
Transport Layer Communication Protocols for Grid Computing

Miguel A. Labrador

Assistant Professor

University of South Florida

Department of Computer Science and Engineering

labrador@csee.usf.edu

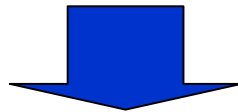
<http://www.csee.usf.edu/~labrador>

Outline

- Why Grid computing?
- What Do we need for Grid computing?
- Do we really have or will have very fast and reliable networks?
- Will we really be able to transfer data at the required speeds?
- TCP and TCP in high bandwidth delay product networks
- Current research

Why Grid computing?

- Foreseen applications play a major role
 - Scientists expect to collaborate manipulating and sharing petabytes (10^{15}) of data [1,2]
- Computing power, network speeds, and storage capacity are expected to double at an average period of 18, 9, and 12 months [1]
- Advances in storage will allow us to build petabyte archives
- Optical communication networks will transfer data at terabits per second
- Improvements in computer power won't keep up with advances in storage and communications
 - Won't be able to compute what we need at a single location



Grid Computing

What Do we need for Grid computing?

- Need to use many resources, in many different places, with many different access policies, hardware and software, etc.
- Scalable, secure, high performance mechanisms to discover and negotiate access to remote resources
 - This is what this workshop is mainly about
- Also need very fast and reliable communication networks so applications running elsewhere look like running locally
 - Grid computing is based on the assumption that communication is free
 - This is what I am going to talk about

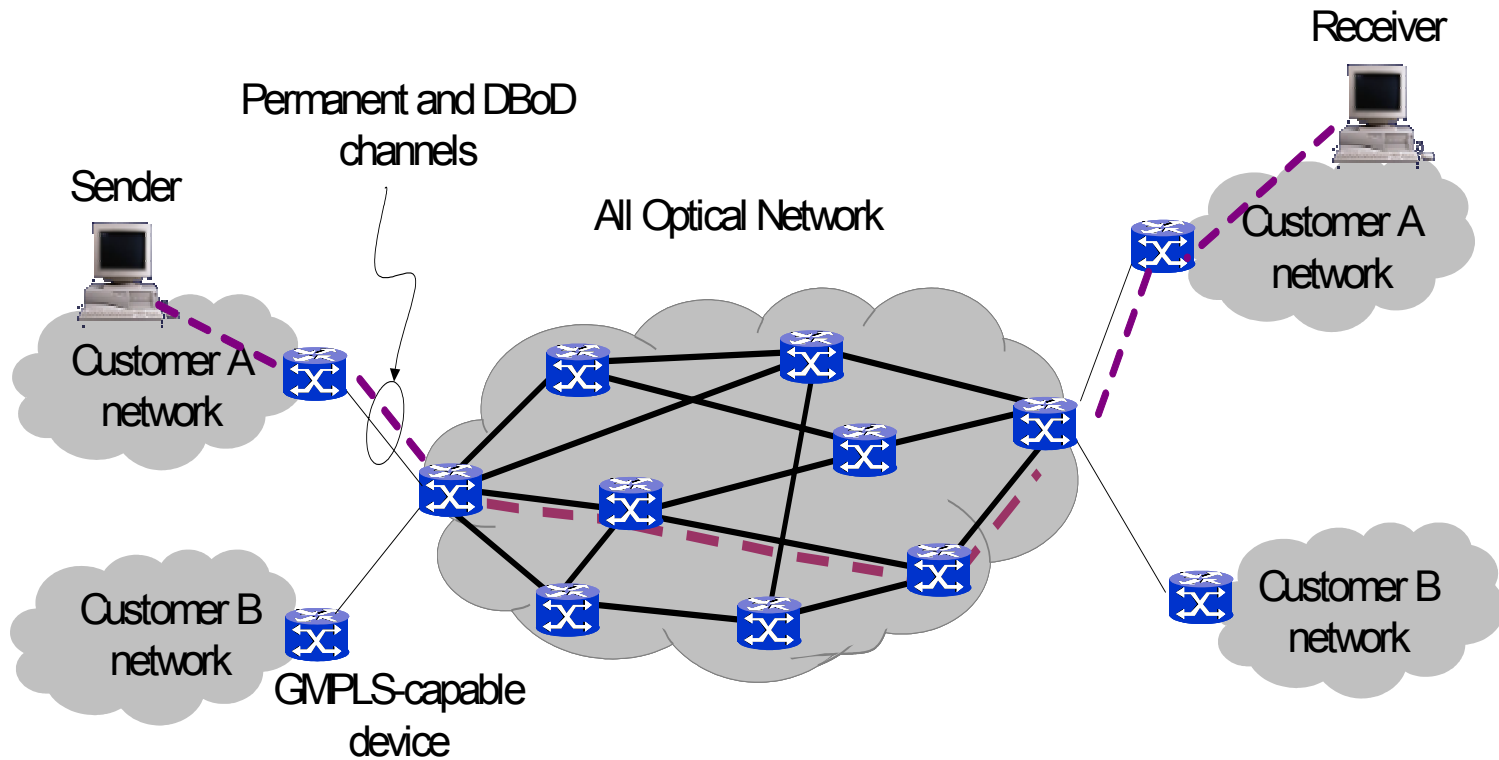
Do we really have or will have those very fast and reliable networks?

Will we really be able to transfer data at these speeds?

Do we really have or will have those very fast and reliable networks?

- Yes! We have gone from 56 Kbps in the 80's, to 155 Mbps in the 90's to 2.5 Gbps now, to 100's of Gbps and Tbps in the near future
- Next generation optical networks will make users available lambdas on demand capable of transferring Gbps
 - A very hot area of research
 - Huge savings potential because infrastructure will be simplified considerably
 - New important services, such as fast provisioning
 - Opaque, hybrid and all optical switches
 - Optical burst and packet switching
 - Control plane signaling protocols
 - GMPLS
 - Network management
 - Many other important aspects such as survivability schemes, virtual topology design and dynamic reconfiguration, new routing and wavelength assignment algorithms, etc.

Next generation optical networks



Will we really be able to transfer data at the required speeds?

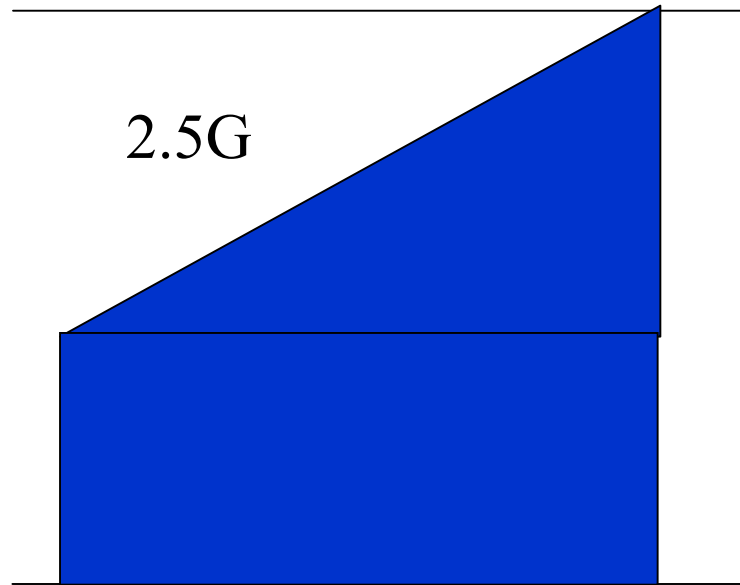
- Optical networks will provide the pipes
 - However, this is just the raw transmission speed
- Communication protocols are in charge of the information exchange and control the final rate at which data is sent over these high-speed pipes
- Between the application and the physical pipe there are several communication protocols that actually make the data transfer possible
- Nowadays everything is based on TCP-UDP/IP
 - And it will continue for a while
- So, can TCP transfer data in this environment?
- Is TCP suitable for this new environment?
- Do we have to continue using a flow and congestion control mechanism based on TCP's congestion window or we need a new paradigm?

TCP

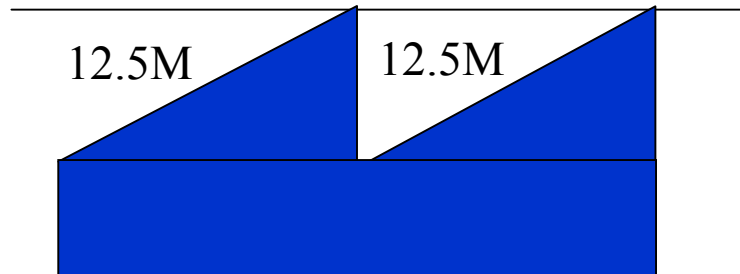
- This problem is being investigated as we speak under different names
 - TCP over High Bandwidth-Delay Product Networks or TCP over all optical channels
- The question is the same
 - How does TCP performs when called to run over a very long and fat pipe?
- TCP will not work well in this new environment
 - Congestion window will take a very long time to reach maximum available capacity during the congestion avoidance phase
 - In CA one packet per RTT!
 - If packet loss occurs, congestion window is reduced by half!
 - TCP stays in CA despite the materialization of new bandwidth
 - During slow start many packets can be dropped
 - Very large congestion window
 - One TCP connection might not be able to use the whole available bandwidth and more importantly, sustain the data rates needed

TCP

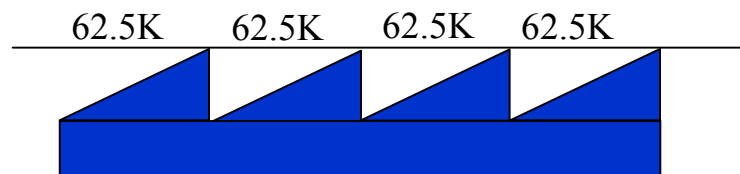
10000 Mbps



100 Mbps



1 Mbps



TCP

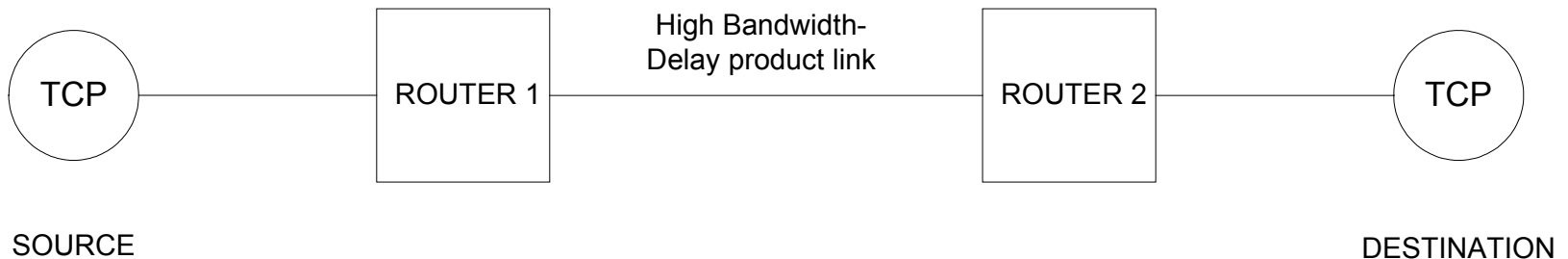
- With RTT=108ms and 1000-byte pkts, filling a 1 Gbps pipe corresponds to a congestion window of 13500 pkts
 - 13500 RTTs or roughly half an hour to reach congestion window value that TCP had before the CA phase was initiated
 - Increases to 4 hrs if link is 10Gbps
- TCP's response function places an upper bound on achievable congestion windows, given some underlying packet drop rate

$$T = \frac{1.2}{\sqrt{p}} \quad W = \frac{\sqrt{1.5}}{\sqrt{p}}$$

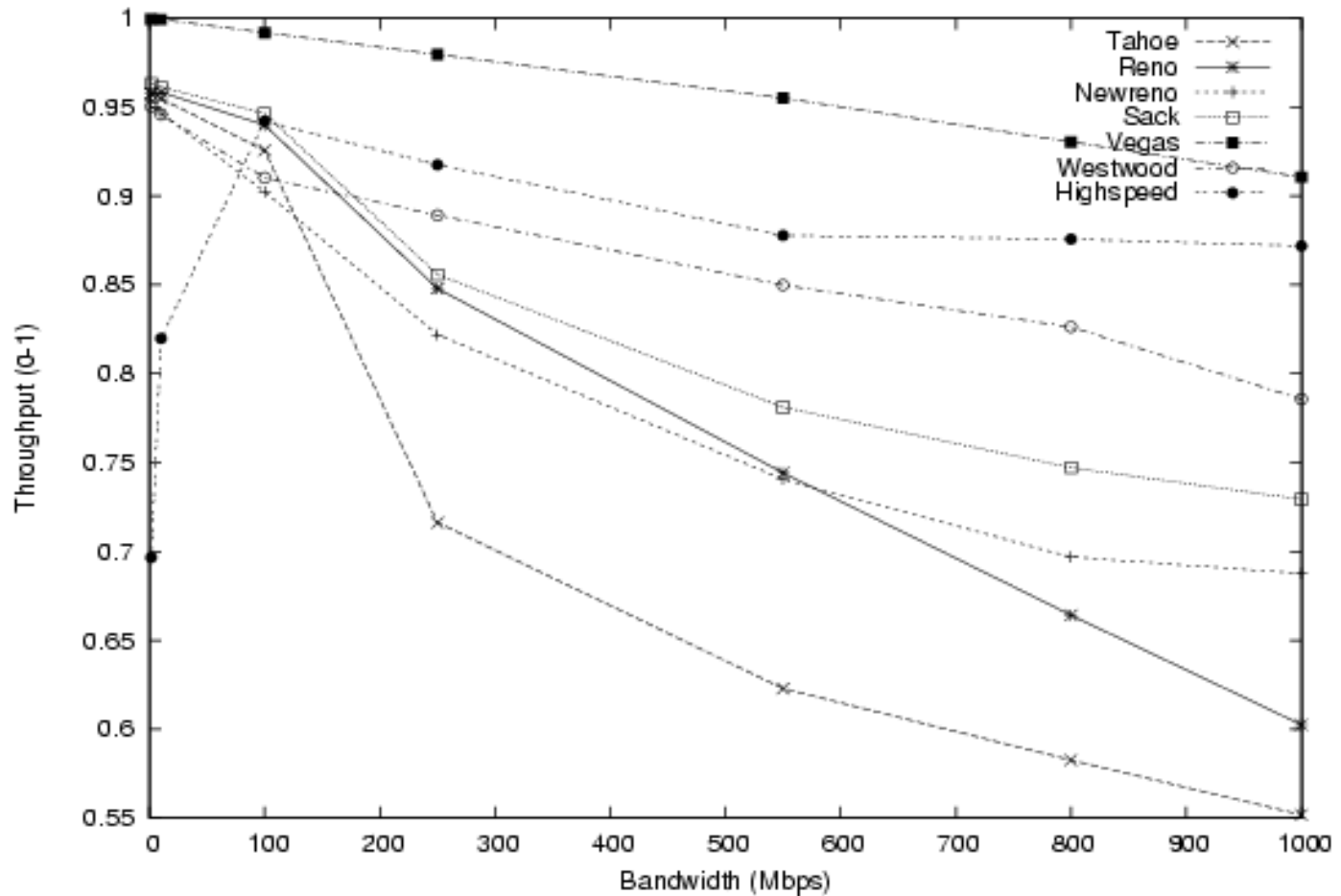
- A TCP connection with 1500-byte packets and a 100 msec RTT would require a CongWind=83333 pkts to fill a 10 Gbps pipe and a packet drop rate of at most one drop every 5,000,000,000 pkts
 - This passes the limits of achievable fiber error rates

TCP in High Bandwidth-Delay Product Networks

- Typical dumbbell scenario in NS-2
 - One TCP source, one bottleneck link at variable speed and Round Trip Time propagation delay of 25 ms
 - Queue size=200 pkts, pkt size=1000 bytes



Simulation results [7]



TCP is NOT Scalable!

Current research

- TCP over high speed links is a very recent topic of research
 - Several new TCP versions have been proposed to address the problems
 - HighSpeed TCP [3], FAST TCP[4], Scalable TCP [5], Enhanced TCP [6]
 - More to come
- All these schemes are still based on TCP's congestion window mechanism
 - Main problem is that TCP is “blind” and these protocols will never ever achieve full channel utilization
 - FAST utilizes queueing delay
- Need a new paradigm
 - Bandwidth estimation techniques might be part of the solution
 - TCP Vegas and Westwood
 - Our simulations showed that these protocols perform better than HighSpeed TCP
 - Other schemes are under current investigation
- Not a solved problem yet!

References

- [1] I. Foster, “The Grid: A New Infrastructure for the 21st. Century Science,” Physics Today, February 2002. Available at <http://www.aip.org/pt/vol-55/iss-2/p42.html>
- [2] First International Workshop on Protocols for Fast Long-Distance Networks (PFLDnet) 2003
- [3] S. Floyd, HighSpeed TCP (HSTCP) web page, <http://www.icir.org/floyd/hstcp.html>
- [4] S. Low, FAST Protocols for Ultrascale Networks, <http://netlab.caltech.edu/FAST/>
- <http://datatag.web.cern.ch/datatag/pfldnet2003/>
- [5] T. Kelly, “Scalable TCP:Improving Performance in Highspeed Wide Area Networks, in Proceedings of PFLDnet, 2003
- [6] A. Kamra, V. Misra, and D. Towsley, “Achieving High Throughput in Low Multiplexed, High Bandwidth, High Delay Environments,” in Proceedings of PFLDnet, 2003
- [7] X. Jianxuan, S. Kerkar, M. A. Labrador and Mohsen Guizani, “Performance Evaluation of TCP over Optical Links”, to appear in Proceedings of IEEE ICC 2004.

Thanks!