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
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
End-to-End Session

Dynamic Net

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

![Diagram of network layers](image)
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 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
The eXtensible Session Protocol (XSP) can be used to manage a multi-layer connection.
Deployment Status

Internet2 Network - IP Network

- Seattle, WA
  - Westin Bldg (and 1000 Denny Way)
- Portland, OR
  - 335 NW Northrop
- Sunnyvale, CA
  - 1380 Kifer
- San Francisco, CA
  - 30 Montgomery St
- Los Angeles, CA
  - 818 W 7th
- Denver, CO
  - 1850 Pearl
- Salt Lake City, UT
  - 572 DeLong
- Kansas City, MO
  - 1100 Walnut
- Tulsa, OK
  - 18 W Archer
- El Paso, TX
  - 501 W Overland
- Albuquerque, NM
  - 104 Gold Ave SE
- Baton Rouge, LA
  - 9987 Burbak Dr
- Houston, TX
  - 12001 N I-45
- New York, NY
  - 32 AOA (and 11 8th)
- Philadelphia, PA
  - 401 N Broad
- Washington, DC
  - 1755 Old Meadow
- Atlanta, GA
  - 345 Courtland, NE
- Jacksonville, FL
  - 4914 Phillips Hwy
- Raleigh, NC
- Pittsburgh, PA
  - 143 S 25th
- Indianapolis, IN
  - 1902 S East St
- Cleveland, OH
  - 4000 Chester
- Chicago, IL
  - 600 W Chicago (and StarLight)
- Boston, MA
  - 300 Bent St
- Internet2 Phoebus Gateway
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
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
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 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 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
Windows Support

- SOCKS proxy support in development
- Java COGKit GridFTP
  - globus-url-copy
- Firefox Plugin jTopaz
  - available, uses Java implementation
50ms Latency, .001% loss

- TCP
- TCP * 8 streams
- Phoebus–TCP
- Phoebus–TCP * 8 streams
50ms Latency, .01% loss

- **TCP**
- **TCP * 8 streams**
- **Phoebus-TCP**
- **Phoebus-TCP * 8 streams**

**Graph:**
- **Y-axis:** Mb/s
- **X-axis:** Transfer time (s)
- **Legend:**
  - TCP (green square)
  - TCP * 8 streams (red circle)
  - Phoebus-TCP (cyan cross)
  - Phoebus-TCP * 8 streams (violet diamond)
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
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
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?