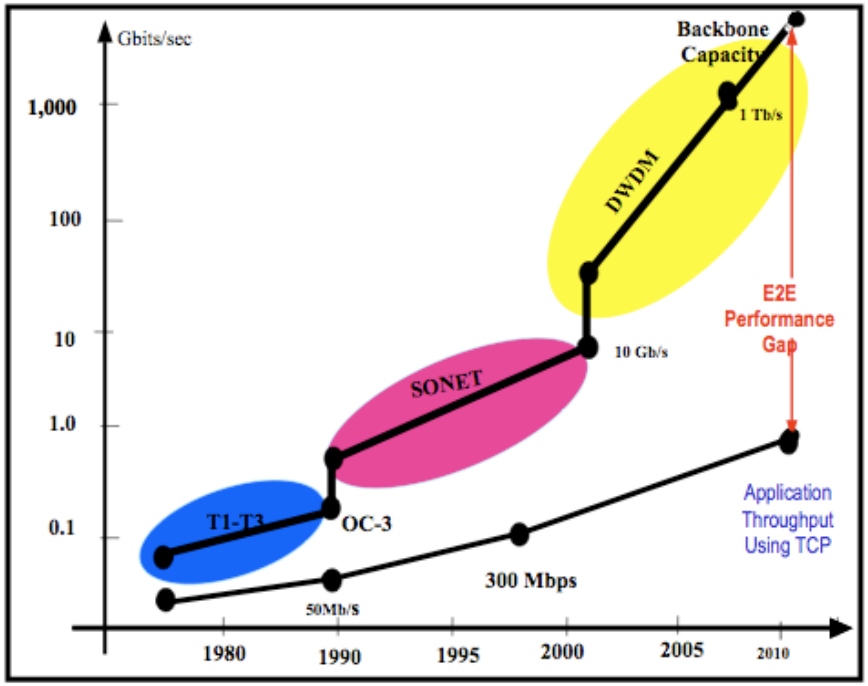
Network Middleware for High-Performance Networking

Martin Swany

Introduction and Motivation

- Networks are increasingly critical for science and education
- Data Movement is a key problem
- Network speeds can increase dramatically but users' throughput increases much more slowly
  - Source: DOE



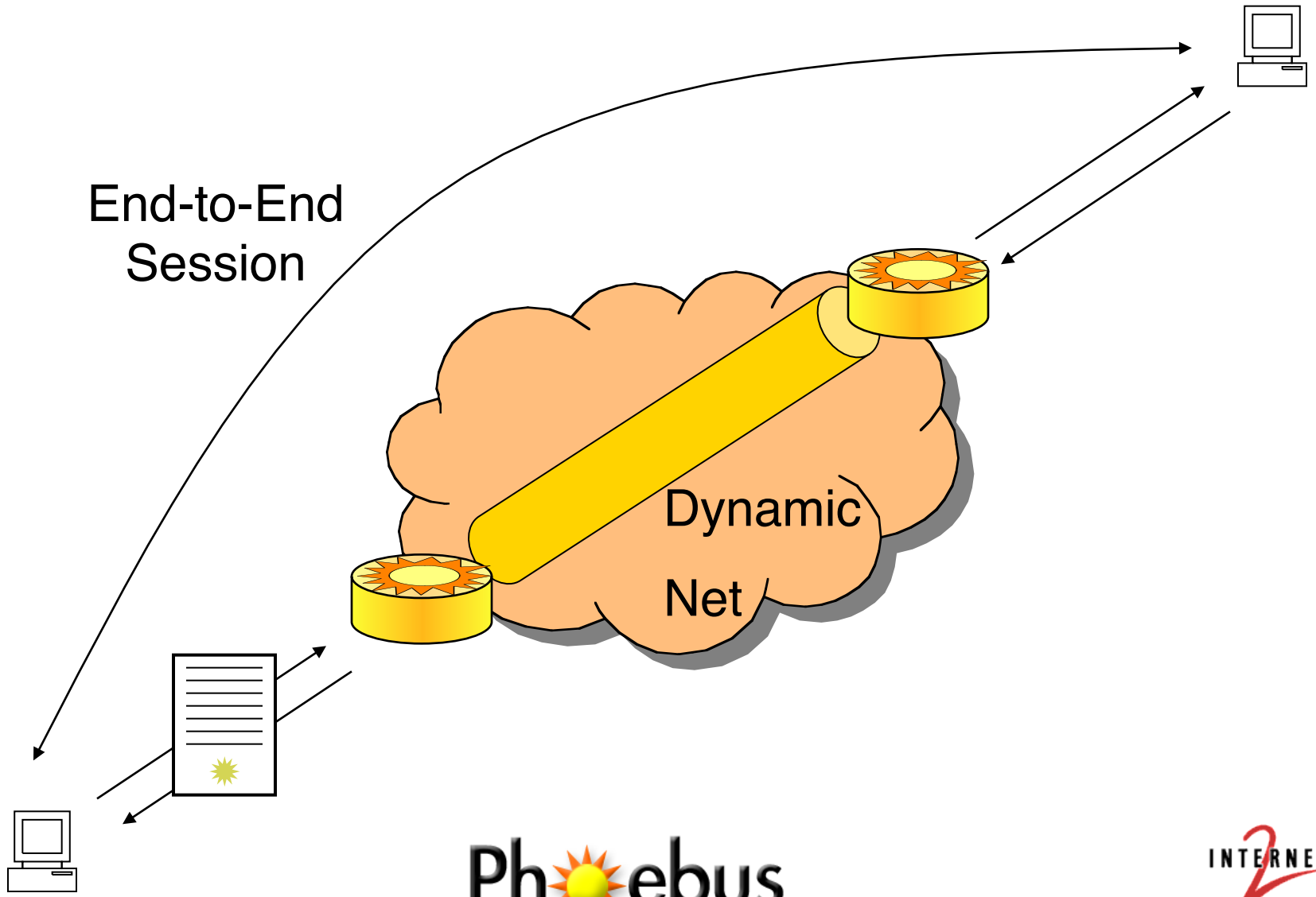
Phoebus

INTERNET2

Phoebus

- The Phoebus project aims to help bridge the performance gap by bringing revolutionary networks to users
  - Phoebus is another name for the mythical Apollo in his role as the “sun god”
- Phoebus is based on the concept of a “session” that enables multiple adaptation points in the network to be composed
- Phoebus provides a gateway for legacy applications to use advanced networks
- Open source WAN accelerator

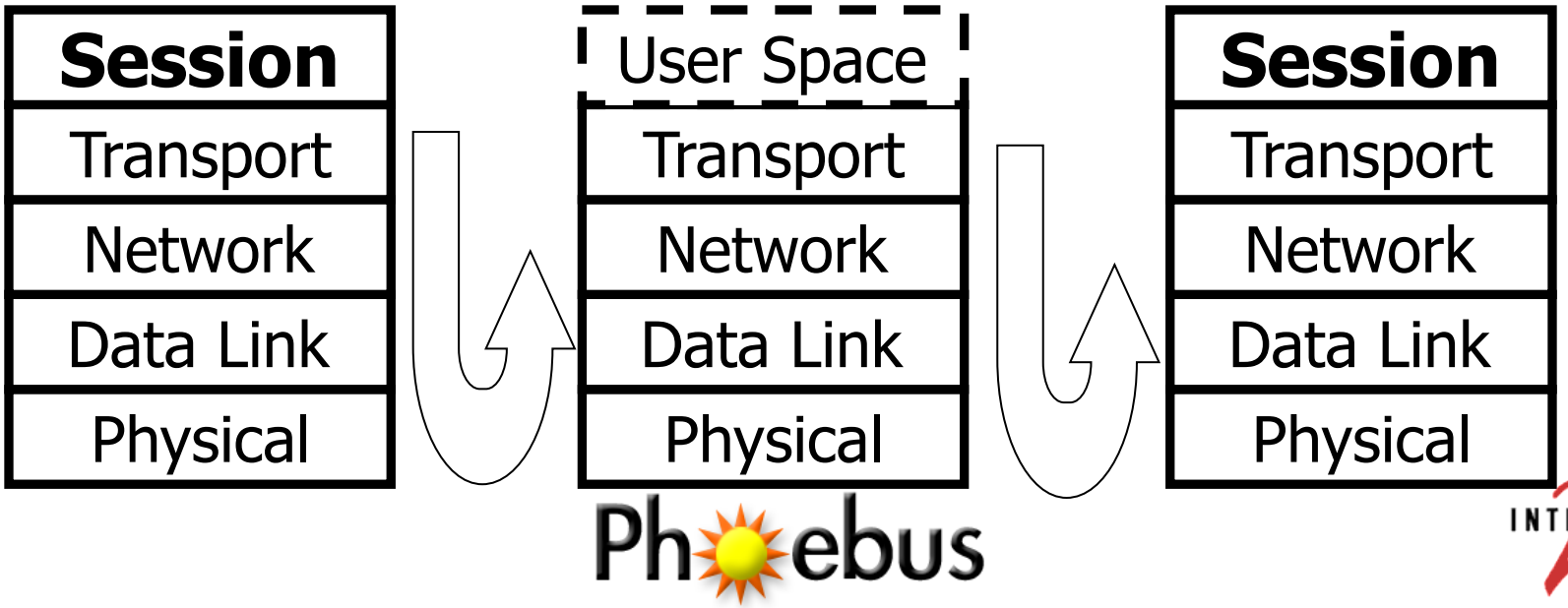
Phoebus



Phoebus

Session Layer

- A *session* is the end-to-end composition of *segment-specific* transports and signaling
  - More responsive control loop via reduction of signaling latency
  - Adapt to local conditions with greater specificity
  - Buffering in the network means retransmissions need not come from the source



Session Layer Benefits

- A session layer provides explicit control over ***adaptation points*** in the network
- Transport protocol
  - Rate-based to congestion based
  - Shorter feedback loops
- Traffic engineering
  - Map between provider-specific DiffServ Code Points / VLANs
- Authorization and Authentication
  - Rich expression of policy via e.g. the Security Assertion Markup Language (SAML)

Phoebus

INTERNET2

Phoebus Signaling

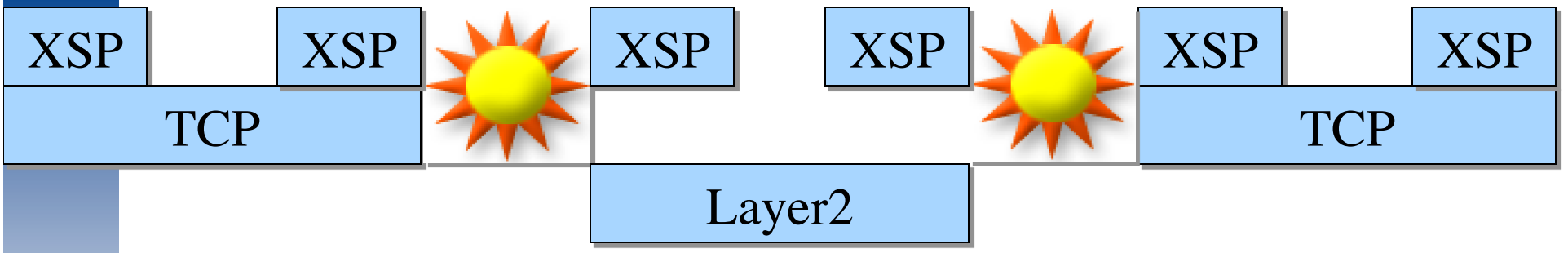
- Phoebus speaks to the control plane to provision network resources
  - Can allocate circuits from the OSCARS IDC
    - Which underlies ION
  - Also, direct communication with DRAGON
- Once the connection is established to the Phoebus node, traffic can begin to flow
  - Could be sent over an existing link if unable to provision
- Phoebus can finish the connection over the commodity network if the allocation times out

Phoebus

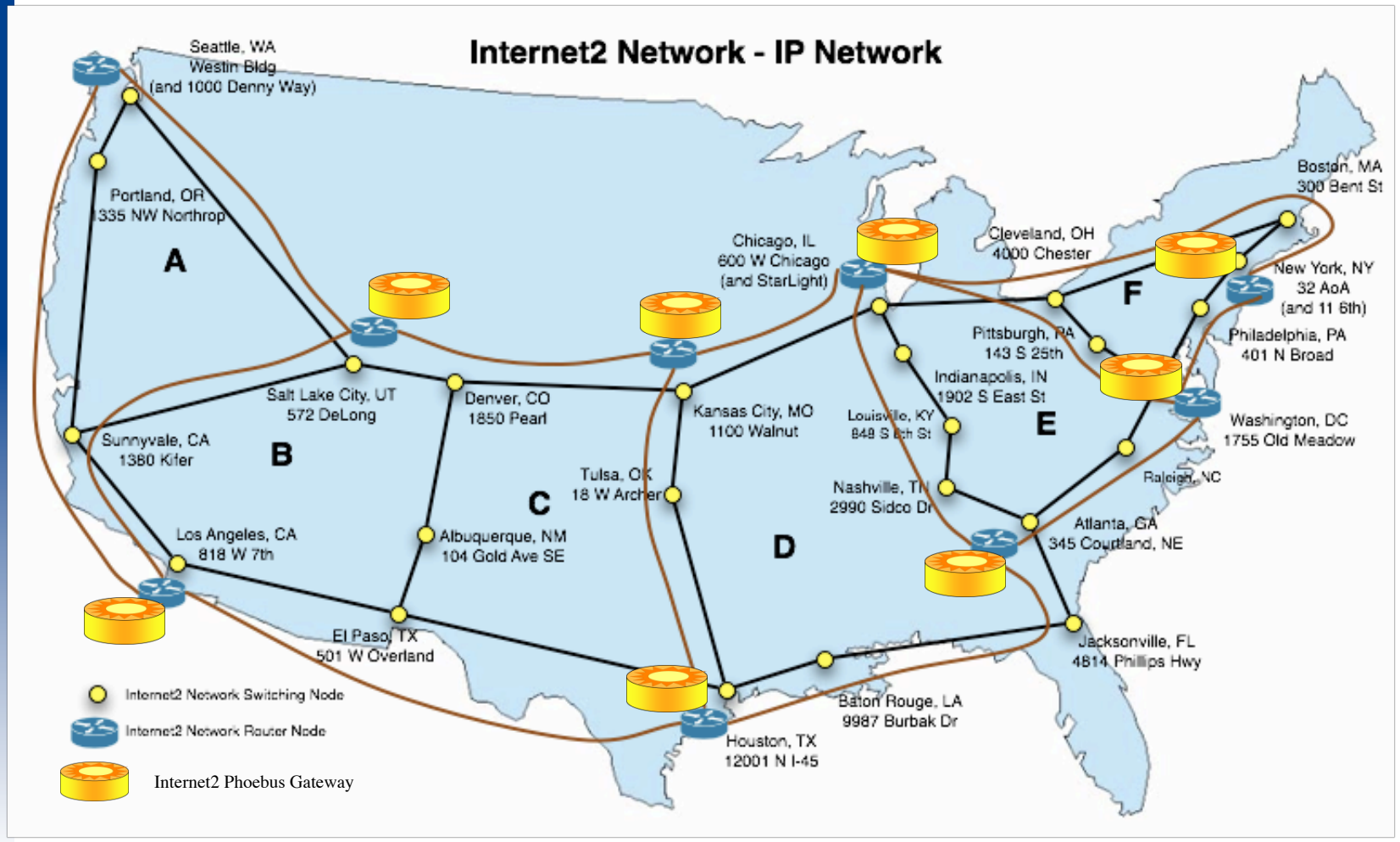
INTERNET2

Session Layer Protocol

- The eXtensible Session Protocol (XSP) can be used to manage a multi-layer connection



Deployment Status



Phoebus

Phoebus Authentication

- Password
  - SQLite/MySQL/File backends
- Trusted Host/Subnet
- GSI
  - Globus-based
- Anonymous
  - The session has no identifying information
- Accepted authentication handler can be set on a per host/per subnet basis

Phoebus

INTERNET2

Implementation - Library

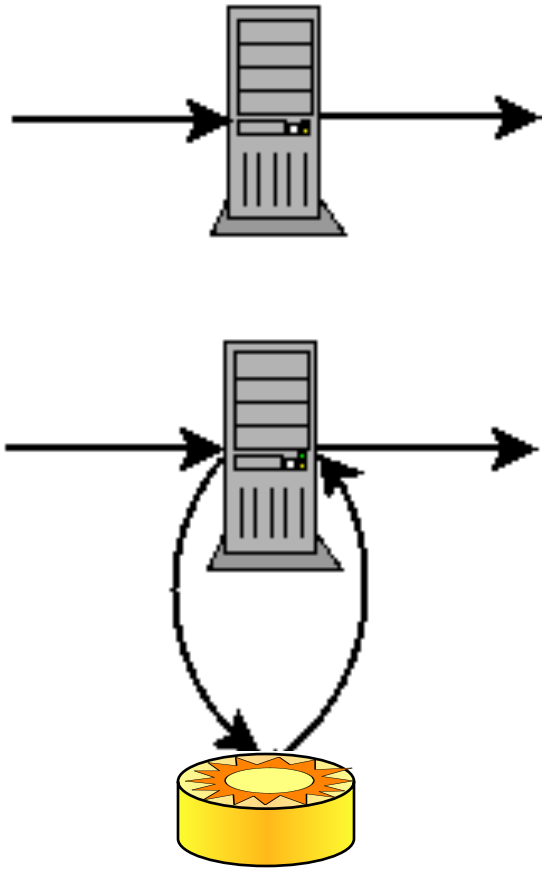
- The client library provides compatibility with current socket applications
  - AF_LSL
- On Linux, LD_PRELOAD is used for function override
  - `socket()`, `bind()`, `connect()`, `setsockopt()`…
  - Allows Un*x binaries to use the system without recompilation
- Prototype working on MacOS X

Phoebus

INTERNET2

Implementation - Intercept

- Intercept the TCP connection with IP Tables (on Linux)
- Redirect to local forwarding process
- Establish connection with appropriate service nodes or end node
  - Based on policy
- Transparent to end hosts

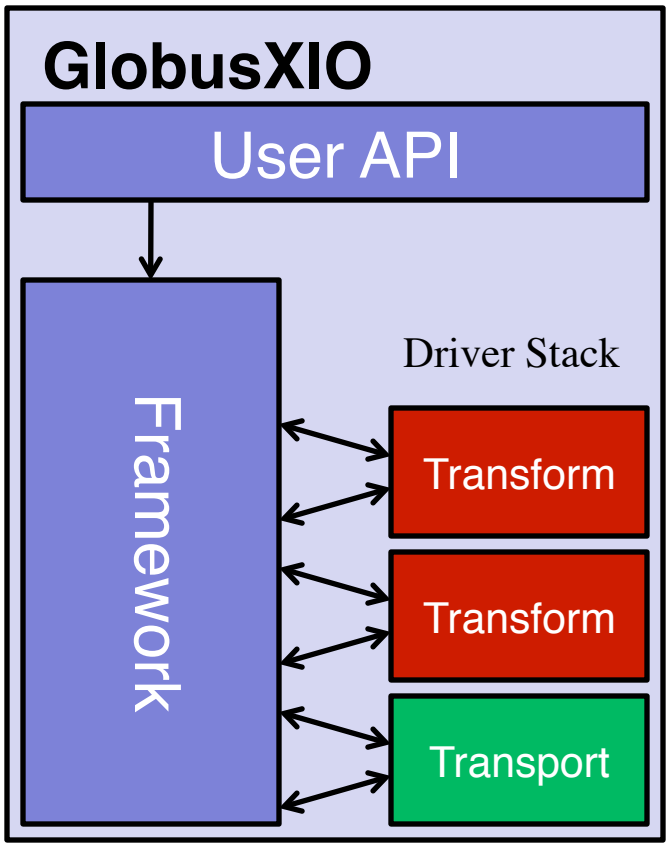


Phoebus

INTERNET2

Phoebus XIO Driver

- Provides a modular Phoebus transport driver for use with the Globus Toolkit
- Based on the TCP XIO driver
- Simplifies use of Phoebus Gateways
  - Eliminates need for shim library



Phoebus

INTERNET2

Phoebus and GridFTP

- *globus-gridftp-server* loads the Phoebus XIO driver when requested
- *globus-url-copy* extended to support Phoebus-based transfers
  - with -ph flag or explicitly with –dcstack
- Support for advanced features
  - 3rd party transfers
  - Parallel streams

Phoebus

INTERNET2

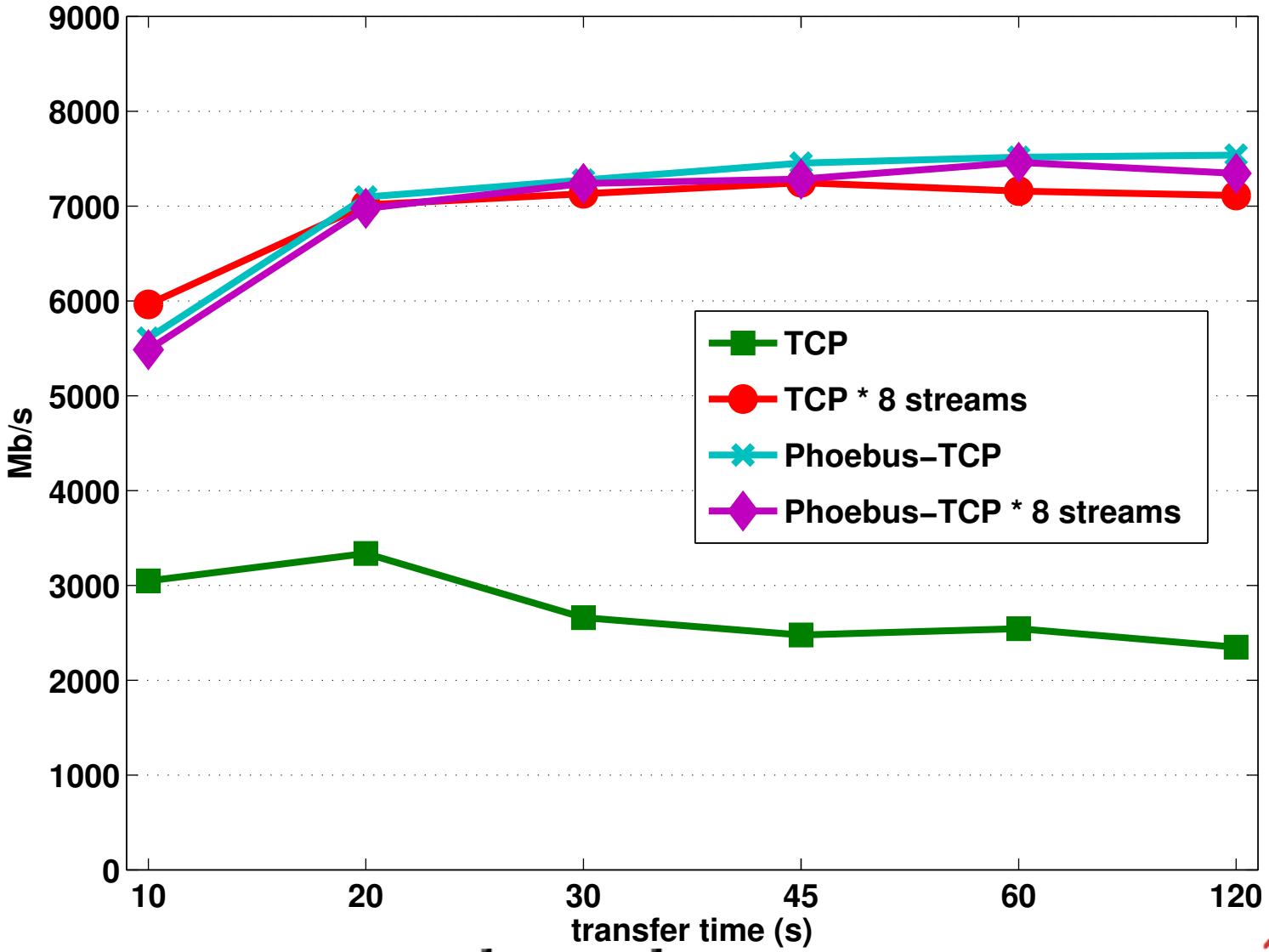
Windows Support

- SOCKS proxy support in development
- Java COGKit GridFTP
  - globus-url-copy
- Firefox Plugin jTopaz
  - available, uses Java implementation

Phoebus

INTERNET2

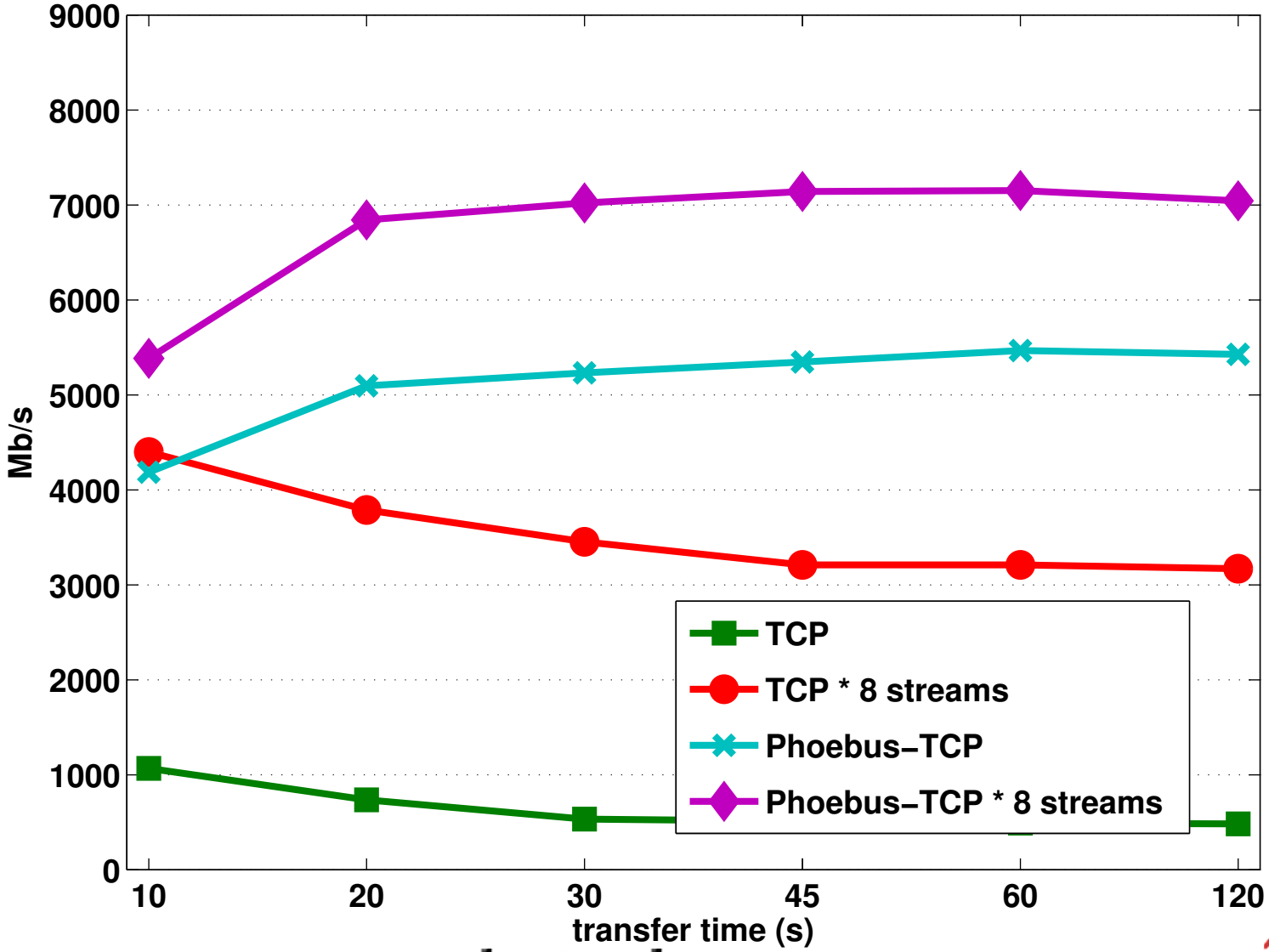
50ms Latency, .001% loss



Phoebus

INTERNET2

50ms Latency, .01% loss

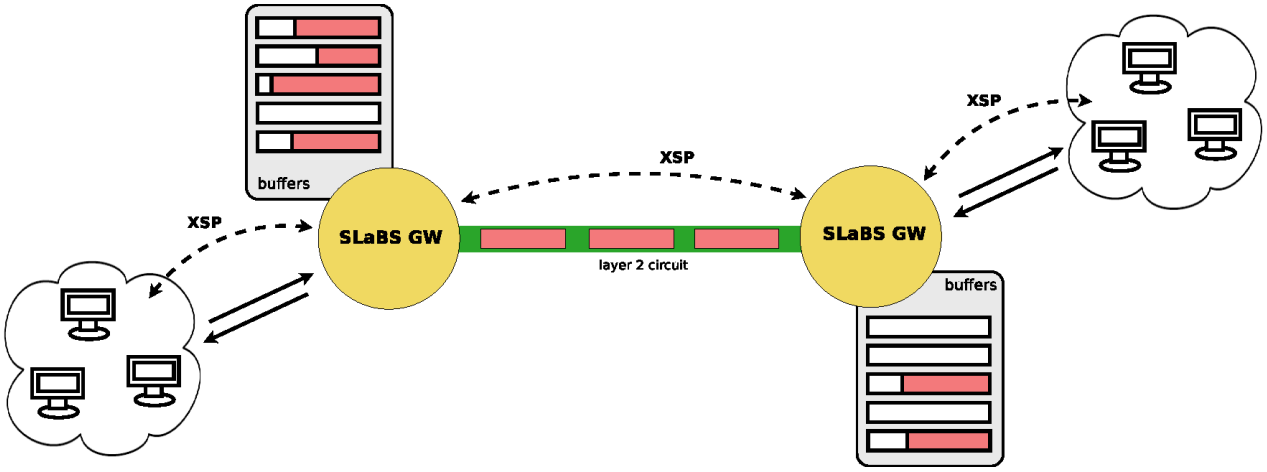


Phoebus

INTERNET2

SLaBS

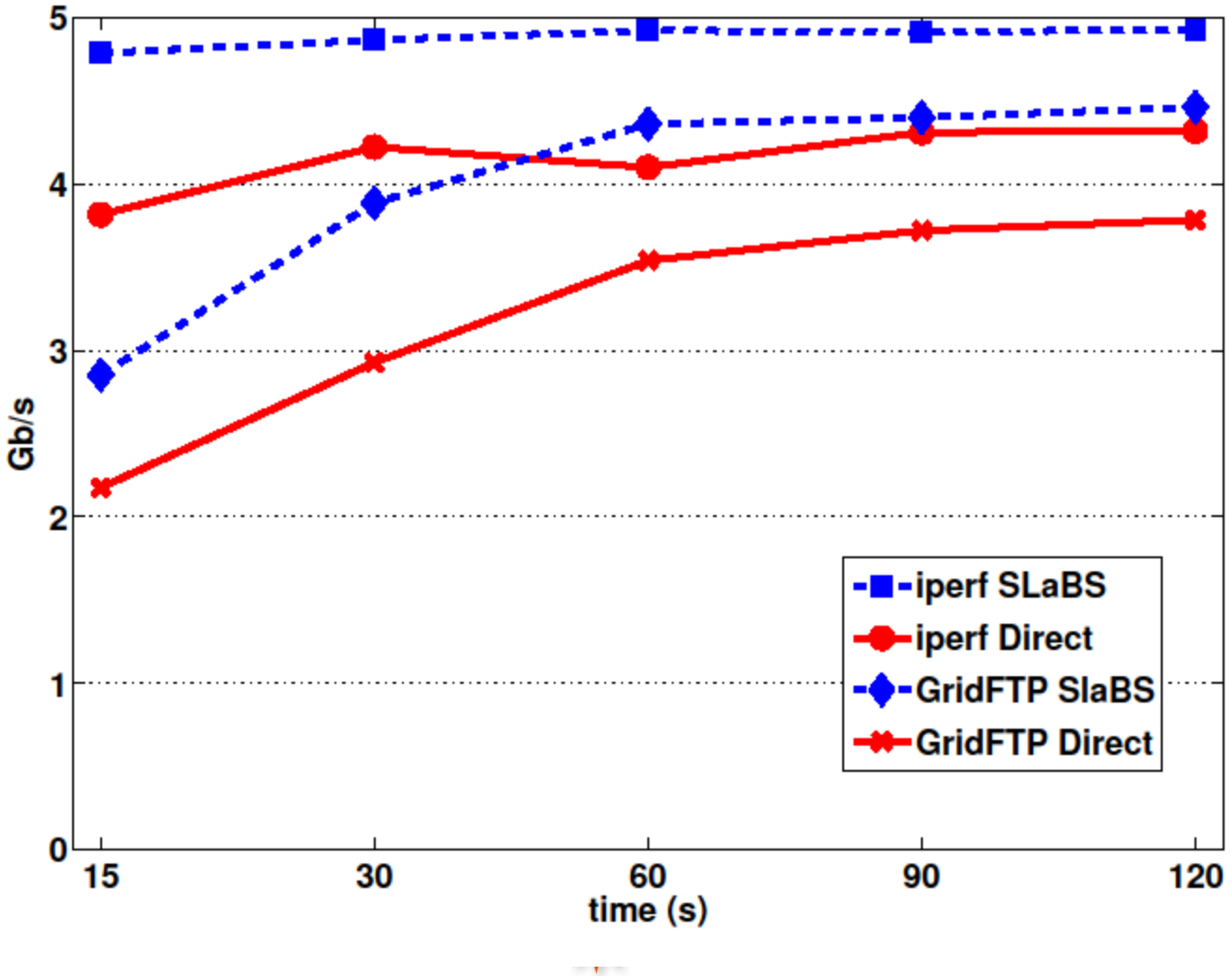
- Apply burst switching concepts at session-enabled gateways
  - Send relatively large PDUs versus small layer-3, layer-4 PDUs common today
- Schedule and optimize bursts over dedicated resources
- Reduce protocol overhead
- Hide provisioning latencies



Phoebus

GridFTP and iperf

4 parallel streams over 5G WAN bottleneck with 115ms RTT



Ongoing Efforts

- Implementation on Multi Service PIC/DPC on Juniper
- Optimized Myrinet-based forwarding
- Refactoring into standalone xspd for End Site Control Plane Services (DOE Project with Phil Demar and Dantong Yu)
  - A circuit maps pretty well to a session

Phoebus

INTERNET2

Acknowledgements

- UD Students
  - Ezra Kissel, Omer Arap, Miao Zhang
- Internet2:
  - Aaron Brown, Guy Almes (now at Texas A&M), Eric Boyd, Rick Summerhill, John Vollbrecht, Matt Zekauskas, Jason Zurawski, Jeff Boote
- US Department of Energy Office of Science, Mathematical, Information and Computational Sciences (MICS) Program
  - Early Career Principal Investigator program

End

- Thank you for your attention
- Questions?

Phoebus

INTERNET2