DiskRouter: A Mechanism for High Performance Large Scale Data Transfers

George Kola
Computer Sciences Department
University of Wisconsin-Madison
kola@cs.wisc.edu
http://www.cs.wisc.edu/condor

Outline

- Problem
- DiskRouter Overview
- Details
- Real life DiskRouters
- Experiments

Problem

SDSC to NCSA
Bottleneck Bandwidth : 12.5 MBPS
Latency 67 ms

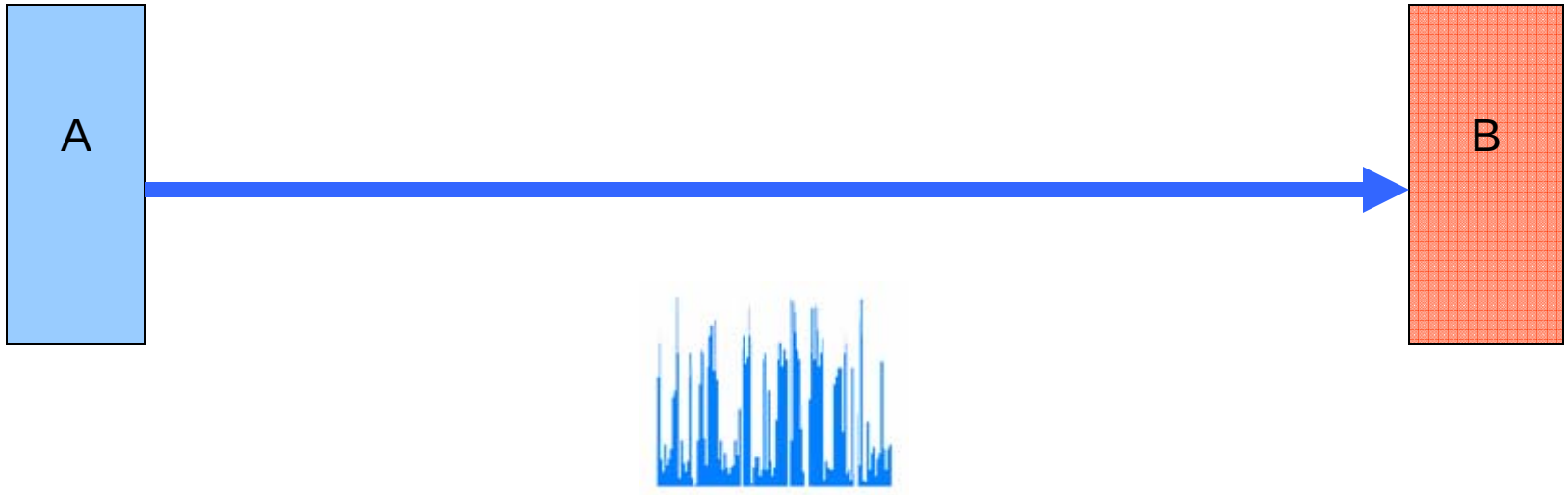
Transfer Rate got by applications for a 1GB file

Scp : 0.66MBPS
GridFTP(1 stream) : 0.85 MBPS
GridFTP(10 streams) : 3.52 MBPS

DiskRouter Overview

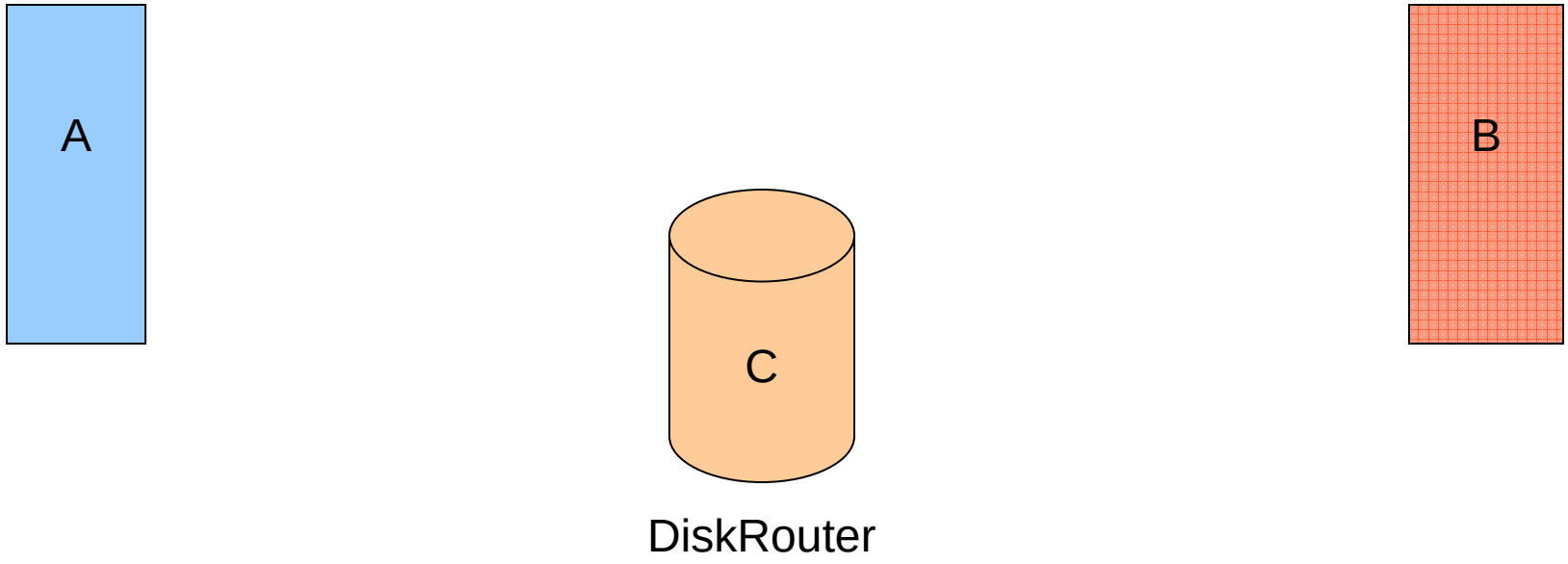
- Mechanism to efficiently move large amounts of data (order of terabytes)
- Uses disk as a buffer to aid in large scale data transfers
- Application-level overlay network used for routing
- Ability to use higher level knowledge for data movement

A Simple Case



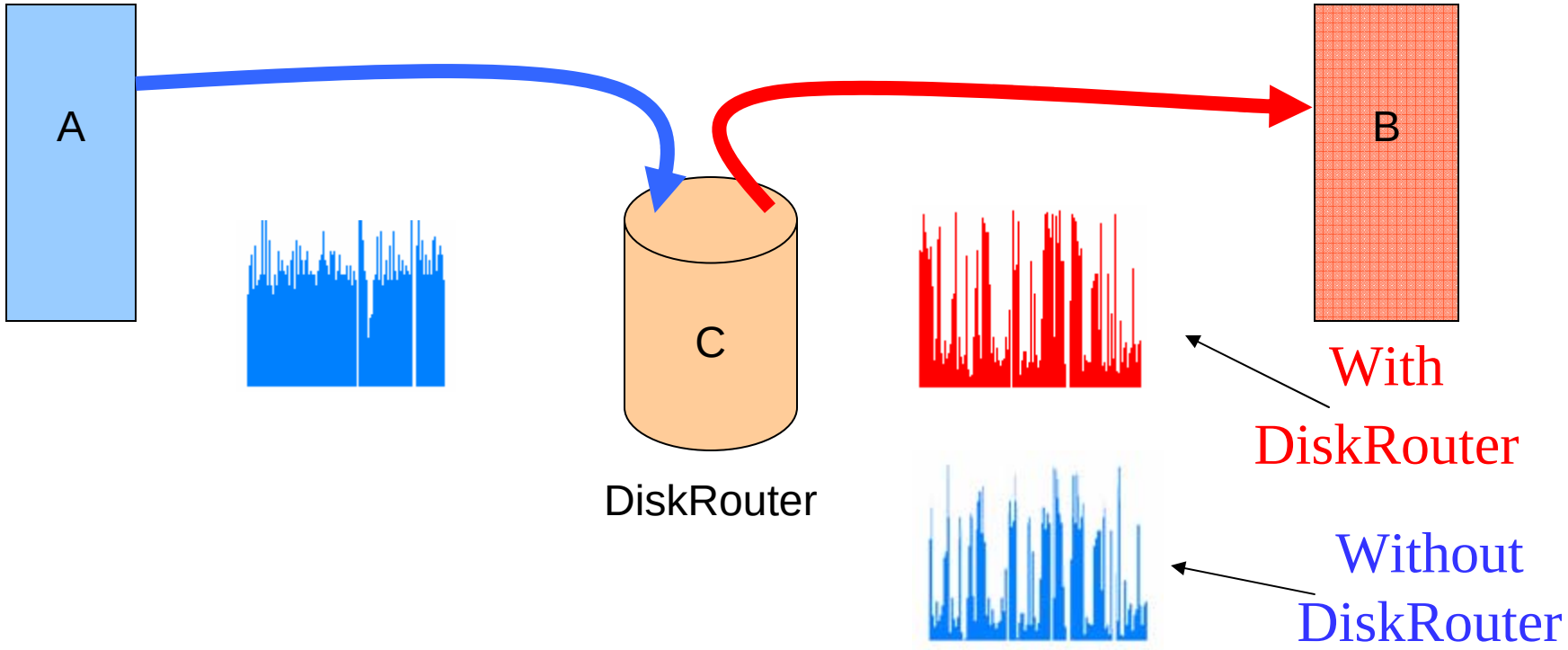
A is transferring a large amount of data to B

A Simple Case



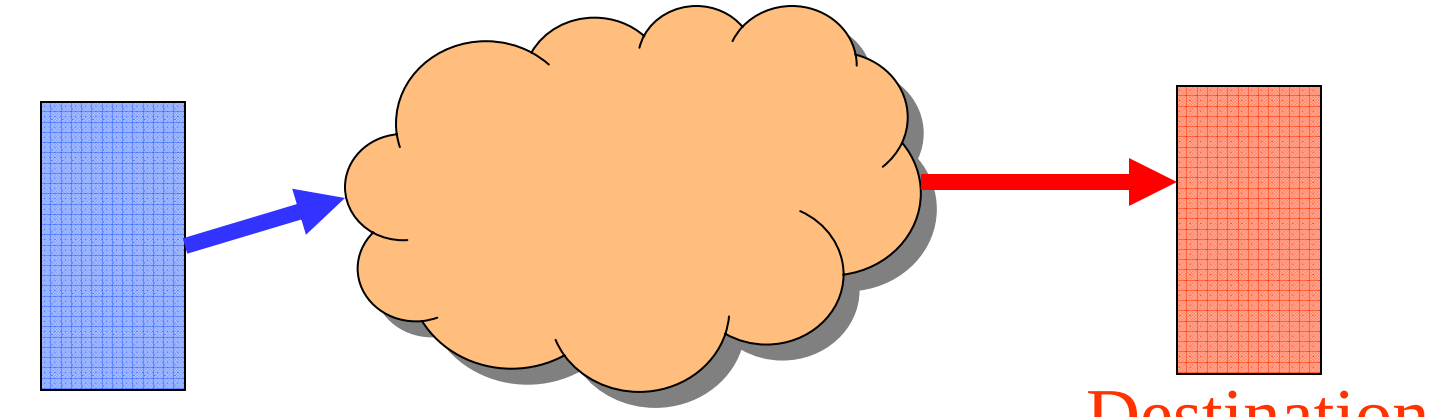
C is an intermediate node between A and B

A Simple Case with DiskRouter



Improves performance when bandwidth fluctuation between **A and C** is independent of the bandwidth fluctuation between **C and B**

Data Mover/Distributed Cache

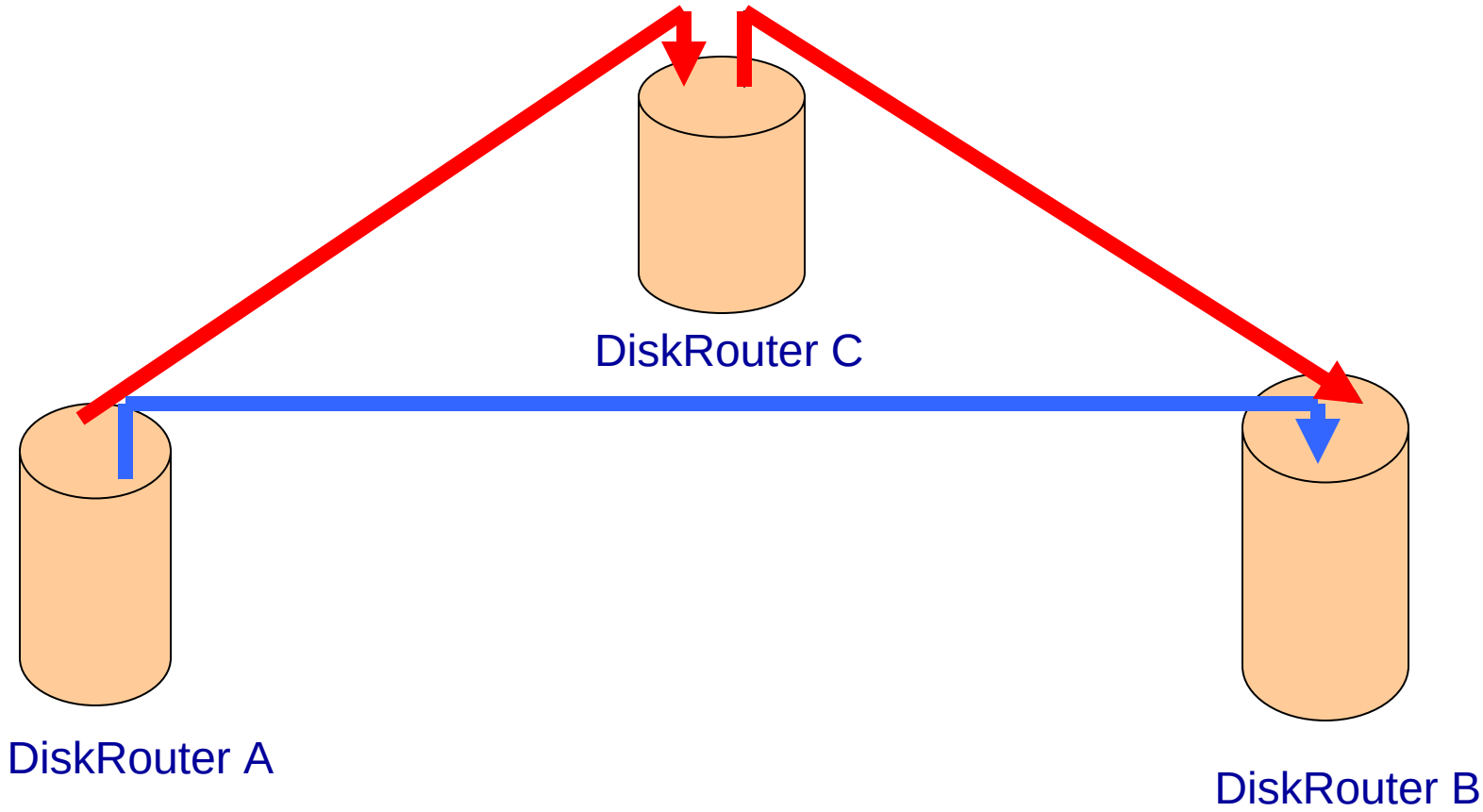


Source writes to the closest DiskRouter and Destination receives it up from its closest DiskRouter

Outline

- Problem
- DiskRouter Overview
- Details
- Real life DiskRouters
- Experiments

Routing Between DiskRouters



C need not be in the path between A and B

Network Monitoring

- Uses 'Pathrate' for estimating network capacity
- Performs actual transfers for measurement
- Logging the data rate seen by different components
- Generate network interface stats on the machines involved in the transfers

Implementation Details

- Uses multiple sockets and explicitly sets TCP buffer sizes
- Overlaps disk I/O and socket I/O

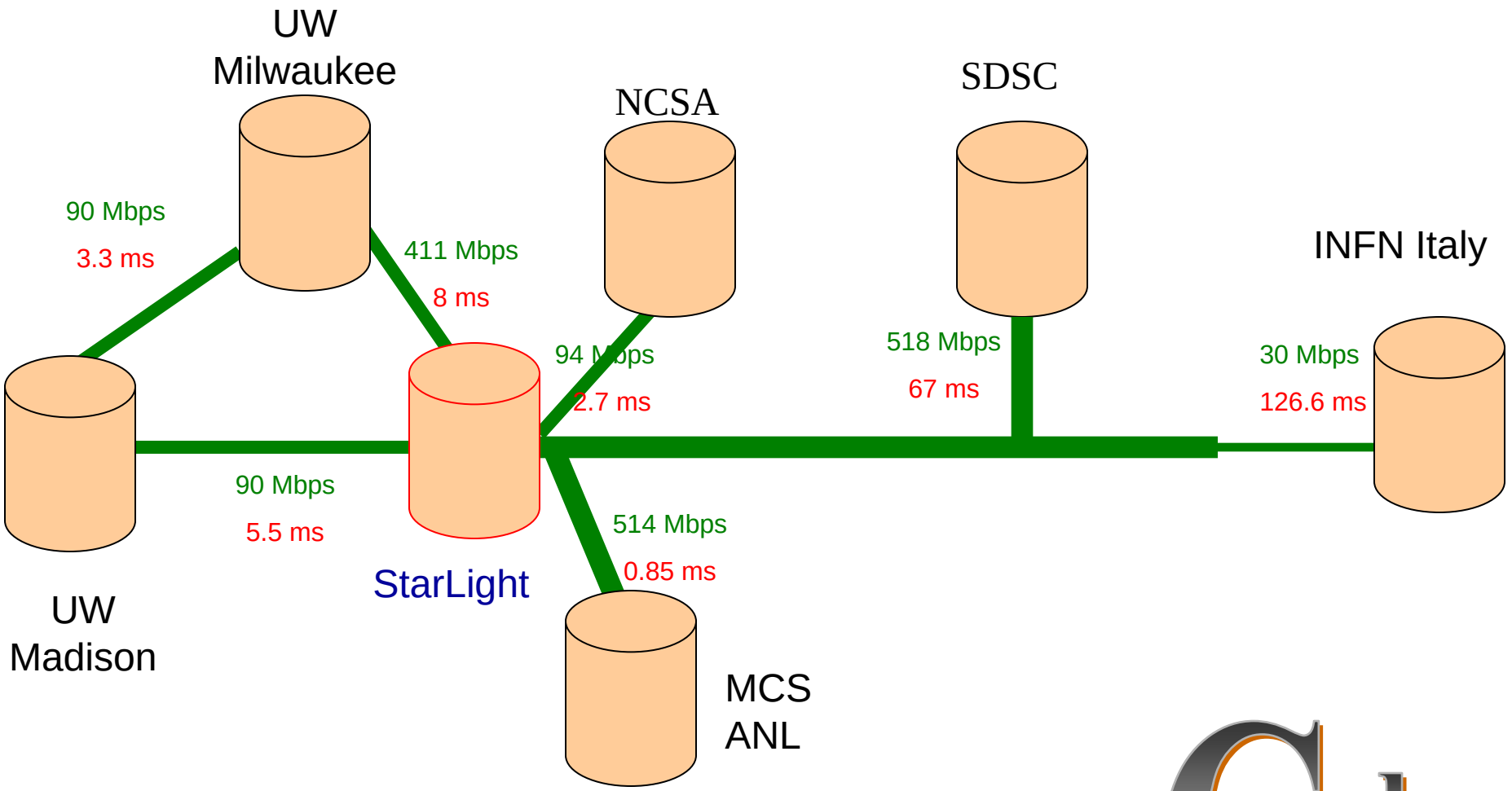
Client Side

- Client library provided
- Applications can call library functions for network I/O
- Functions provided for common case file transfer (overlaps network I/O and disk I/O)
- Third party transfer support

Outline

- Problem
- DiskRouter Overview
- Details
- Real life DiskRouters
- Experiments

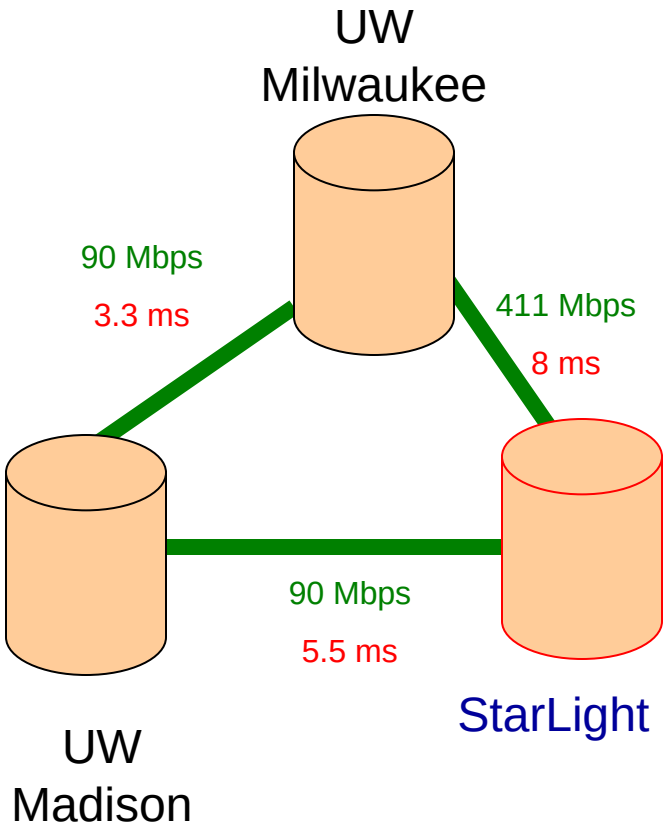
Real Life DiskRouters



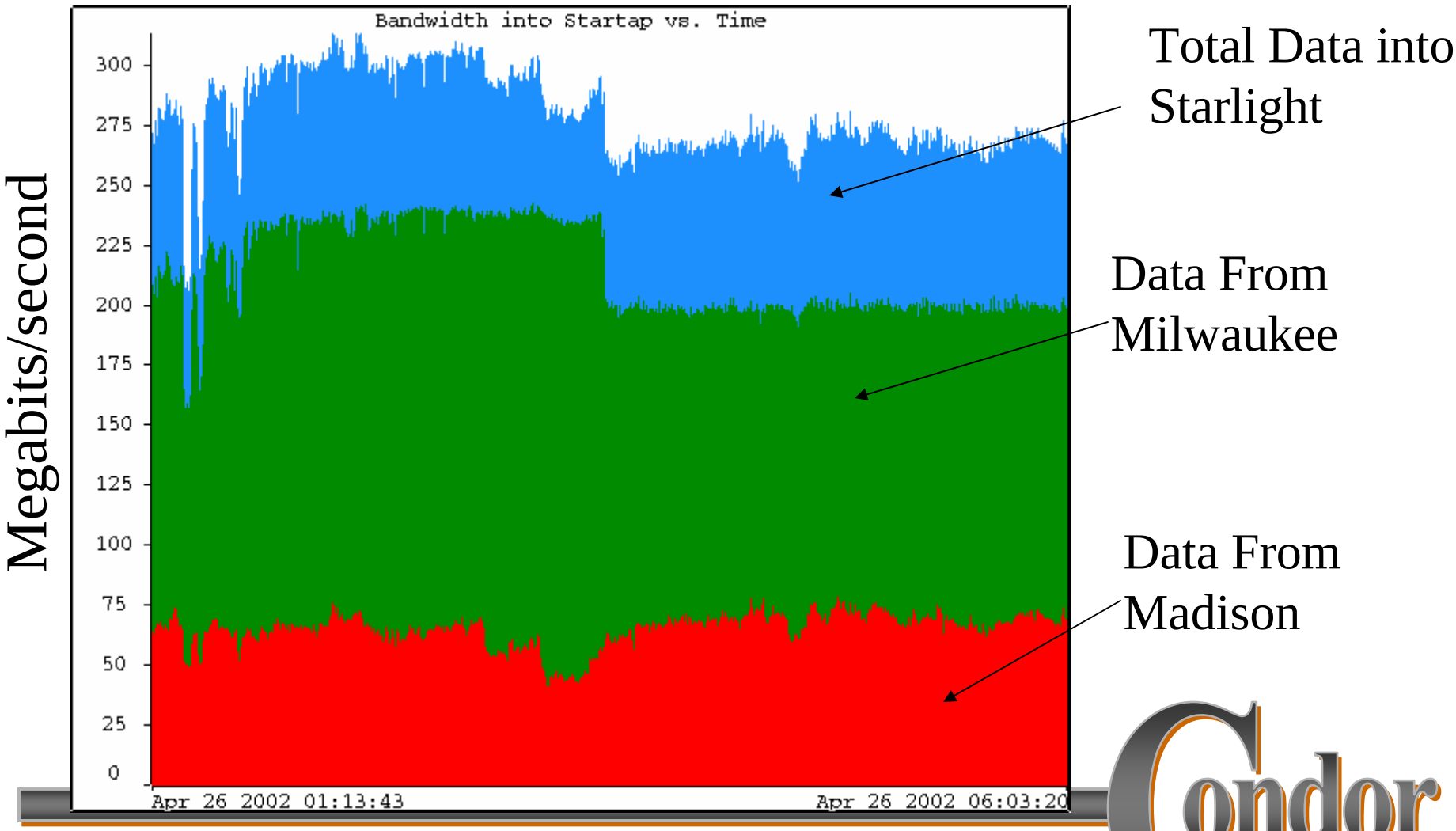
Outline

- Overview
- Details
- Real Life DiskRouters
- Experiments

Testing Multiroute



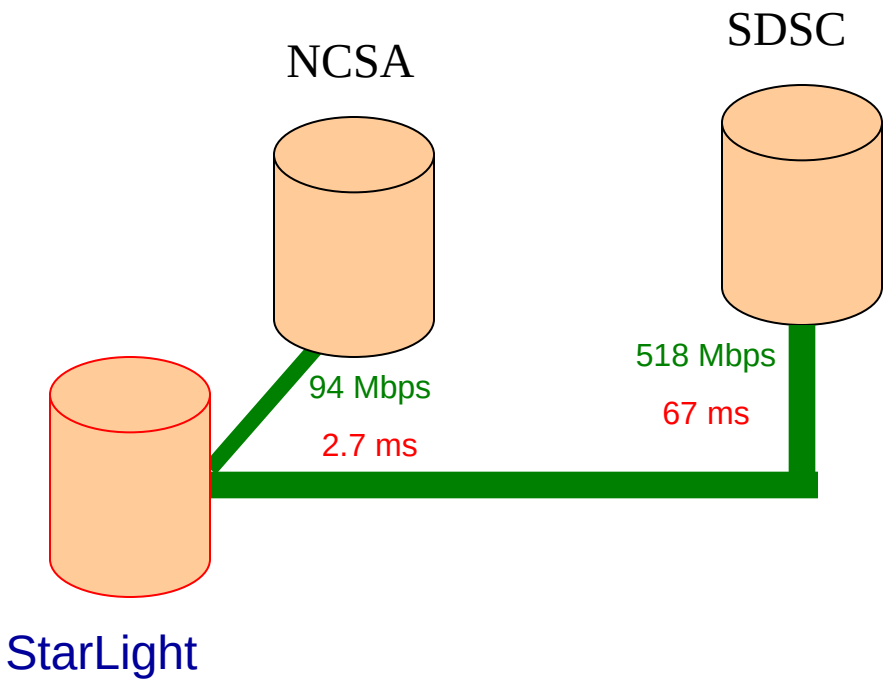
Multiroute Improves Performance



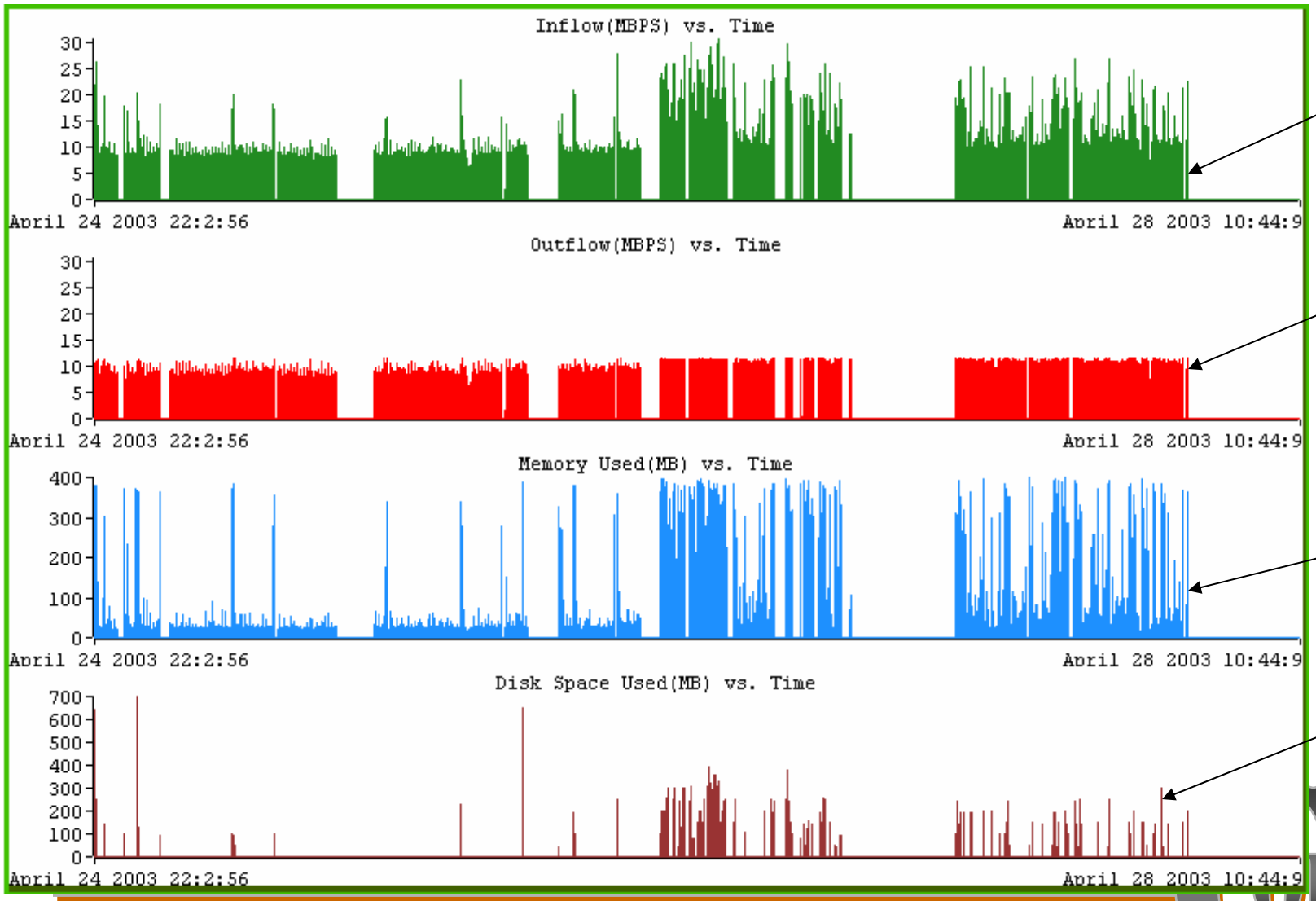
SRB to Unitree Transfer Using Stork

- Data movement from SDSC to NCSA via Starlight (3 TB of data had to be moved)
- Integrated into Stork
- Found significant performance gain

Link between SDSC and NCSA



Starlight DiskRouter Stats



Data Inflow

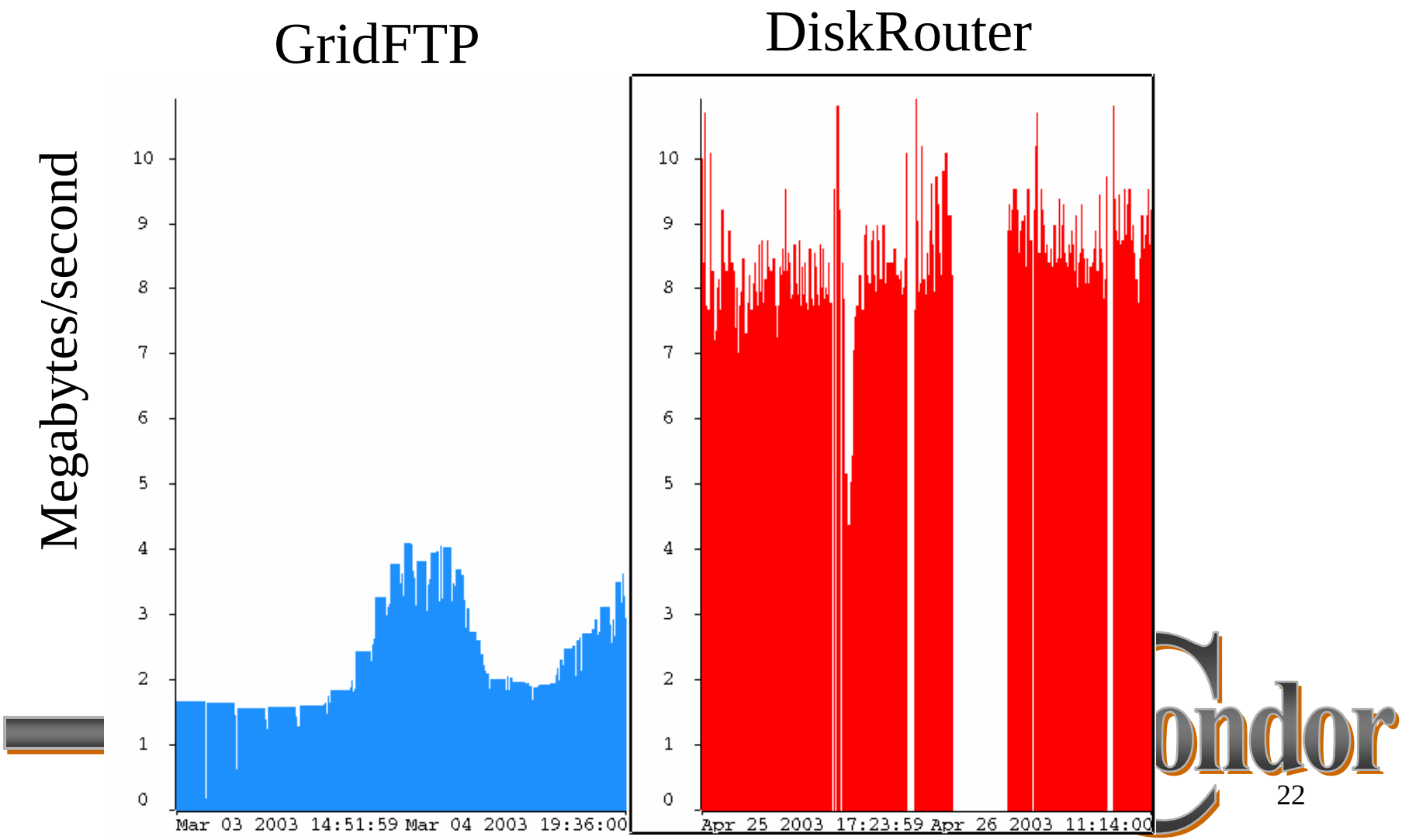
Data Outflow

Memory Used

Disk Used

GridFTP vs DiskRouter

End-to-End Data Rate Seen by Stork(MBPS) vs. Time



A Glimpse of Performance

Transfer of 1 GB file from SDSC (SanDiego) to NCSA (Urbana-Champaign)

Tool	Transfer Rate
Scp	0.66 MBPS
GridFTP(1 stream)	0.85 MBPS
GridFTP(10 streams)	3.52 MBPS
DiskRouter	10.77 MBPS

Work In Progress

- Computation on data streams in the DiskRouter
- Ability to perform computation in the nodes attached locally to the DiskRouter
- Working together with Stork to add intelligence to data movement

Questions

- Thanks for listening