Integrated Control Plane and Scheduler Developments

Nagi Rao,
Oak Ridge National Laboratory, raons@ornl.gov
Tom Lehman
Information Sciences Institute East, tlehman@isi.edu
Chin Guok
ESnet, cguok@lbl.gov

June 20, 2006
Meeting of Joint Engineering Task Force Engineering Taskforce
Washington, DC

Outline

- Introduction to Control Planes
  - USN
  - HOPI-DRAGON
  - OSCARS
- Multi Domain Resource Provisioning
  - A Framework
  - Integrated Path Computation
  - VLAN-Based Channel Alignment

Motivation: Disparate Control Plane Technologies

We area at a critical point in control-plane technologies for advanced networks:

- **MPLS/GMPLS being enabled and/or deployed by vendors**
- **Several control planes are being developed under different paradigms:**
  - **ESnet-OSCARS, CHEETAH, HOPI-DRAGON, UCLP, USN**
- **Underlying challenges and solutions slowly being understood:**
  - **security of control plane access**
  - **bandwidth optimization**
  - **data and control plane alignment**
  - **others**
- **Solutions appear to be diverging, at least on the surface but they address different application domains**

Control Plane Paradigms

Two main paradigms

- On-Demand Provisioning:
  - TL1,CLI: earlier mechanisms primarily for manual configuration
  - MPLS for layer 3, GMPLS for lower layers meant for automated configuration
- In-Advance Provisioning: Using a front-end for
  - Path computation and/or
  - Bandwidth optimization

  followed by mostly in-time signaling

These paradigms address different application needs:

- in-advance provisioning to coordinate with reservations on other facilities such as supercomputers
- on-demand provisioning to provide dedicated bandwidth

Some Control Plane Technologies

- HOPI-DRAGON – on-demand
  - GMPLS front end with CLI/TL1 for lower layers
  - Ethernet switches and routers
- OSCARS-ESnet – in-advance
  - Path computation + MPLS at layer 3
  - Cisco routers
- USN – in-advance
  - Bandwidth optimization + CLI/TL1
  - Ciena CDCI + Force10 E300
- CHEETAH – on-demand
  - On-demand provisioning using GMPLS
  - Sycamore SN16000 switches

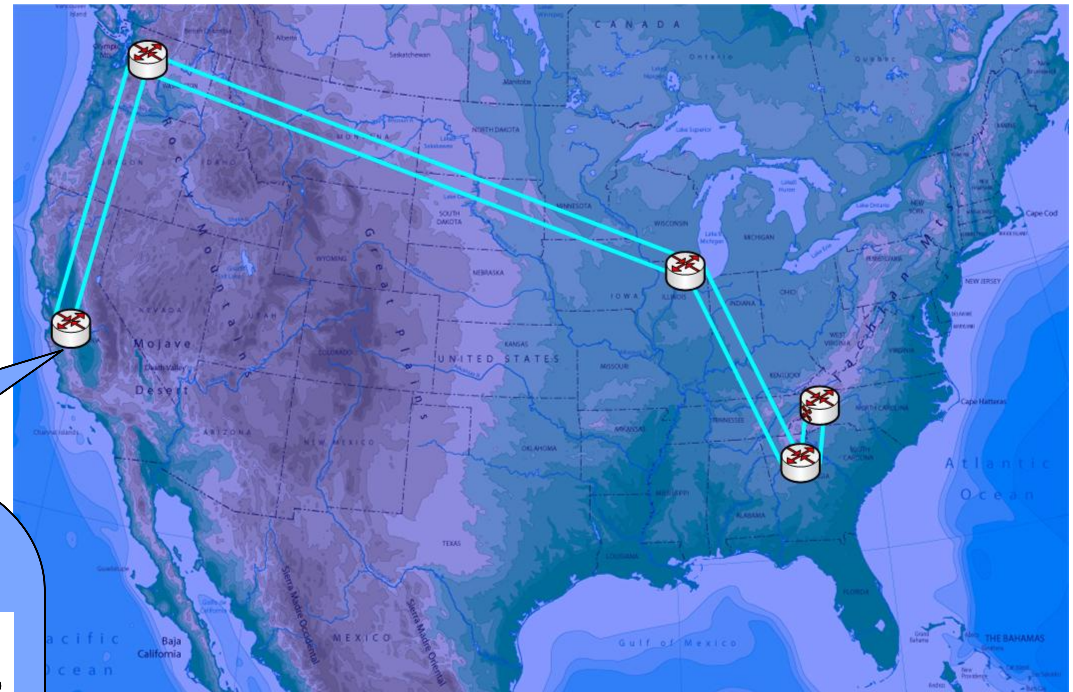
USN Data-Plane: Node Configuration

In the Core:

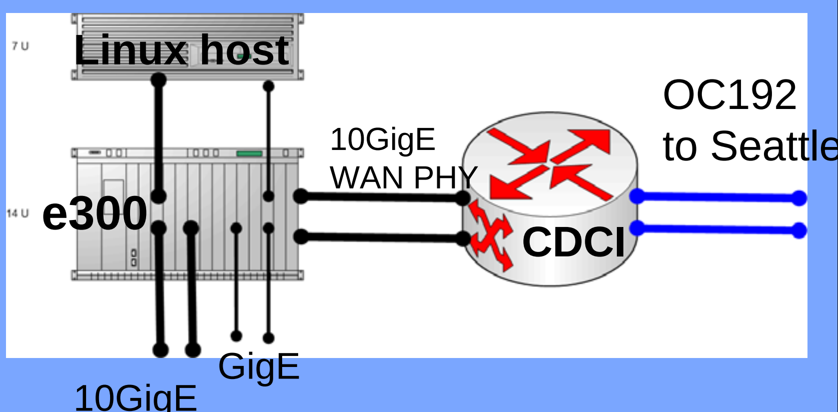
– Two OC192 switched by Ciena CDCIs

At the Edge

– 10/1 GigE provisioning using Force10 E300



Node Configuration



Connections to CalTech and ESnet

Data Plane User Connections:

Direct connections to:

core switches –SONET &1GigE

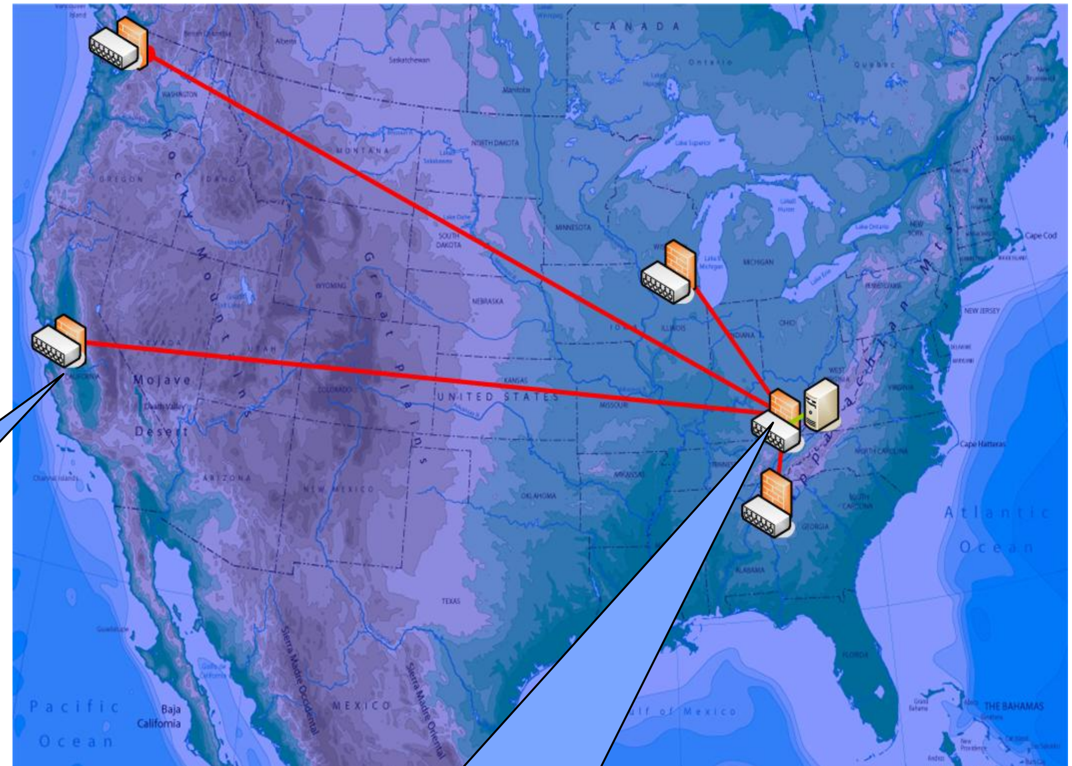
MSPP – Ethernet channels

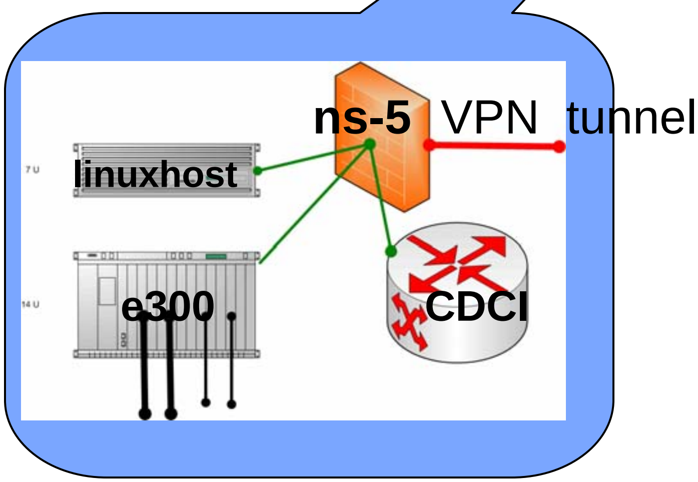
Utilize UltraScience Net hosts

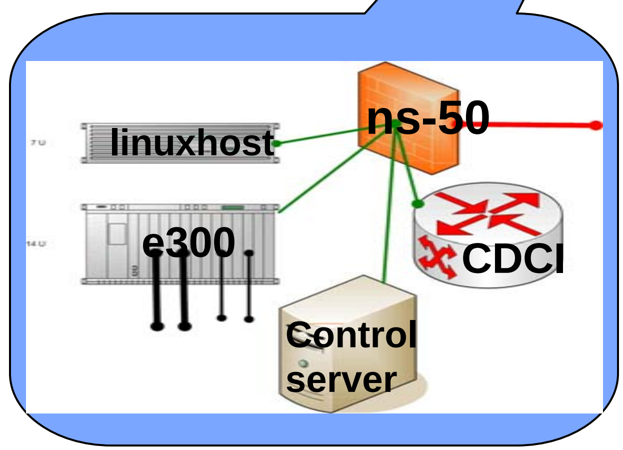
Secure Control-Plane

VPN-based authentication, encryption and firewall

- NetScreen ns-50 at ORNL
  NetScreen-5 at each node
- Centralized server at ORNL
  - **Bandwidth scheduling**
  - **Signaling**

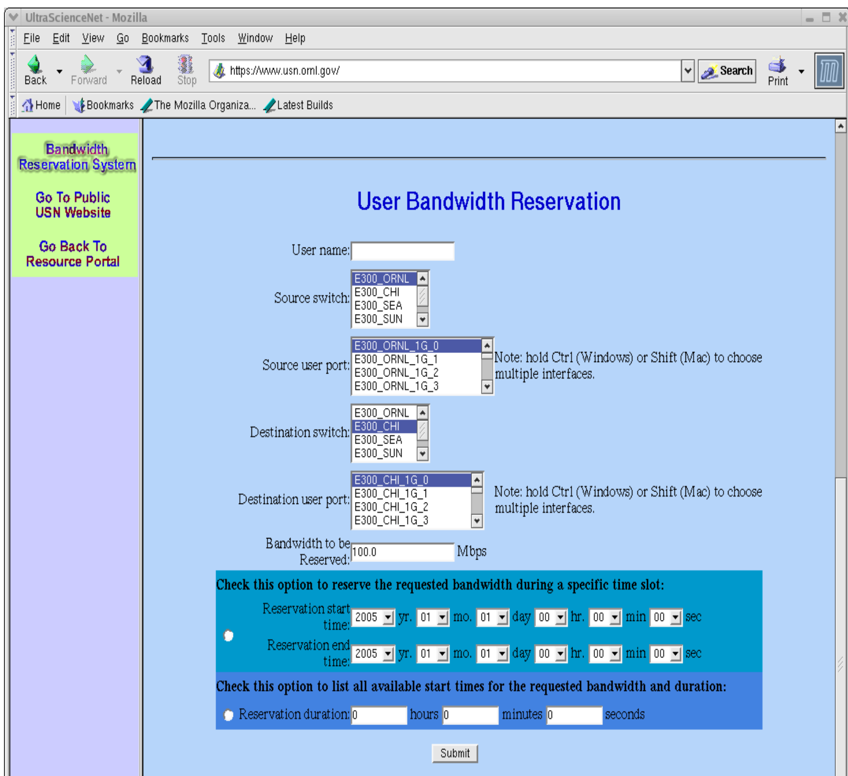


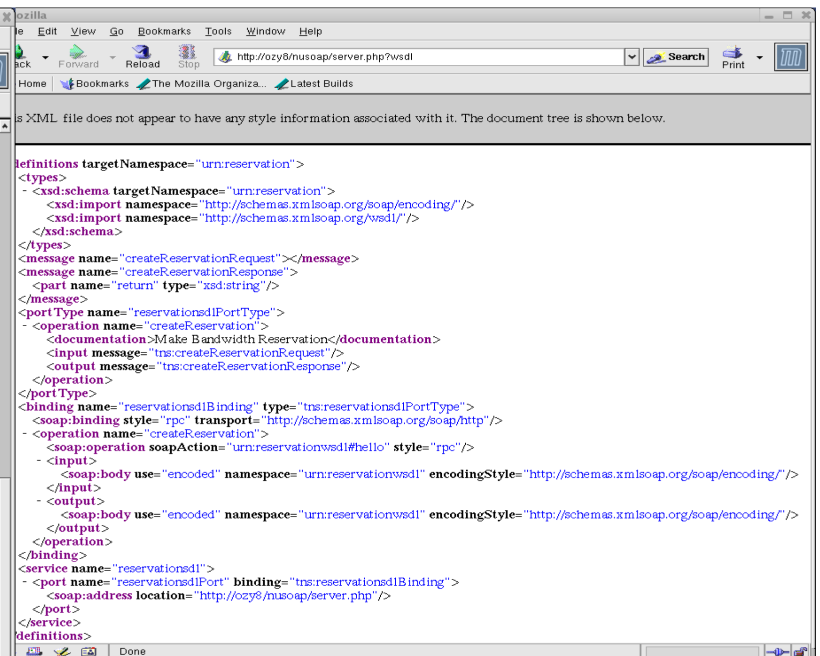




USN Control Plane

- Phase I
  - Centralized path computation for bandwidth optimization
  - TL1/CLI-based communication with CoreDirectors and E300s
  - User access via centralized web-based scheduler
- Phase II (current)
  - GMPLS wrappers for TL1/CLI
  - Inter-domain “secured” GMPLS-based interface
  - Webservices interface





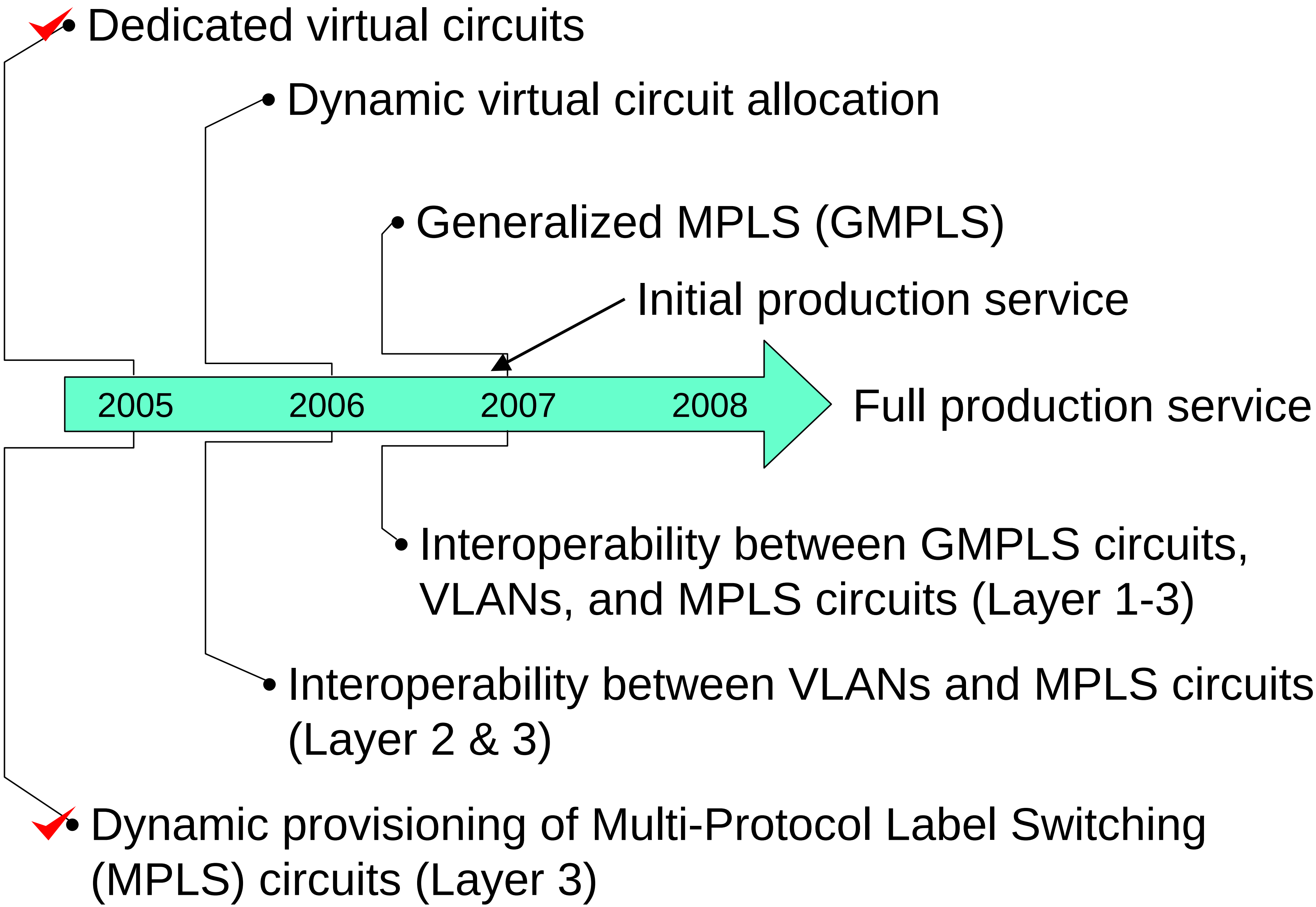
OSCARS: Guaranteed Bandwidth VC Service For SC Science

- A prototype service has been deployed within ESnet as a proof of concept
  - To date more then 20 accounts have been created for beta users, collaborators, and developers
  - More then 100 reservation requests have been processed
- In its current phase this effort is being funded as a research project by the Office of Science, Mathematical, Information, and Computational Sciences (MICS) Network R&D Program

Functional View of OSCARS

- Support user/application VC reservation requests
  - ✓ Users enter reservations via a web-page
  - – Applications uses API to send signed SOAP messages
- Manage allocations of scarce, shared resources
  - ✓ Centralized resource management
  - – Authentication is done using X509 certificates
  - – Authorization TBD
- Provide circuit setup and teardown mechanisms
  - ✓ OSPE-TE for routing
  - ✓ RSVP-TE for signaling
  - ✓ MPLS for switching
- Enable the claiming of reservations
  - ✓ Policy based filtering for traffic destined for VC
- Enforce usage limits
  - ✓ Per VC admission control
  - ✓ Separate router queue for VCs

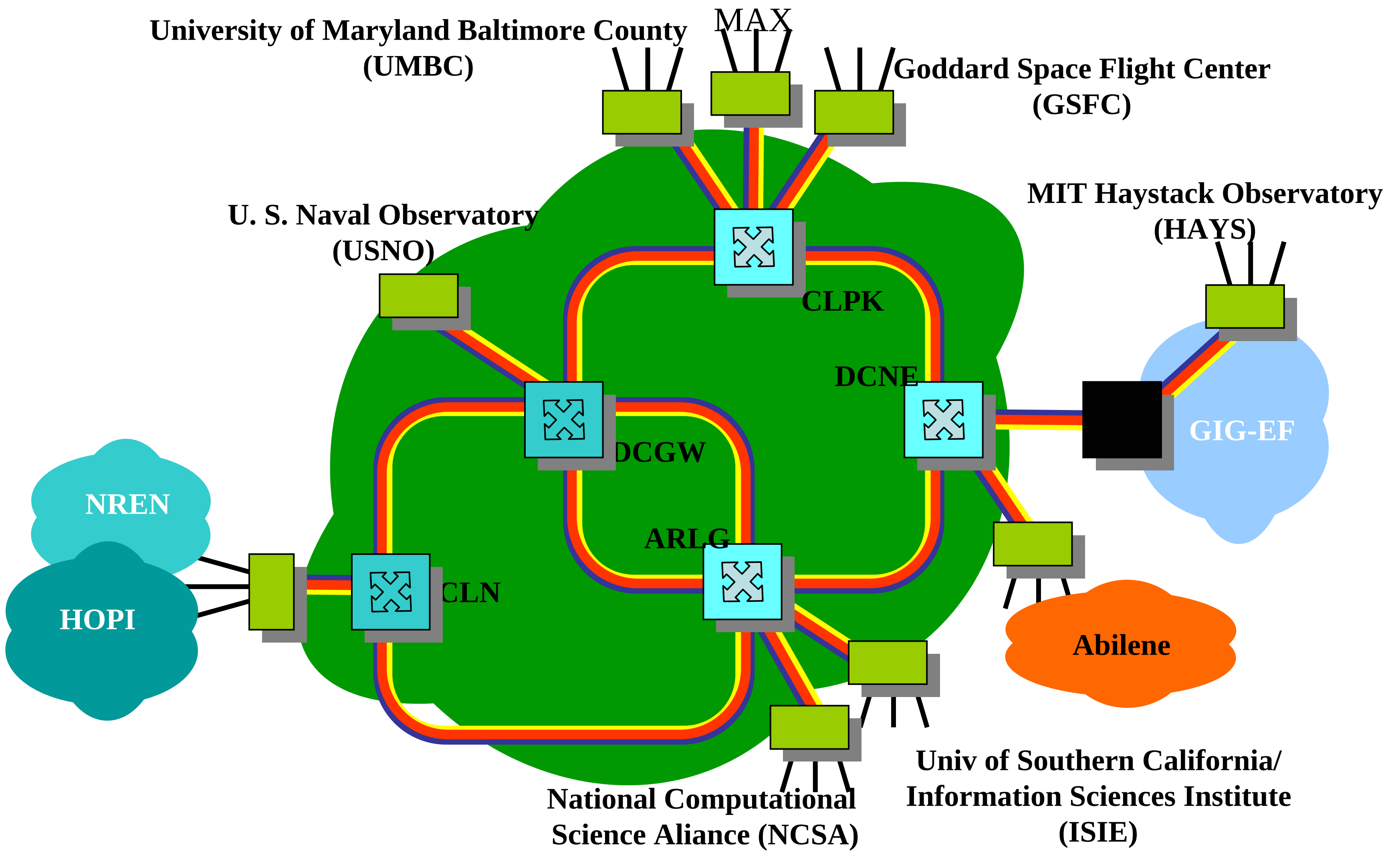
ESnet Virtual Circuit Service Roadmap



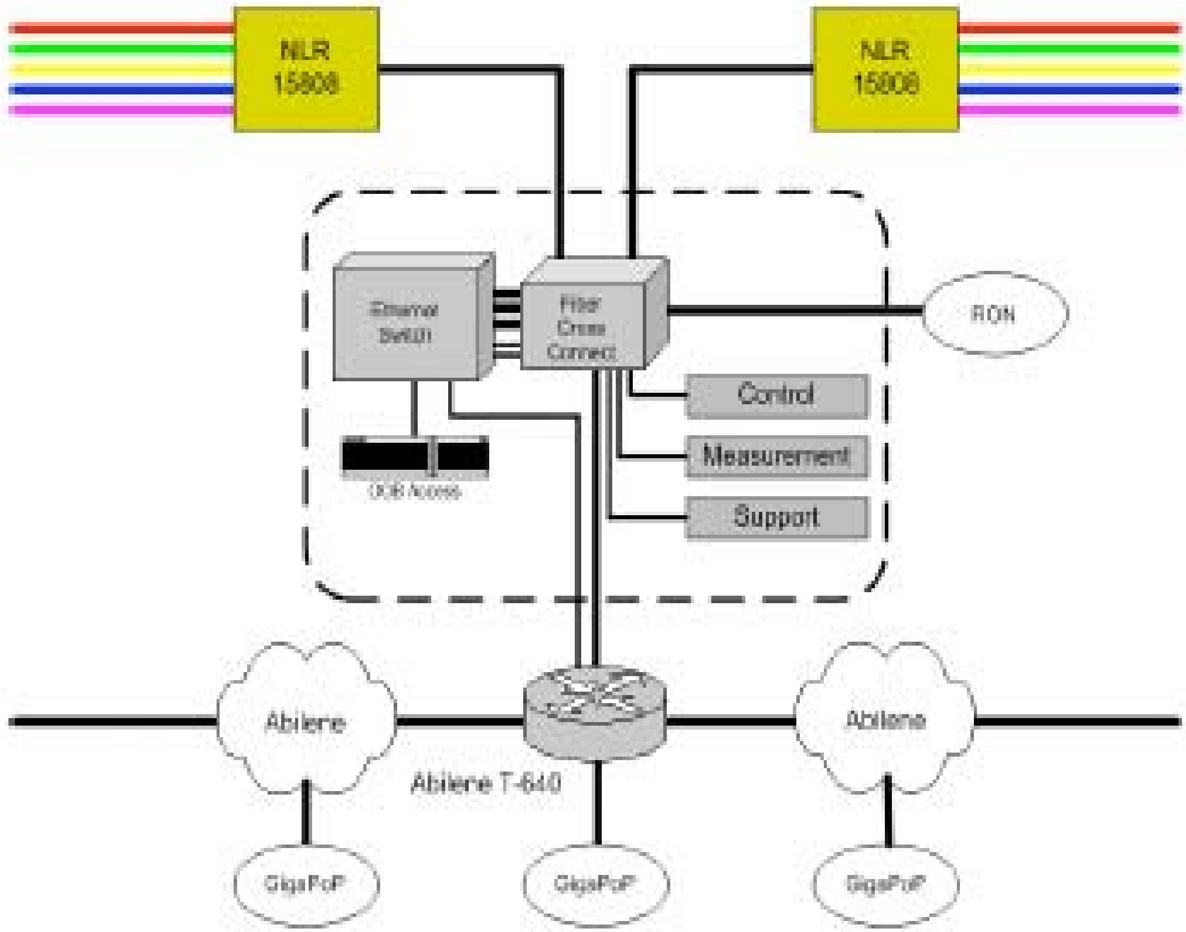
DRAGON: Project Features and Objectives

- **Utilize GMPLS protocols for dynamic provisioning of Light Paths**
  - Addition of CSPF Path Computation algorithms for wavelength routing
- **Inter-domain service routing techniques**
  - Network Aware Resource Broker (NARB) for service advertising, inter-domain ERO generation, AAA
- **Application Specific Topology Description Language**
  - Formalized means to describe the application topology and network service requirements
- **Integration with real applications**:
  - E-VLBI
  - HD-CVAN

The DRAGON Testbed
Washington, DC metro region



HOPI Node


NLR
15808
NLR
15808
RON
Control
Measurement
Support
Abilene
Abilene
Abilene T-640
GigaPoP
GigaPoP
GigaPoP

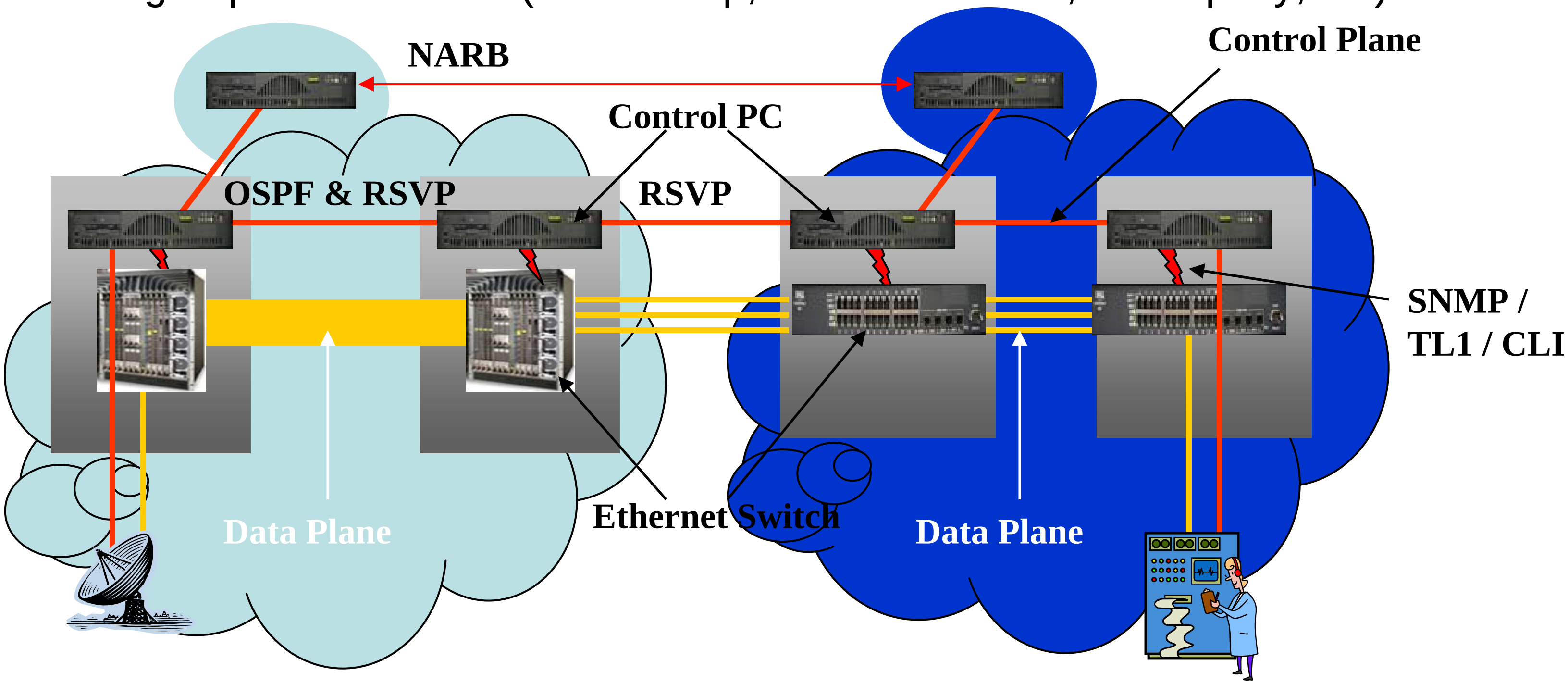
DRAGON Technologies

- Open Source GMPLS Control Plane
  - Virtual Label Swapping Router – **VLSR**
    - Open source OSPF-TE & RSVP-TE to control Ethernet switches and fiber switches
  - Network Aware Resource Broker – **NARB**
    - GMPLS-OSPF-TE listener
    - Performs the inter-domain routing, AAA, scheduling, PC
  - Advanced Constrained Shortest Path First Path Computation Element – CSPF PCE
    - Domain selectable abstraction levels and end-to-end LSP
- Application Specific Topologies - **ASTB**
  - Formalization of application's resource requirements – particularly the network resources.
- Photonic metro-scale wavelength services
  - Reduce/eliminate unnecessary OEO,
  - allow user generated ITU waves to transit the metro network (alien waves)
  - Framing and encoding agnostic

DRAGON Control Plane

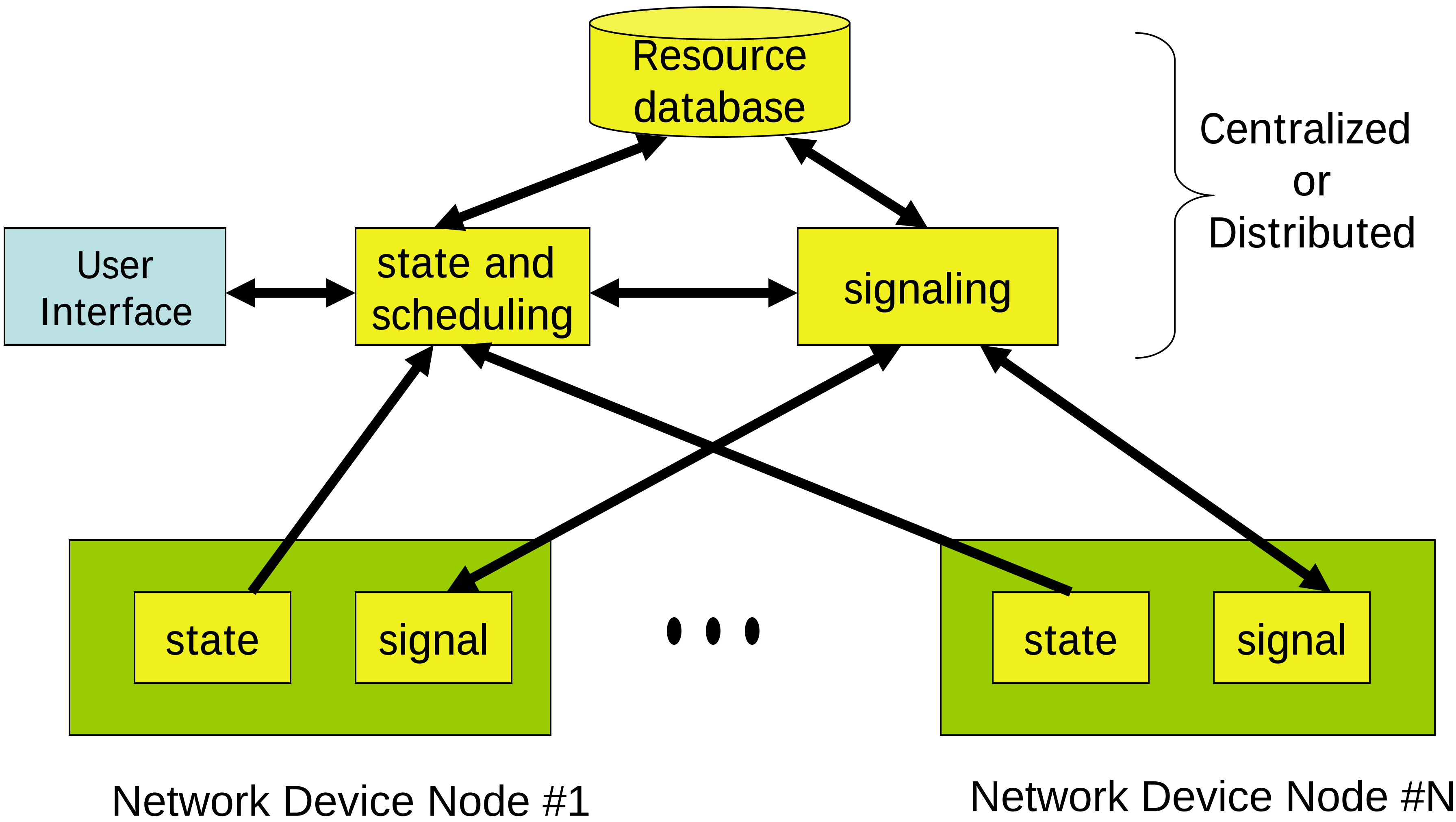
Virtual Label Switching Router (VLSR)

- Open source protocols running on PC act as GMPLS signaling entity
- Control PCs participate in protocol exchanges and reprovision covered switch according to protocol events (PATH setup, PATH tear down, state query, etc)



CHEETAH control-plane is quite similar – uses Sycamore SN16000

A General Control-Plane Architecture



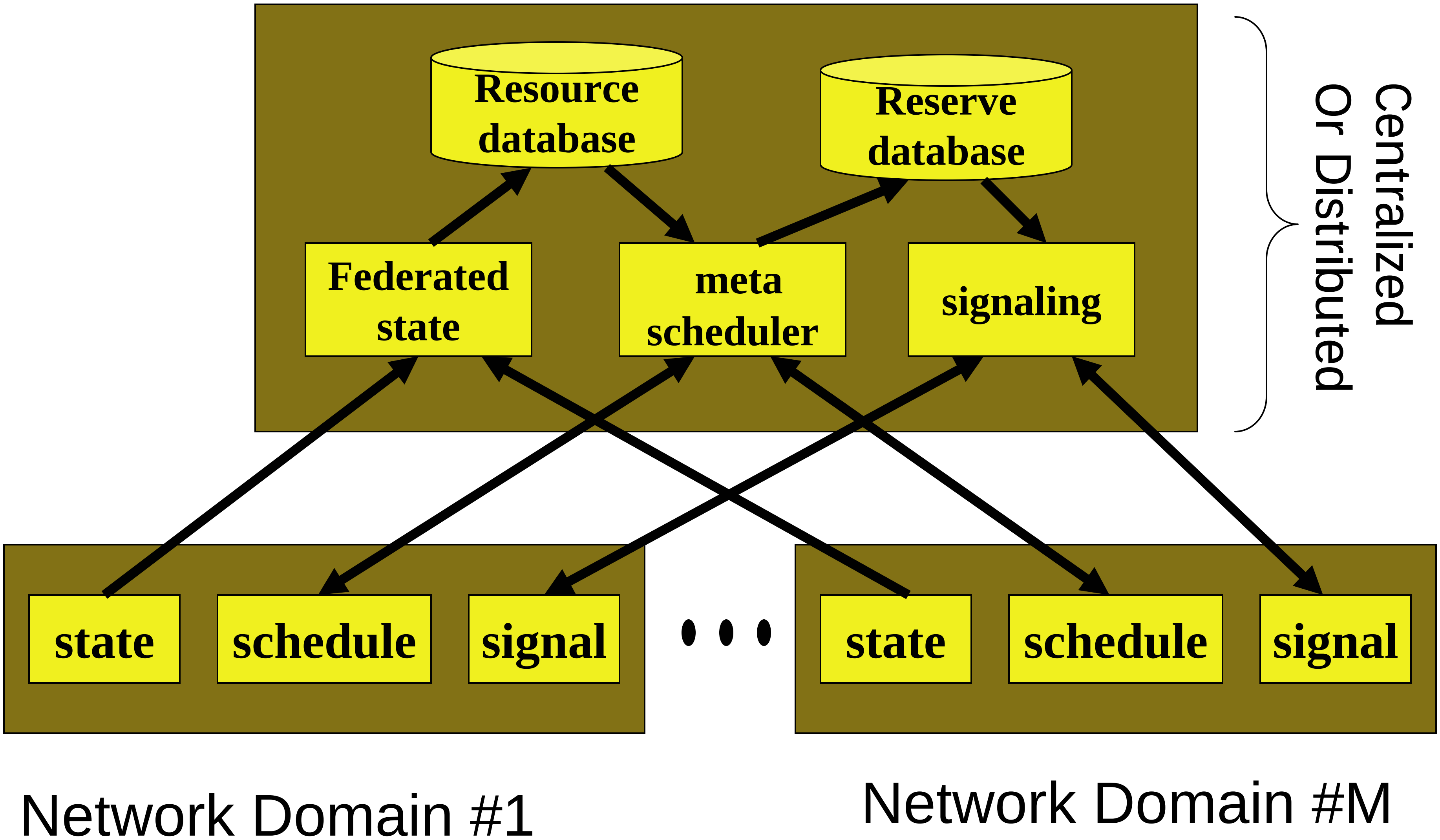
Some modules might be trivial in some control planes

Integrated Multi-Domain Resource Provisioning – A Perspective

- **Front-End website/webservice**
  - Integrates all data bases; SOAP interface
- **Meta-Scheduler**
  - Path Composition and Alignment
    - Computes multi-domain path; Queries domain schedulers
  - May include host scheduling
- **Coordinated Signaling**
  - Sends path requests to domain signaling modules
- **Multi-Domain Authentication, Authorization and Accounting**

A proposal is under consideration with DOE High-Performance Networking program on this topic

Multi-Domain Resource Provisioning Architecture



USN Path Computation – Bandwidth Optimization

Different paths may be computed:

- **(i) A specified bandwidth in a specified time slot,**
- **(ii) Earliest available time with a specified bandwidth and duration,**
- **(iii) Highest available bandwidth in a specified time slot,**
- **(iv) All available time slots with a specified bandwidth and duration.**

All are computed by extending the shortest path algorithms using a closed semi-ring structure defined on sequences of real intervals

- **(i)-(iii): Variation of Dijkstra's shortest path algorithm**
- **(iv): Variation of Bellman-Ford algorithm;**
  - **- previously solved using transitive-closure algorithm**

$$\left(S, \oplus, \otimes, \overline{0}, \overline{1}\right)$$

S: Sequence of disjoint real intervals $\left\{[l_1, h_1], \cdots, [l_p, h_p]\right\}$

$\oplus$: Point-wise union

$\otimes$: Point-wise intersection

$\overline{0}$: $\{R^+\}$

$\overline{1}$: $\{R^+\}$

All-Slots Algorithm

Given network with band width allocations on all links

ALL-SLOTS returns all possible starting times for a connection with bandwidth *b* duration *t* between source node *s* and destination node *d*

Modified Bell-Ford algorithm:
Time-complexity: $O(mn)$

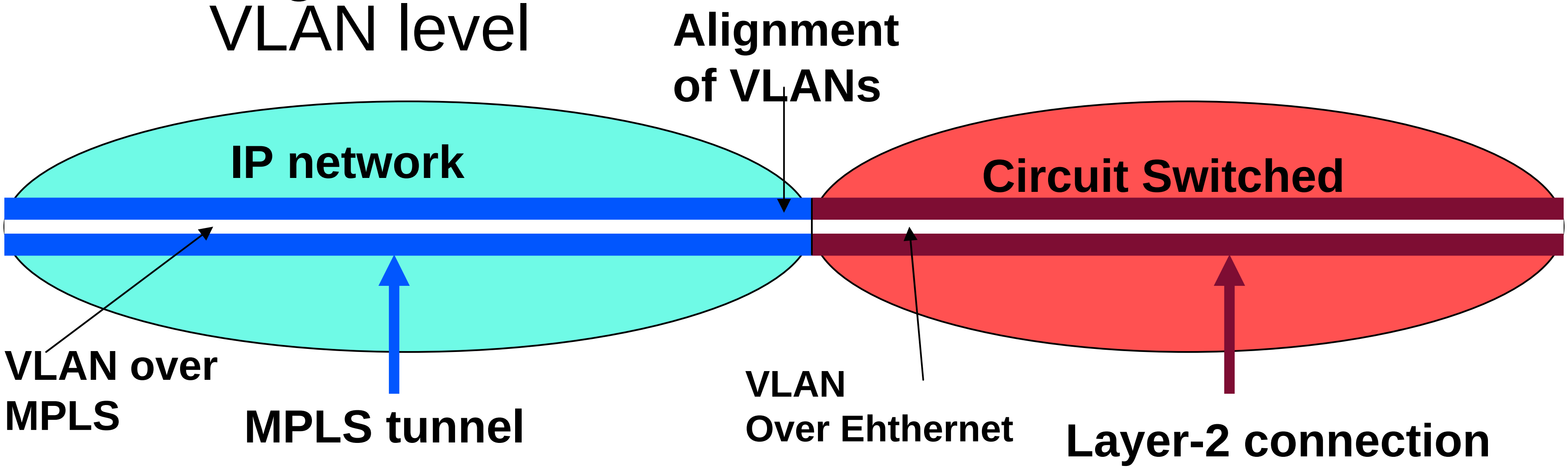
More efficient than transitive-closure algorithm: $O\left(n^3\right)$

Algorithm ALL-SLOTS

1. $\tau(s) \leftarrow \{\Re\}$;
2. $\tau(v) \leftarrow \{\varnothing\}$ for all $v \neq s$;
3. for $k = 1, 2, \ldots, n-1$ do
4. for each edge $e = (v, w)$ do
5. $\tau(w) \leftarrow \tau(w) \oplus \{\tau(v) \otimes L_e\}$;
6. return $(\tau(d))$.

VLAN – Unifying Data-Plane Technology

- IP networks
  - VLANs Implemented in MPLS tunnels
- Circuit switched networks
  - VLANs Implemented on top of Ethernet channels
- Align IP and circuit connections at VLAN level



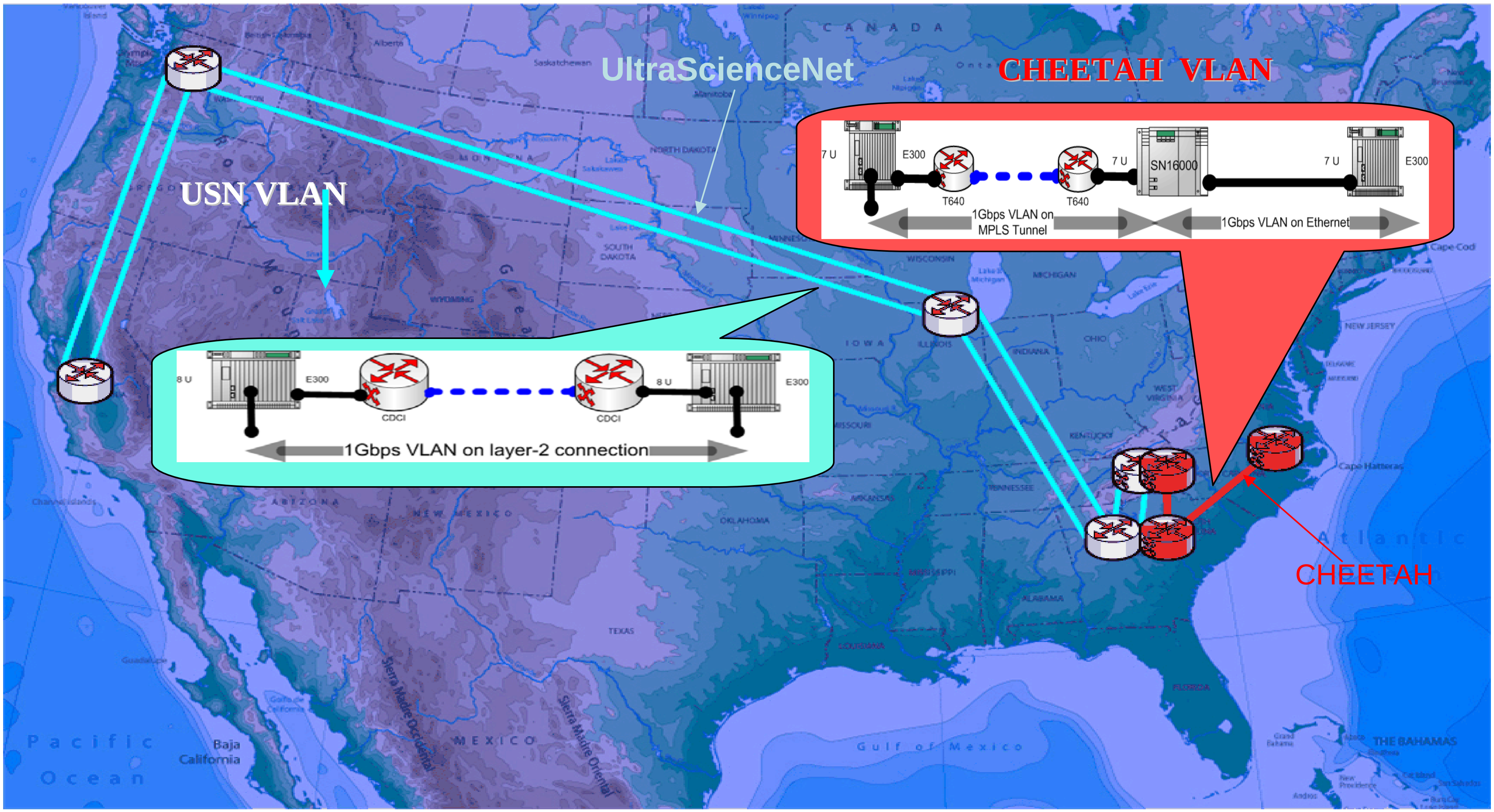
USN– CHEETAH VLAN Alignment

UltraScience Net: Layer-2

VLAN: E300 – CDCI - … - CDCI – E300

CHEETAH: layer-3 + layer 2

VLAN: T640-T640 – SN1600 – Cisco 3750



Technical Tasks

- AAA System
  - User/Client authentication over multiple domains
- Meta-Scheduler
  - Composite graph
    - Nodes: peering points
    - Edge: in-domain connections or resources
  - Path computation and composition
    - Issue individual domain path requests
    - Agglomerate domain paths
- Integrated Signaling System
  - Issue reservation commands
  - Exception handling

Domain Technical Tasks

- USN
  - Webservice interface
  - Integrated VLAN specification
- OSCARS
  - VLANs over MPLS tunnels
  - Provisioning to Chicago and Sunnyvale routers
- HOPI-DRAGON
  - Webservice interface
  - Advanced Scheduling
  - Peering in Chicago and Seattle

Integrated Scheduler for Hybrid Networks

- MPLS tunnel over IP network
- Layers 1-2 over switched network

Approach

- Higher level scheduler composes end-to-end VLAN consisting of portions of MPLS tunnels and layer 1-2 channels
- Extensions of Dijkstra and Bellman-Ford Algorithms
- Coordinated daemons will be launched to setup the needed portions of path

Layer 3 connection	**Layer 2/1 circuit**	**Layer 3 connection**	
MPLS	**GMPLS**	**MPLS**	**On-demand**
OSCARS+	**USN+(TL1,CLI)**	**OSCARS+**	**In-advance**

Conclusions

- Coming Together of Control-Plane Technologies
  - Different application domains are optimally supported by different control-planes
  - Multiple-domain paths are needed for several applications
  - Several new technologies exist but some more are needed, and they all must be integrated
- Hybrid Integrated Architecture
  - Networks that combine best of shared IP and dedicated channels are feasible but need further development
- Peering Multiple Control-Planes is complex
  - Synchronization of Schedulers
  - Multi-domain path composition
  - Coordinated signaling
  - Multi-Domain AAA

Thank you

Need for Secure Control Plane

- Security of control plane is extremely important
  - USN switches (Ciena, Force10, Turin, Sycamore, Whiterock) do not support IPSec – do not know of any that do
  - TL1/CLI and GMPLS commands sent in the “clear”
    - Can be sniffed to profile the network
    - Can be injected to “take over” the control
  - Following cyber attacks could be easily launched
    - Hijack the dedicated circuits; sustain a DOS flood to prevent recovery
    - Takeover/flood UltraScienceNet end hosts and switching gear
- USN control-plane is out-of-band and secure
  - Uses VPN-based control channels and firewalled enclaves