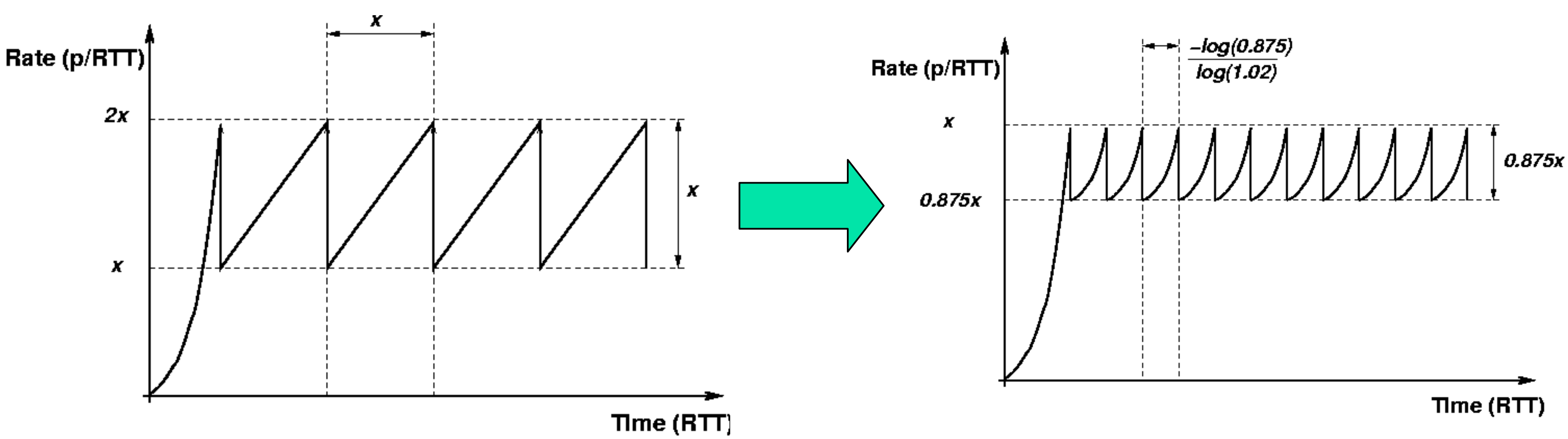
QoS and TCP 2

- Active research area illustrated by the availability of several protocols
  - HS-TCP: HighSpeed TCP S. Floyd http://www.icir.org/floyd/hstcp.html
  - Scalable TCP: T. Kelly http://www-lce.eng.cam.ac.uk/~ctk21/scalable/

- Other include
  - FAST-TCP: S. Low http://netlab.caltech.edu/FAST/
  - H-TCP: Hamilton institute http://hamilton.ie/net/
  - BIC-TCP: Binary Increase Control I. Rhee http://www.csc.ncsu.edu/faculty/rhee/export/bitcp/
  - XCP: Explicit Congestion Control Protocol http://www.ana.lcs.mit.edu/dina/XCP/
  - SABUL: Simple Available Bandwidth Utilization Library http://www.evl.uic.edu/eric/atp/sabul.pdf

TCP Improvement

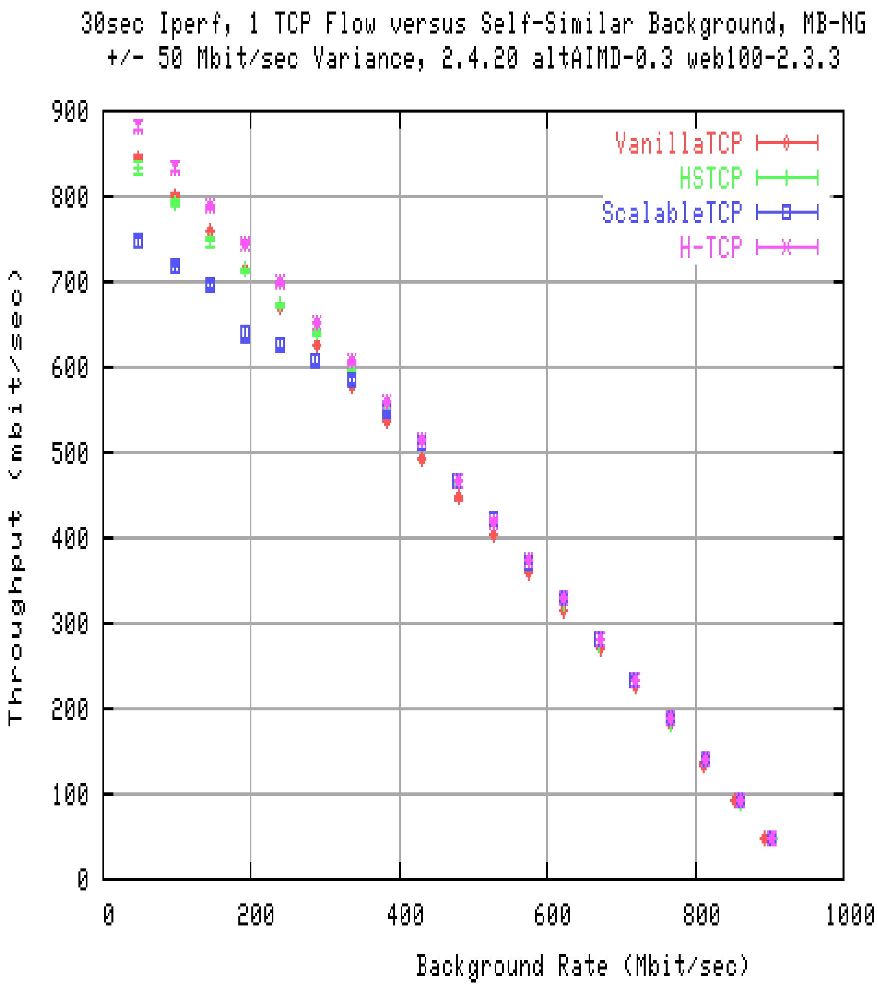
- Basic idea is to change the behaviour in Congestion avoidance
  - Reduce the decrease parameter compared to normal TCP
  - Increase the increase parameter compared to normal TCP

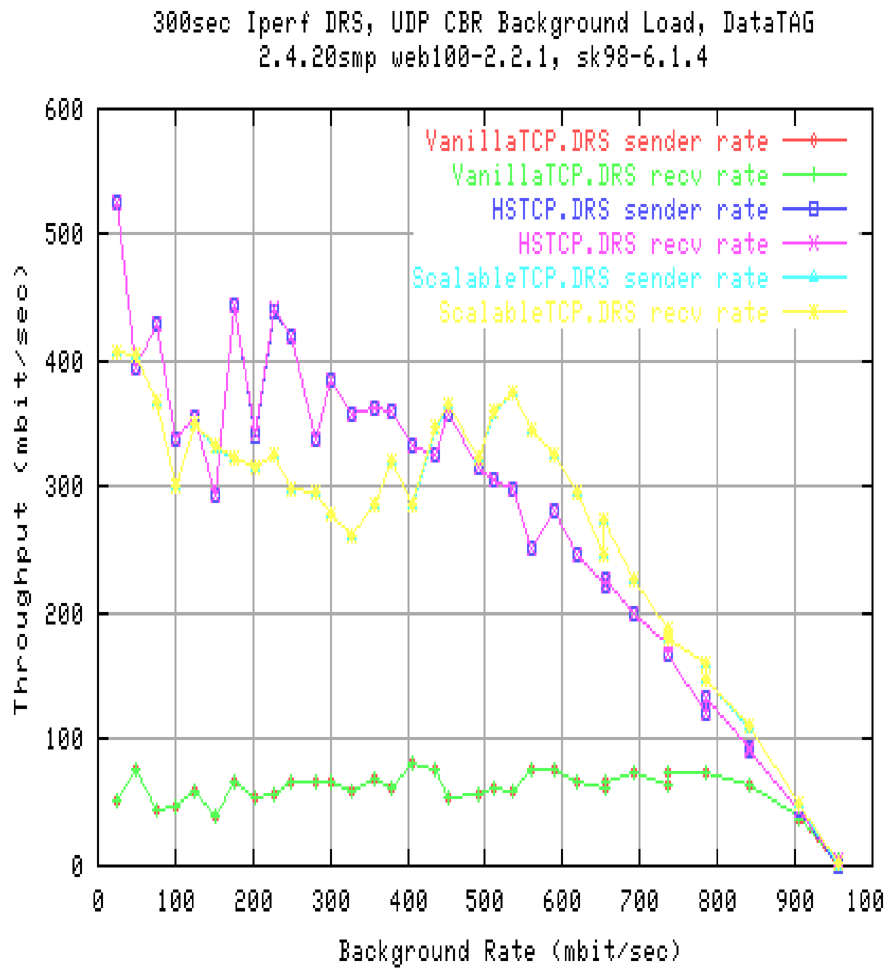


Standard TCP

Scalable TCP

Performance of various protocols





QoS and TCP 4

- Not for fairness, below a certain CWND, they behave like standard TCP.
- Above this CWND, they are unfair
- Possible to put these protocols into a separate class

High throughput

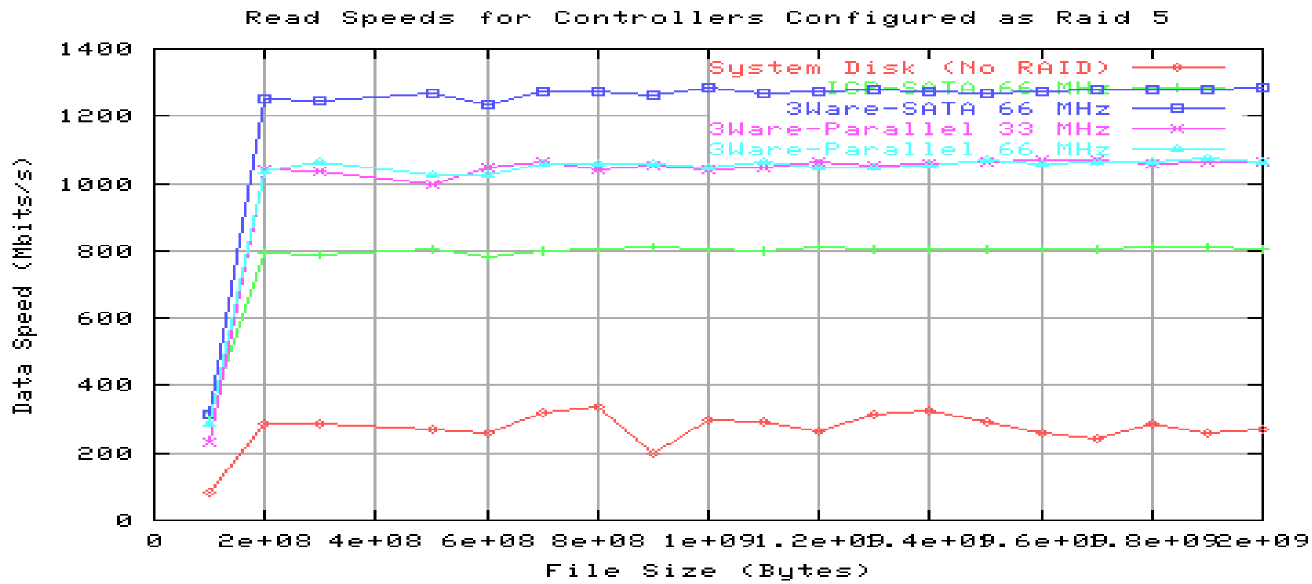
- Investigating bulk data transfers
- Investigating performance of raid0 and raid5

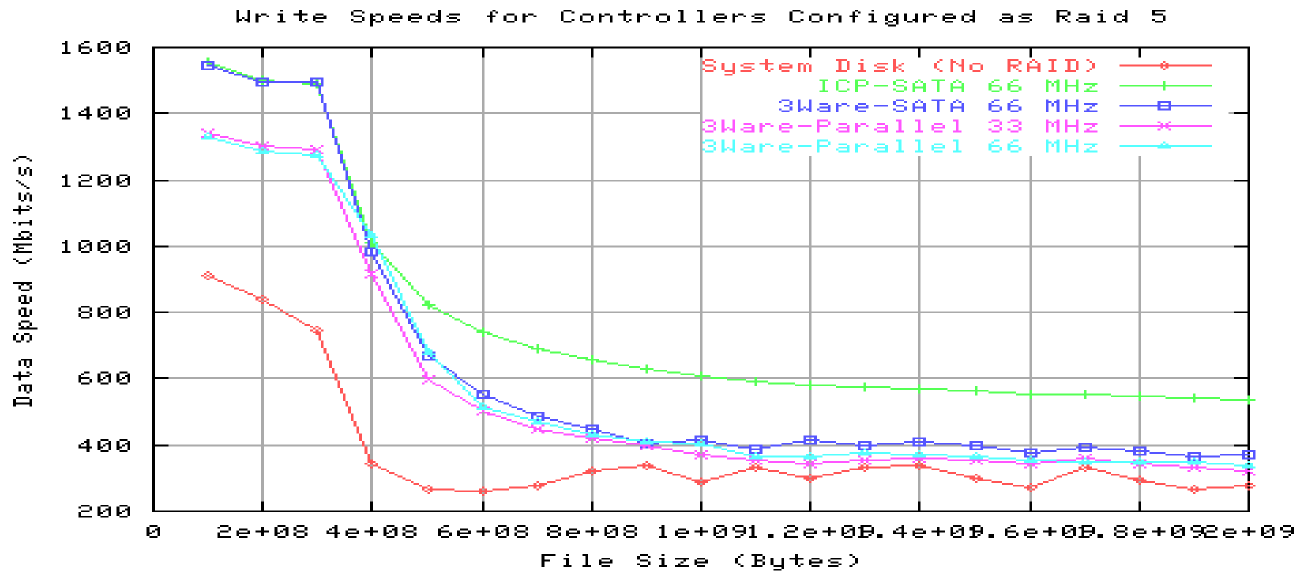
Raid studies

- RAID 0 known as disk striping.
  - Each disk has 64kbytes
  - Disks are in an array
  - The writing is sequential in a round-robin fashion.
- RAID 5 disk arrays give high read/write speeds together with built in redundancy to ensure fault tolerance
  - First disk 1^{st} block real data
  - 2^{nd} disk 1^{st} block compressed data (parity data)
  - 2^{nd} disk 2^{nd} block real data
  - 3^{rd} disk 1^{st} block compressed data (parity data) etc…

Performance of raid controllers: Disk read & write speeds

- 4 disk in the array (maxtor 160Gbytes 2Mbytes cache 7200 rpm) No network transfer.
- Improves asymptotically with increasing number of disks.
- Max read speed ~ 1300 Mbits/s
- Write speed (large files) ~ 600 Mbits/s





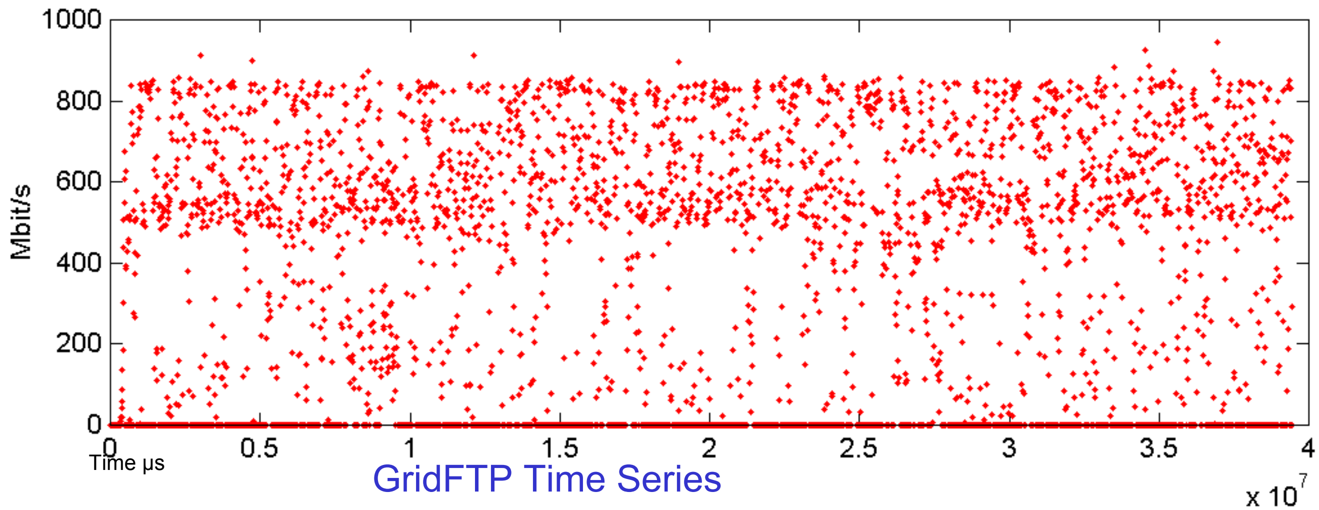
GridFTP vs APACHE disk-to-disk (web100)

- Raid0 with 4 disks. Transfer of 2GByte file.
- 3ware Serial ATA
- Web100 measurement at 10ms resolution
- Going through MB-NG network London to Manchester

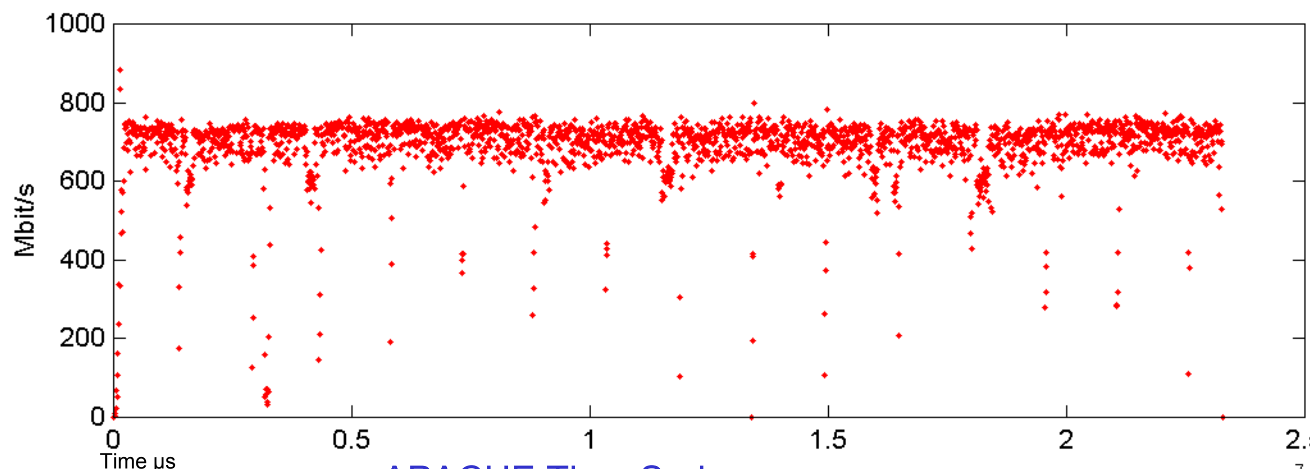
- GridFTP Average of 520 Mbit/s

- APACHE Average of 710 Mbit/s

- Actual disk rate will be lower. At this point data is memory.



GridFTP Time Series

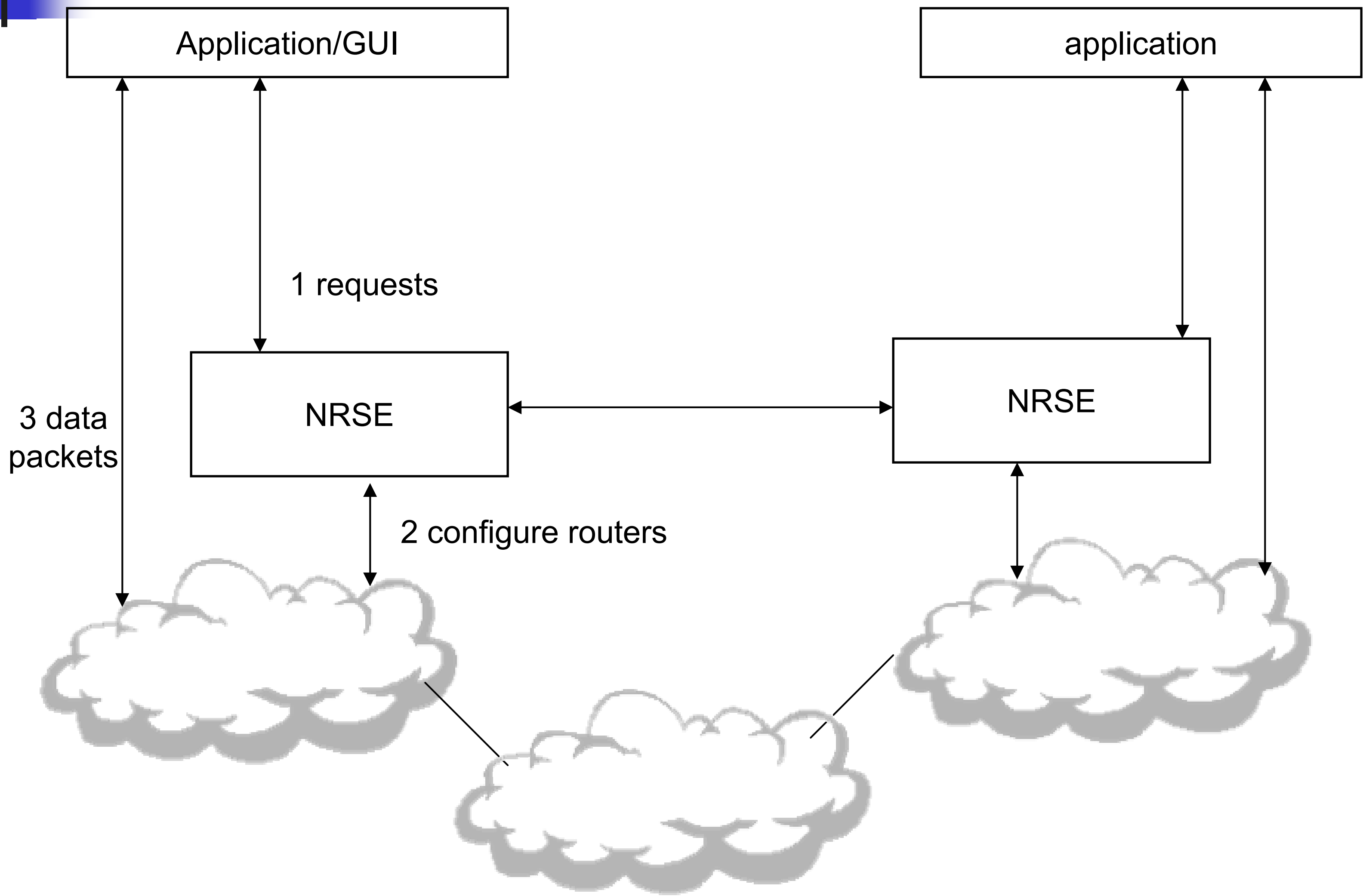


APACHE Time Series

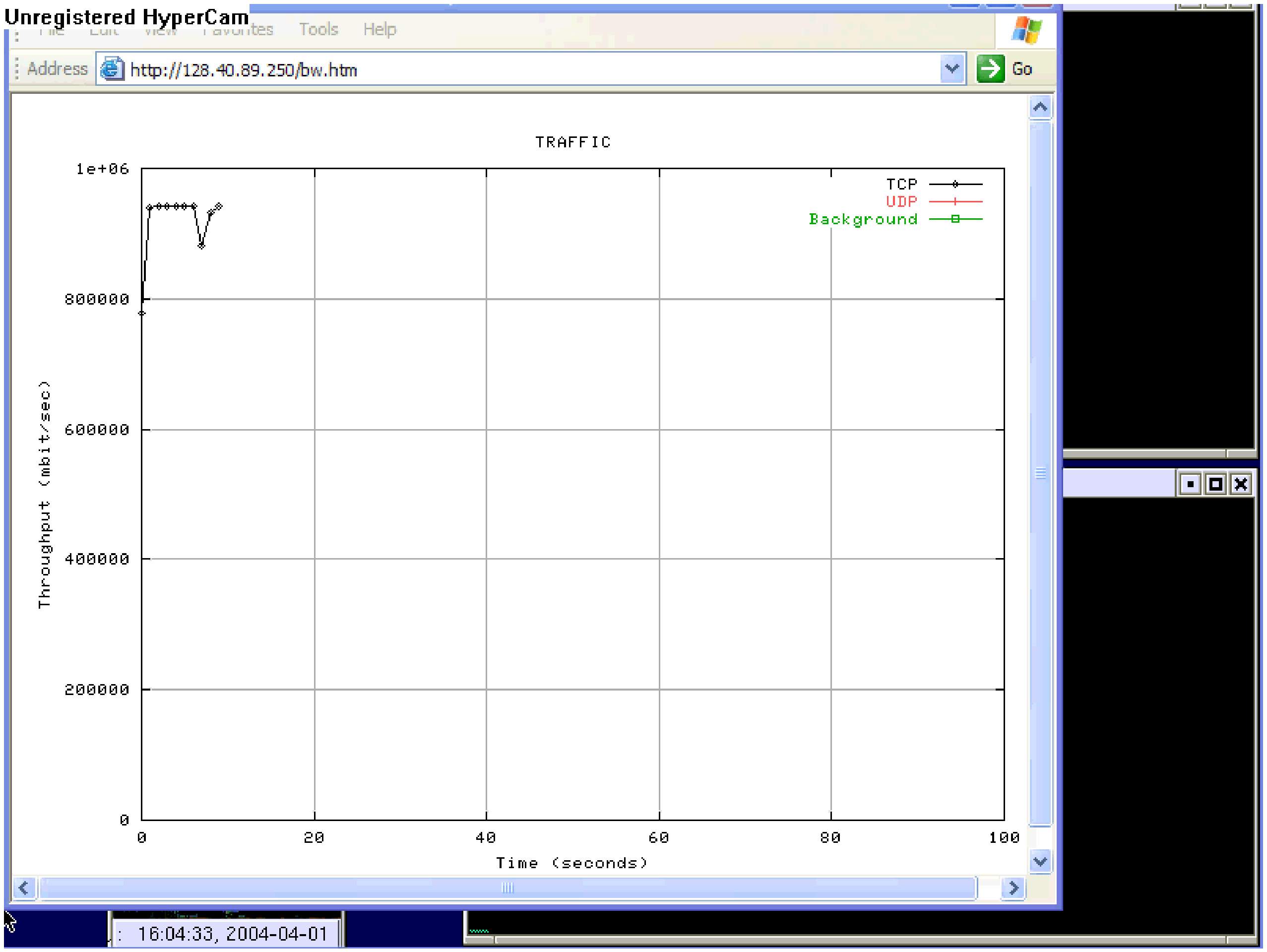
Middleware

- Started off looking at two middleware solutions:
  - GRS (Grid Resource Scheduling)
    - http://www.cs.ucl.ac.uk/research/grs/
  - GARA (General-purpose Architecture for Reservation and Allocation)
    - GARA provides **end-to-end QoS** to the applications using three types of Resources Managers (RM), in our case, we just make use of the Network RM (Differentiated Services). It allows **immediate and advance** reservations.
- So far, GRS has been deployed in MB-NG.

GRS Architecture



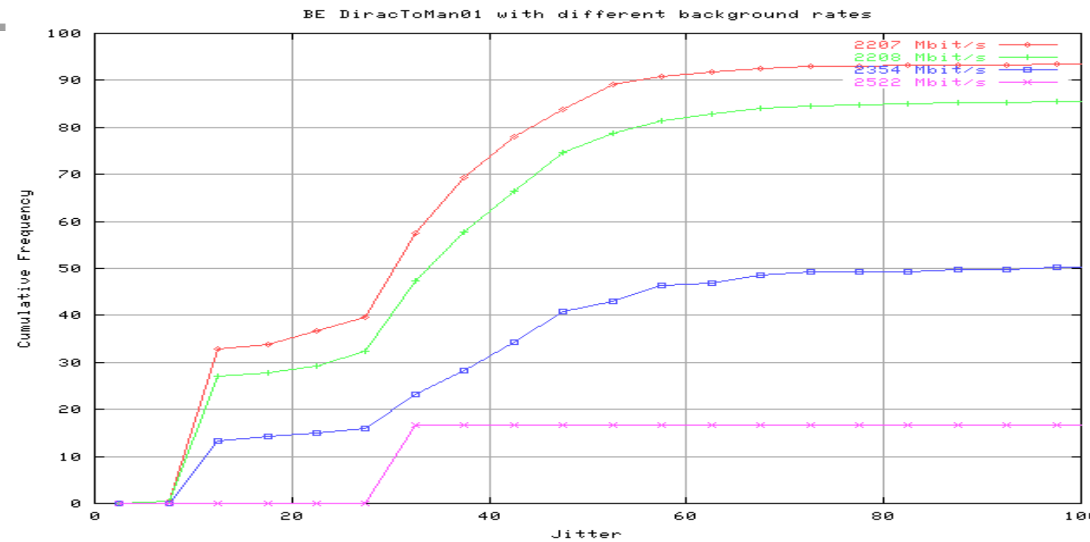
GRS Demo

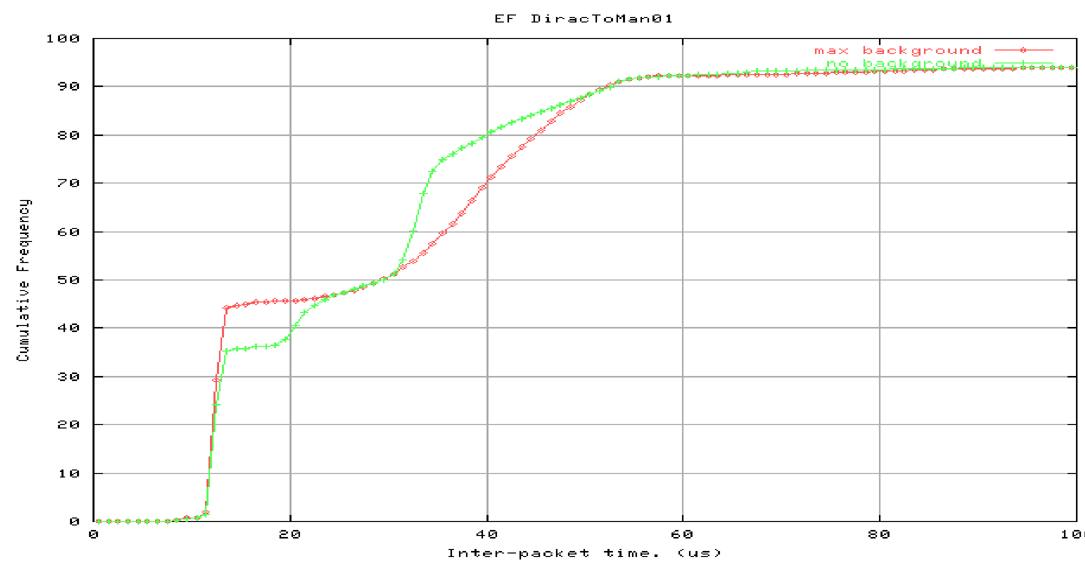


Unregistered HyperCam
Tools
Help
Address
http://128.40.89.250/bw.htm
Go
TRAFFIC
TCP
UDP
Background
1e+06
800000
600000
400000
200000
0
Throughput (mbit/sec)
0
20
40
60
80
100
Time (seconds)
16:04:33, 2004-04-01

Application performance: Collaboration with Reality Grid

- Application was VizServer demo “Smoke and Mirrors”: **Real-time remote visualisation.**
- Processing in London, visualised in Manchester
- Performance with various background rates
- Performace with QoS and maximum background traffic.





Future Activities

- End-to-end reservation removing over-provisioned core requirement
- Managing MPLS tunnels with GRS
- Web-services interface for GRS
- End-to-End network services to Europe and US

Achievements

MB-NG: Managed Bandwidth - Next Generation

- An end-to-end multi-domain technology trial, bringing together:
  - Multi-gigabit network
  - Multi-domain aspects
  - Protocols for high bandwidth transfers
  - Middleware…QoS to the user

Thank You…