Dynamic Services Control Plane Overview and Status

February 15, 2007
GLIF Meeting
Minneapolis, Minnesota

Tom Lehman
University of Southern California
Information Sciences Institute (USC/ISI)

DRAGON Control Plane Team at ISI, MAX, GMU

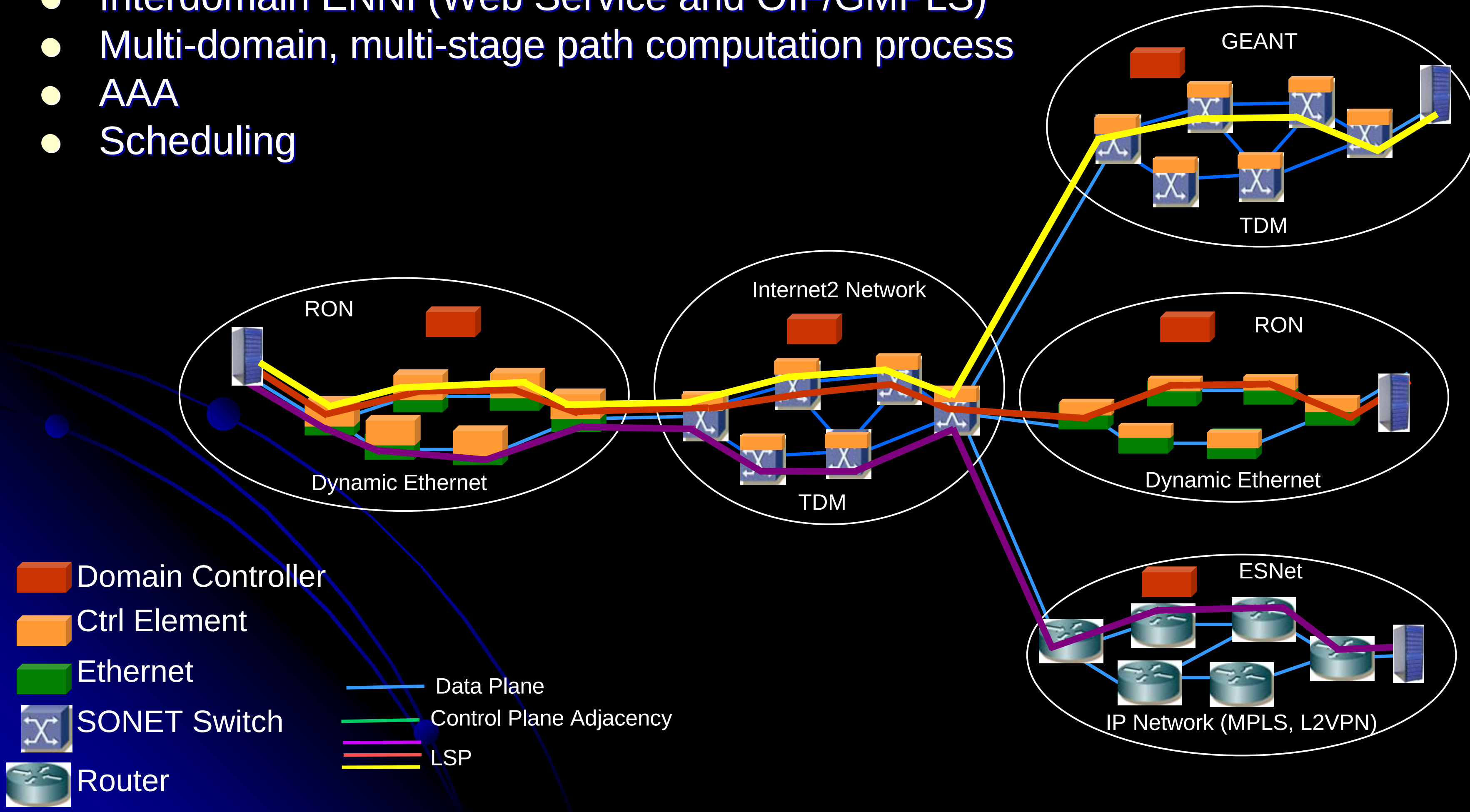
- Jerry Sobieski (MAX)
- Chris Tracy (MAX)
- Jarda Flidr (MAX)
- Fiona Leung (MAX)
- Bill Babson (MAX)
- Bijan Jabbari (GMU)
- Xi Yang (ISI)
- Tom Lehman (ISI)

Control Plane Objectives

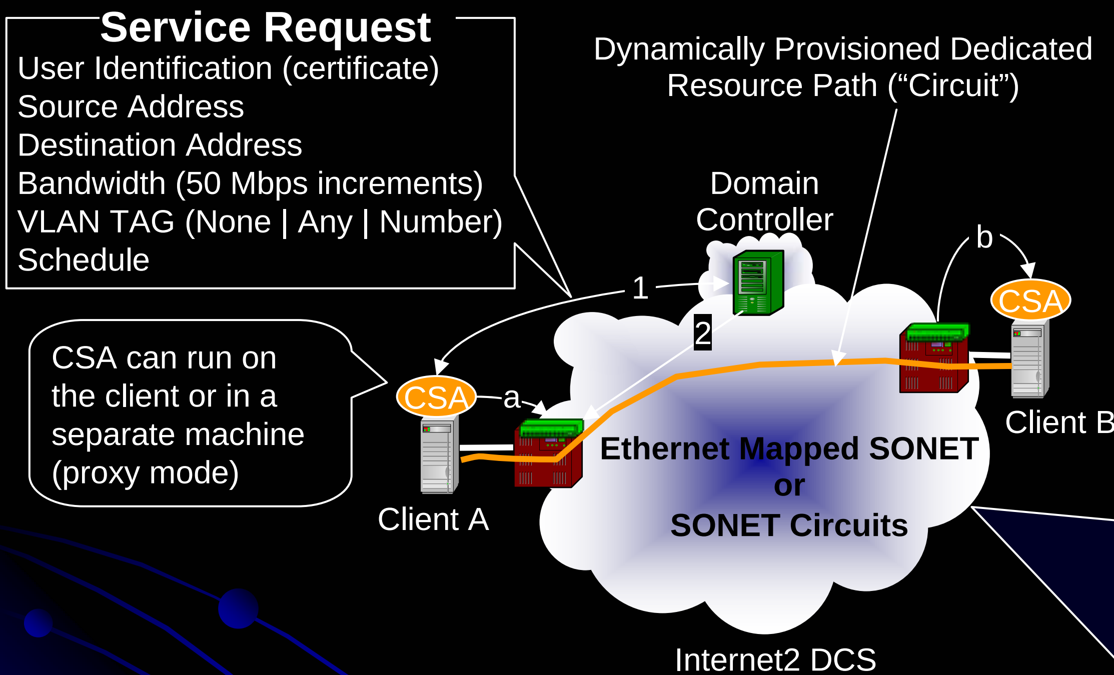
- Multi-Service, Multi-Domain, Multi-Layer, Multi-Vendor Provisioning
  - Basic capability is the provision of a “circuit” in above environment
- In addition, need control plane features for:
  - AAA
  - Scheduling
  - Easy APIs which combine multiple individual control plane actions into an application specific configuration (i.e., application specific topologies)

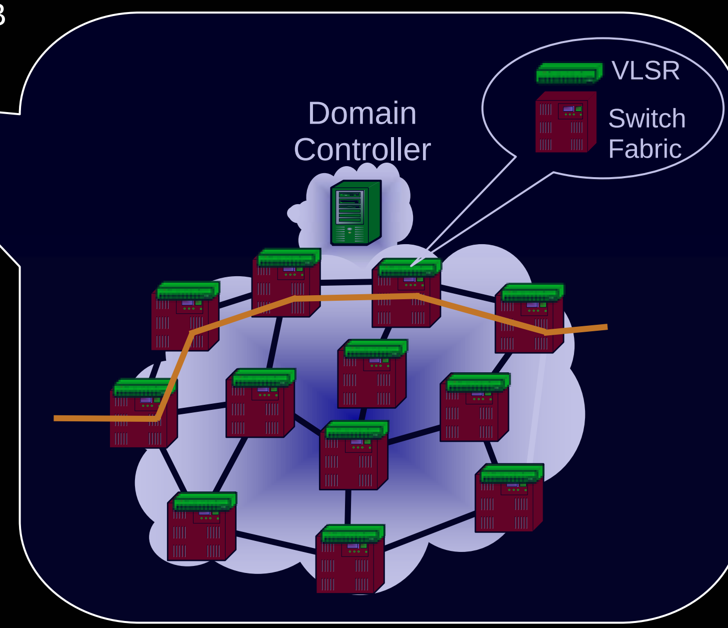
Multi-Domain Control Plane
The (near-term) big picture

- Multi-Domain Provisioning
- Interdomain ENNI (Web Service and OIF/GMPLS)
- Multi-domain, multi-stage path computation process
- AAA
- Scheduling



Client “Service” View IntraDomain



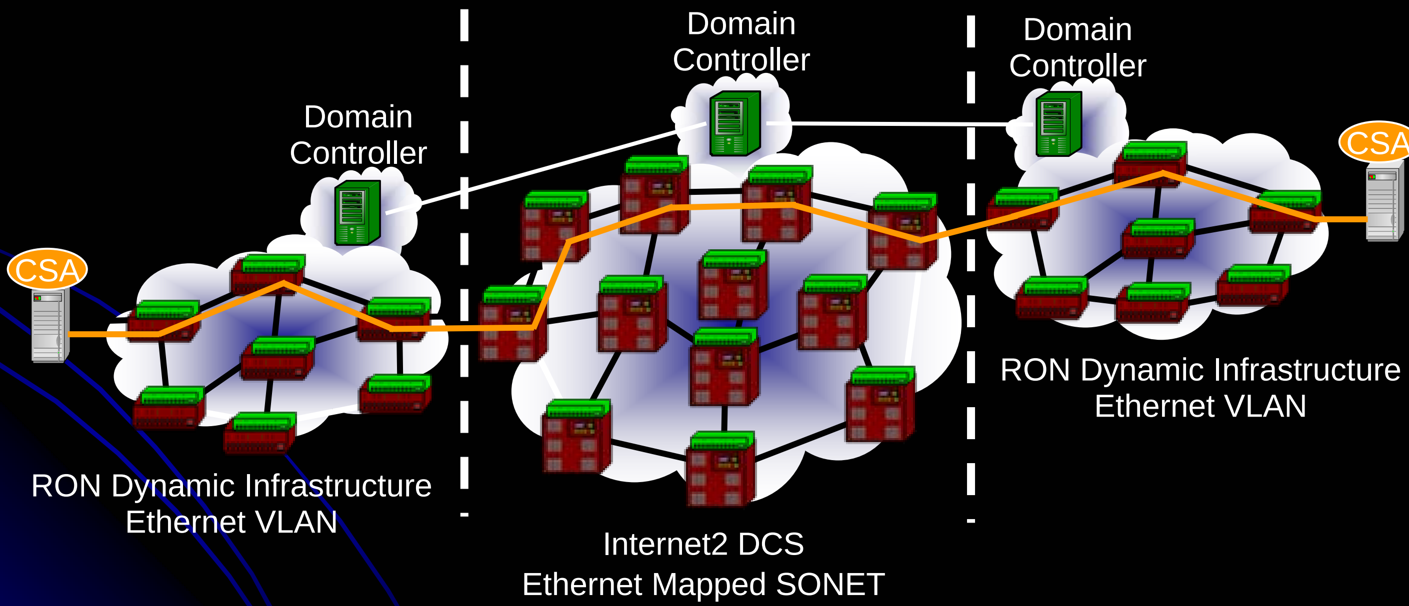


- Items 1,2 represent service request/approval

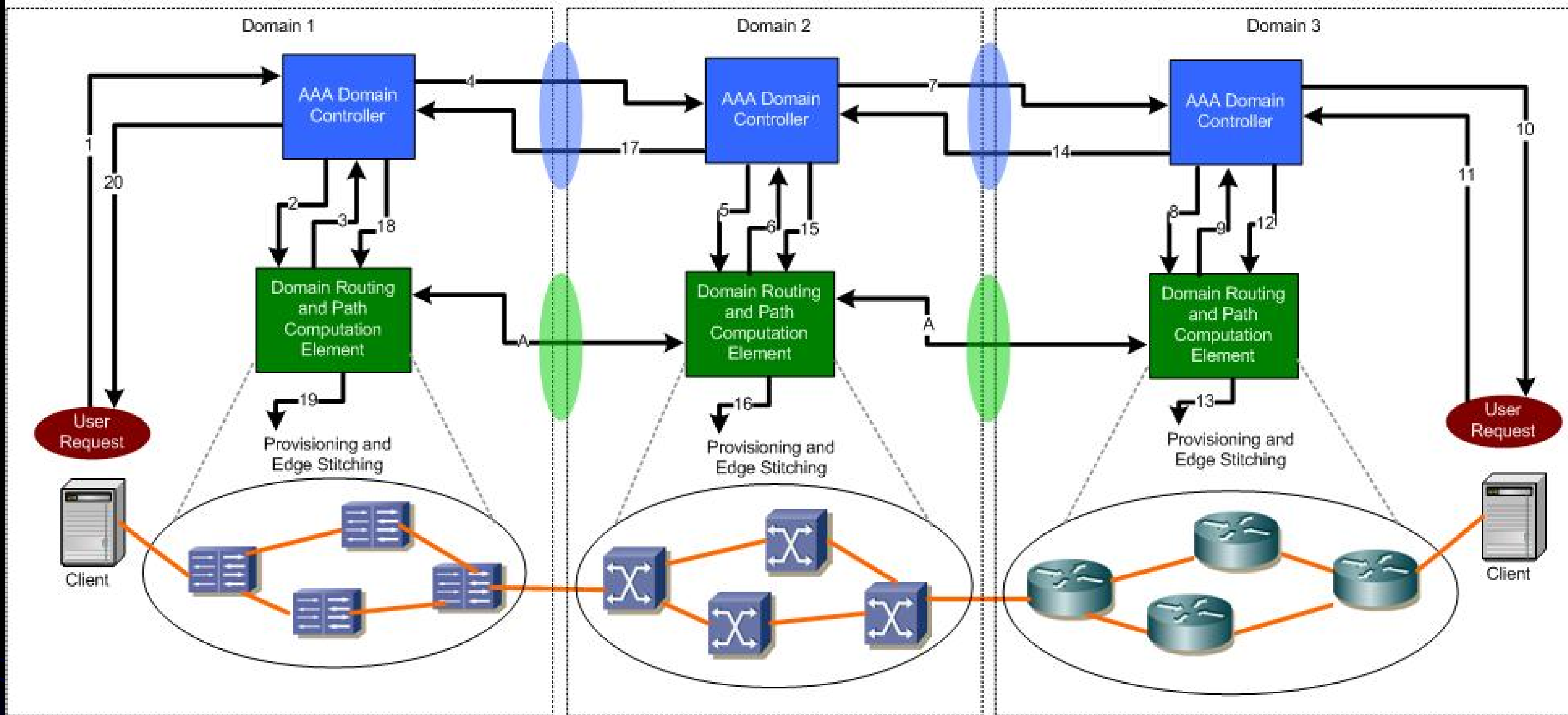
- Items a,b represent service instantiation (signaling)

InterDomain

- From a client perspective, an InterDomain provisioning differs little as compared to IntraDomain
- However, additional work for Domain Controllers



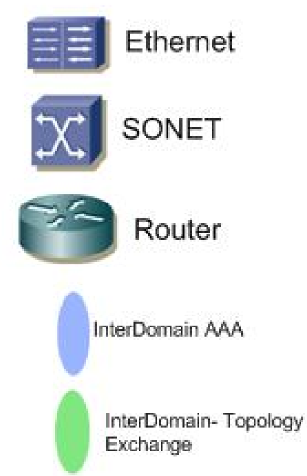
Domain by Domain Authorization, Path Computation, Provisioning (Signaling)



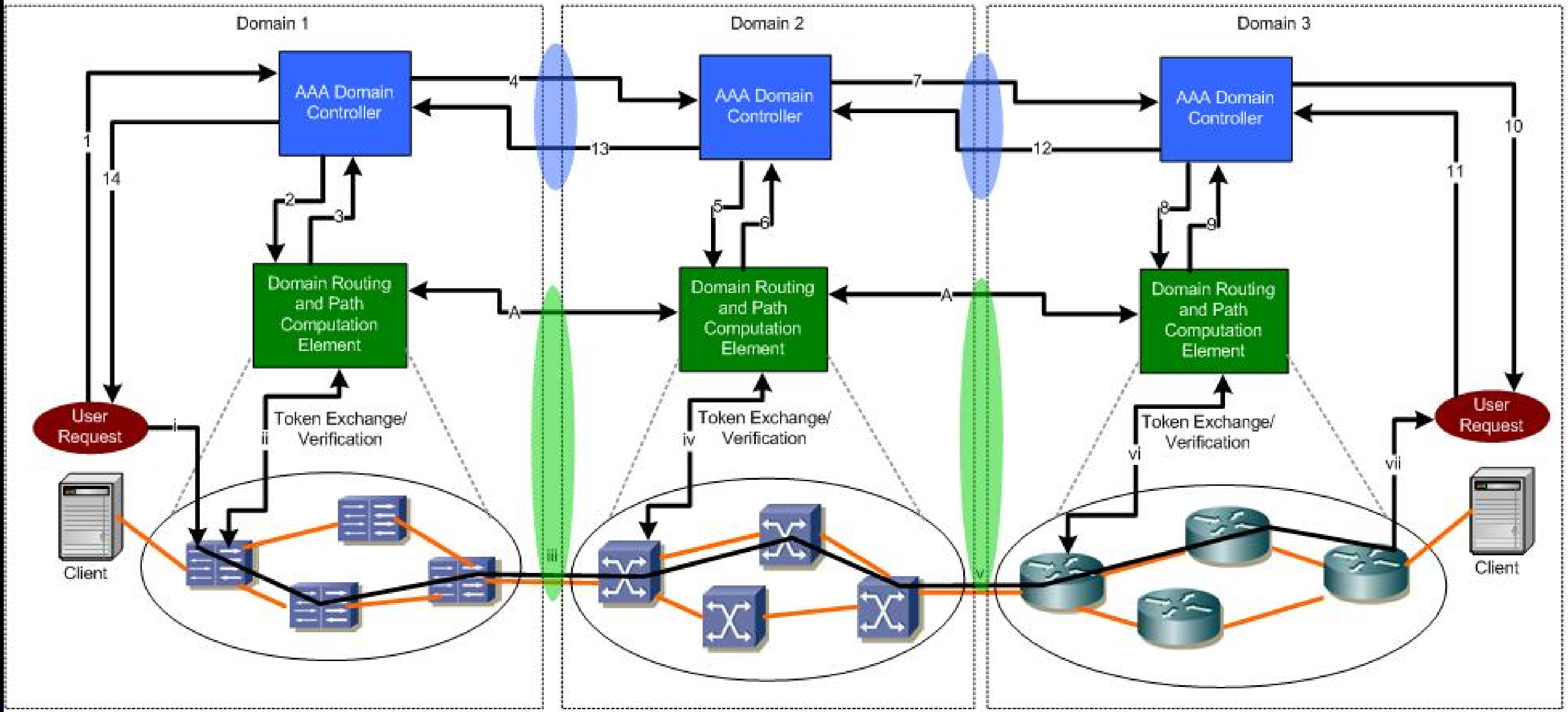
Path Authorization, Computation, and Provisioning (Signaling)

1. Client Request (Source, Destination, Bandwidth, Time, User Certificate)
2. Client Request Approval (Token Provided)
3. Multi-Domain Path Computation Result (based on abstracted topology)
4. Inter-Domain Request (Source, Destination, Bandwidth, Time, Domain1 Certificate, Path Computation Result, Approval Indicated for Domain 1)
5. Path Request Approval (Token Provided)
6. Single Domain Path Computation Approval (details held in DRPCE)
7. Inter-Domain Request (Source, Destination, Bandwidth, Time, Domain1 Certificate, Path Computation Result, Approval Indicated for Domain 1 and Domain 2)
8. Path Request Approval (Partial Token Provided)
9. Single Domain Path Computation Approval (details held in DRPCE)
10. Remote Client Request
11. Remote Client Response
12. Provision Request (Token Provided)
13. Domain Specific Provisioning and Edge Stitching
14. Inter-Domain Response (Approval Indicated for Domain 1, Domain 2, and Domain 3)
15. Provision Request (Token Provided)
16. Domain Specific Provisioning and Edge Stitching
17. Inter-Domain Response (Approval Indicated for Domain 1, Domain 2, and Domain 3)
18. Provision Request (Token Provided)
19. Domain Specific Provisioning and Edge Stitching
20. Client Request Result

<u>Topology Exchange</u>
A. Topology Exchange
Notes: Link State, XML or ENNI like Protocols, Very Little Dynamic Information Exchanged



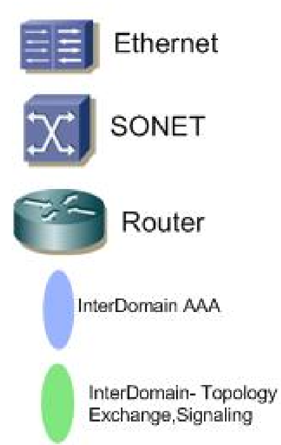
Domain by Domain Authorization, Path Computation
Continuous Provisioning (Signaling)



Path Authorization and Computation
1. Client Request (Source, Destination, Bandwidth, Time, User Certificate)
2. Client Request Approval (Token Provided)
3. Multi-Domain Path Computation Result (based on abstracted topology)
4. Inter-Domain Request (Source, Destination, Bandwidth, Time, Domain1 Certificate, Path Computation Result, Approval Indicated for Domain 1)
5. Path Request Approval (Token Provided)
6. Single Domain Path Computation Approval (details held in DRPCE)
7. Inter-Domain Request (Source, Destination, Bandwidth, Time, Domain1 Certificate, Path Computation Result, Approval Indicated for Domain 1 and Domain 2)
8. Path Request Approval (Partial Token Provided)
9. Single Domain Path Computation Approval (details held in DRPCE)
10. Remote Client Request
11. Remote Client Response
12. Inter-Domain Response (Approval Indicated for Domain 1, Domain 2, and Domain 3)
13. Inter-Domain Response (Approval Indicated for Domain 1, Domain 2, and Domain 3)
14. Client Request Result

Signaling (Provisioning)
i. Provision Request (Token Provided)
ii. Token Verification Processing
iii. Provision Request (Token Provided)
iv. Token Verification Processing
v. Provision Request (Token Provided)
vi. Token Verification Processing
vii. Remote Side Provision Processing

Topology Exchange
A. Topology Exchange
Notes: Link State, XML or ENNI like Protocols, Very Little Dynamic Information Exchanged



Key Control Plane Features
(for Connection Control)

- Routing
  - distribution of "data" between networks. The data that needs to be distributed includes reachability information, resource usages, etc
- Path computation
  - the processing of information received via routing data to determining how to provision an end-to-end path. This is typically a Constrained Shortest Path First (CSPF) type algorithm for the GMPLS control planes. Web services based exchanges might employ a modified version of this technique or something entirely different.
- Signaling
  - the exchange of messages to instantiate specific provisioning requests based upon the above routing and path computation functions. This is typically a RVSP-TE exchange for the GMPLS control planes. Web services based exchanges might employ a modified version of this technique or something entirely different.

Key Control Plane Key Capabilities

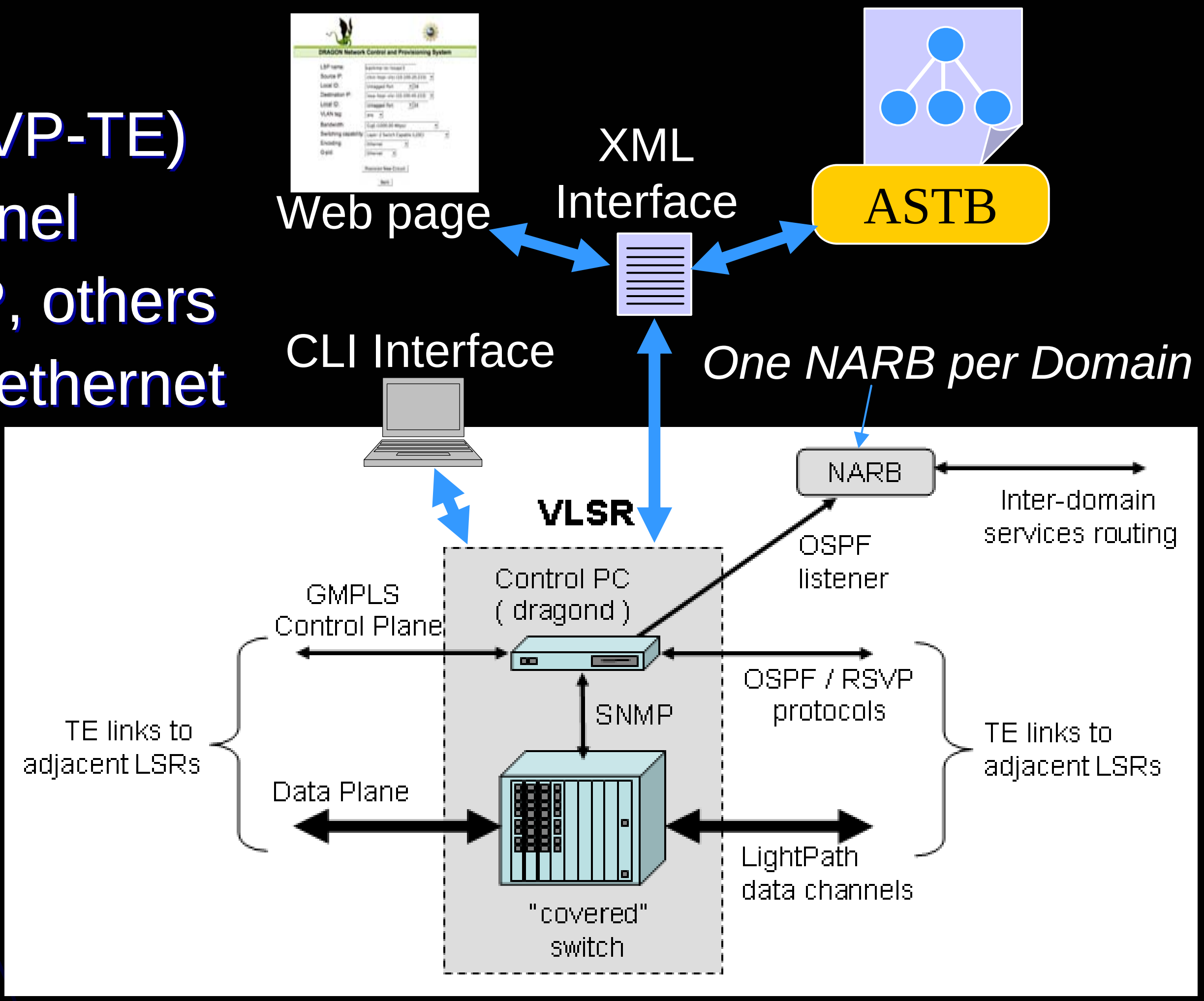
- Domain Summarization
  - Ability to generate abstract representations of your domain for making available to others
  - The type and amount of information (constraints) needed to be included in this abstraction requires discussion.
  - Ability to quickly update this representation based on provisioning actions and other changes
- Multi-layer “Techniques”
  - Stitching: some network elements will need to map one layer into others, i.e., multi-layer adaptation
  - In this context the layers are: PSC, L2SC, TDM, LSC, FSC
  - Hierarchical techniques. Provision a circuit at one layer, then treat it as a resource at another layer. (i.e., Forward Adjacency concept)
- Multi-Layer, Multi-Domain Path Computation Algorithms
  - Algorithms which allow processing on network graphs with multiple constraints
  - Coordination between per domain Path Computation Elements

DRAGON Control Plane Key Components

- Network Aware Resource Broker – NARB
  - Intradomain listener, Path Computation, Interdomain Routing
- Virtual Label Swapping Router – VLSR
  - Open source protocols running on PC act as GMPLS network element (OSPF-TE, RSVP-TE)
  - Control PCs participate in protocol exchanges and provisions covered switch according to protocol events (PATH setup, PATH tear down, state query, etc)
- Client System Agent – CSA
  - End system or client software for signaling into network (UNI or peer mode)
- Application Specific Topology Builder – ASTB
  - User Interface and processing which build topologies on behalf of users
  - Topologies are a user specific configuration of multiple LSPs

VLSR
(Virtual Label Switching Router)

- GMPLS Proxy
  - (OSPF-TE, RSVP-TE)
- Local control channel
  - CLI,TL1, SNMP, others
- Used primarily for ethernet switches
- Provisioning requests via CLI, XML, or ASTB

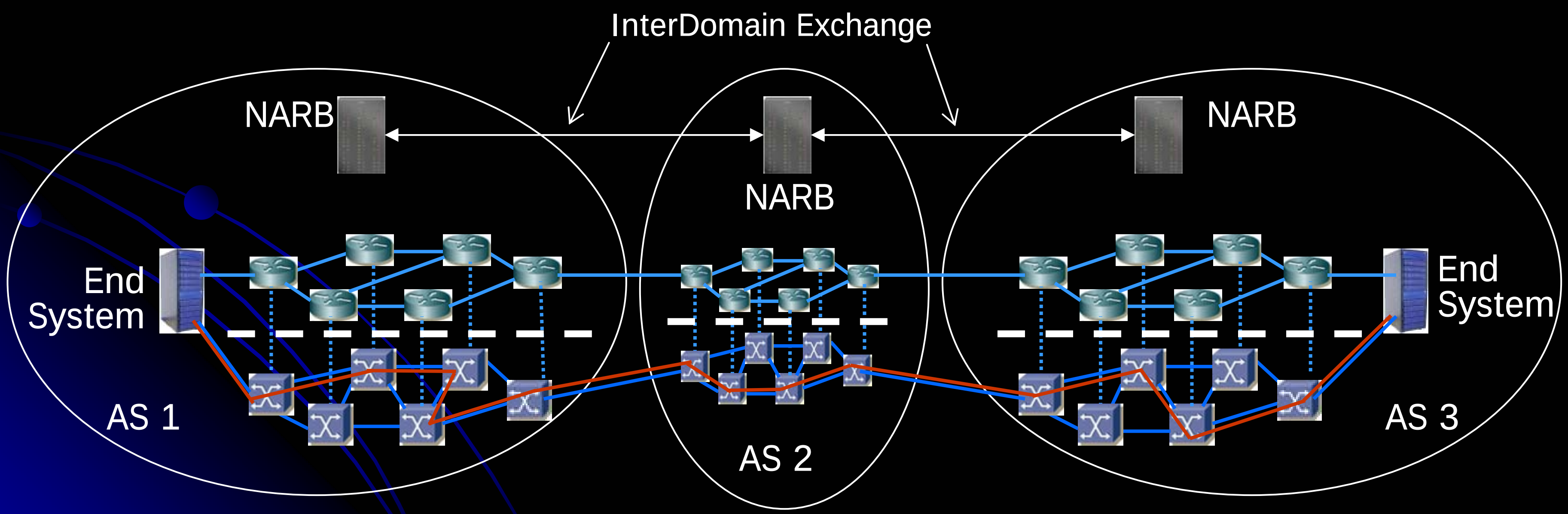


VLSR
(Virtual Label Switching Router)

- RSVP Signaling module
  - Originated from Martin Karsten's C++ KOM-RSVP
  - Extended to support RSVP-TE (RFC 3209)
  - Extended to support GMPLS (RFC 3473)
  - Extended to support Q-Bridge MIB (RFC 2674)
  - For manipulation of VLANs via SNMP (cross-connect)
  - Extended to support VLAN control through CLI
- OSPF Routing module
  - Originated from GNU Zebra
  - Extended to support OSPF-TE (RFC 3630)
  - Extended to support GMPLS (RFC 4203)
- Ethernet switches tested to date
  - Dell PowerConnect, Extreme, Intel, Raptor, Force10

NARB
Network Aware Resource Broker

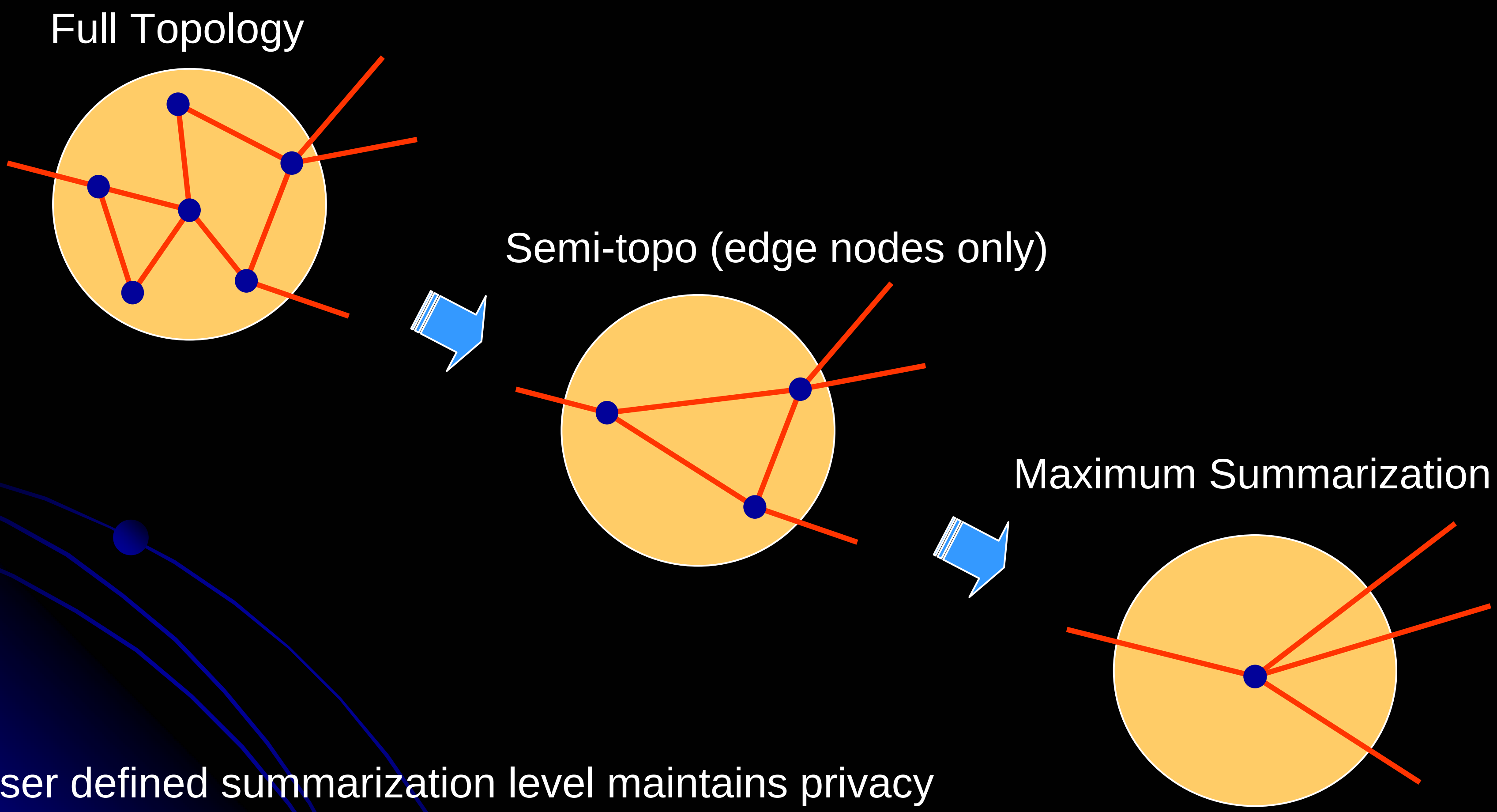
- Interdomain Routing
  - hierarchical link state
- Carries a modified TEDB that can support
  - AAA
  - Scheduling
- Path Computation Element and ERO (loose and strict) generation



NARB
(Network Aware Resource Broker)

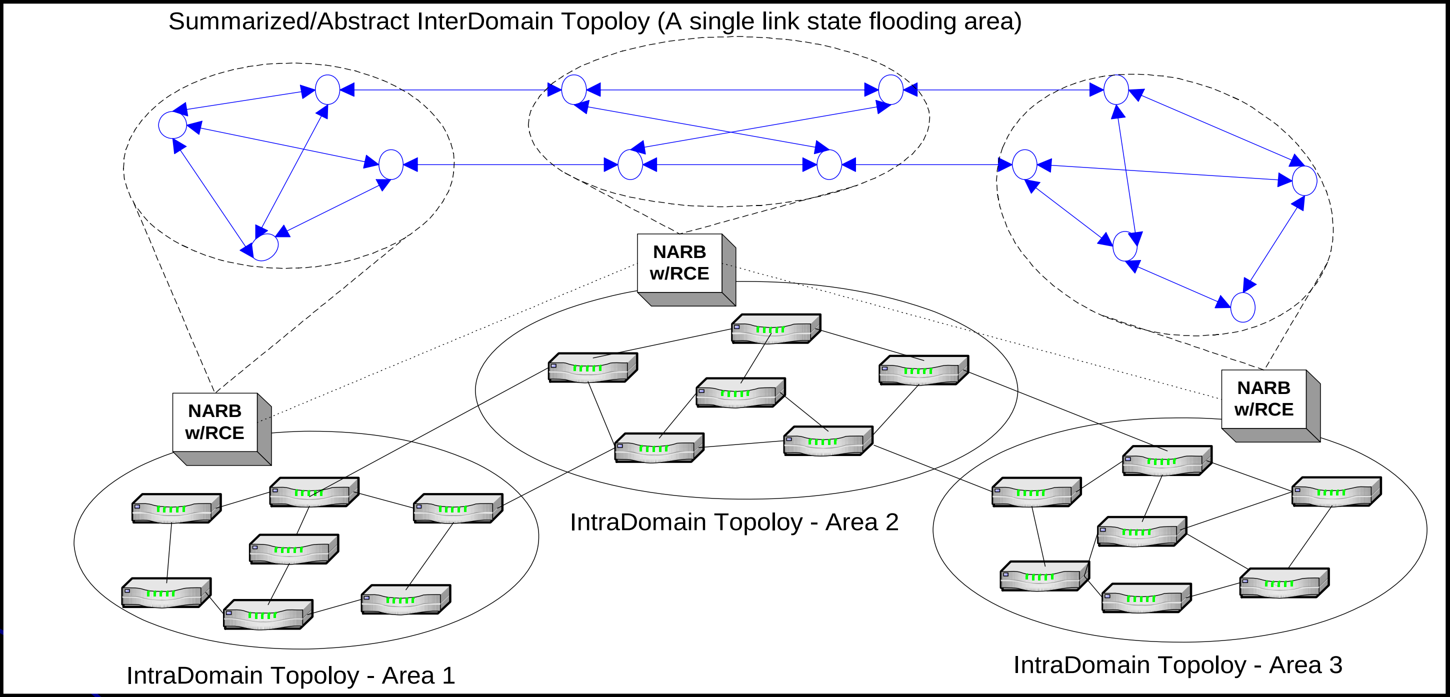
- NARB is an agent that represents a domain
- Intra-domain Listener
  - Listens to OSPF-TE to acquire intra-domain topology
  - Builds an abstracted view of internal domain topology
- Inter-domain routing
  - Peers with NARBs in adjacent domains
  - Exchanges (abstracted) topology information
  - Maintains an inter-domain link state database
- Path Computation
  - Performs intra-domain (strict hop) TE path computation
  - Performs inter-domain (loose hop) TE path computation
  - Expands loose hop specified paths as requested by domain boundary (V)LSRs.
- Hooks for incorporation of AAA and scheduling into path computation via a "3 Dimensional Resource Computation Engine (3D RCE)"
  - The Traffic Engineering DataBase (TEDB) and Constrained Shortest Path Computation (CSPF) are extended to include dimensions of GMPLS TE parameters, AAA constraints, and Scheduling constraints.
  - 3D RCE is the combination of 3D TEDB and 3D CSPF
  - http://dragon.east.isi.edu/data/dragon/documents/dragon-infocom-APBM-workshop-apr282006.pdf

Inter-Domain Topology Summarization



- User defined summarization level maintains privacy
- Summarization impacts optimal path computation but allows the domain to choose (and reserve) an internal path

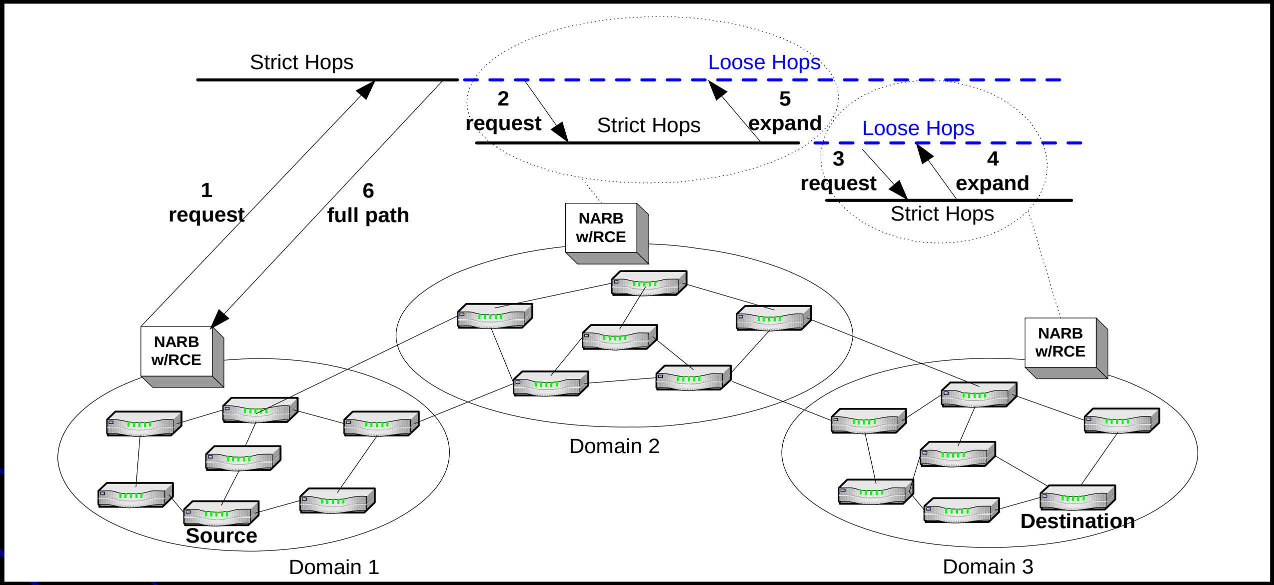
Interdomain Path Computation
A Hierarchical Architecture



- NARB summarizes individual domain topology and advertise it globally using link-state routing protocol, generating an abstract topology.
- RCE computes partial paths by combining the abstract global topology and detailed local topology.
- NARB's assemble the partial paths into a full path by speaking to one another across domains.

E2E Multi-Domain Path Computation Scheme

DRAGON mainly uses Recursive Per-Domain (RPD) interdomain path computation

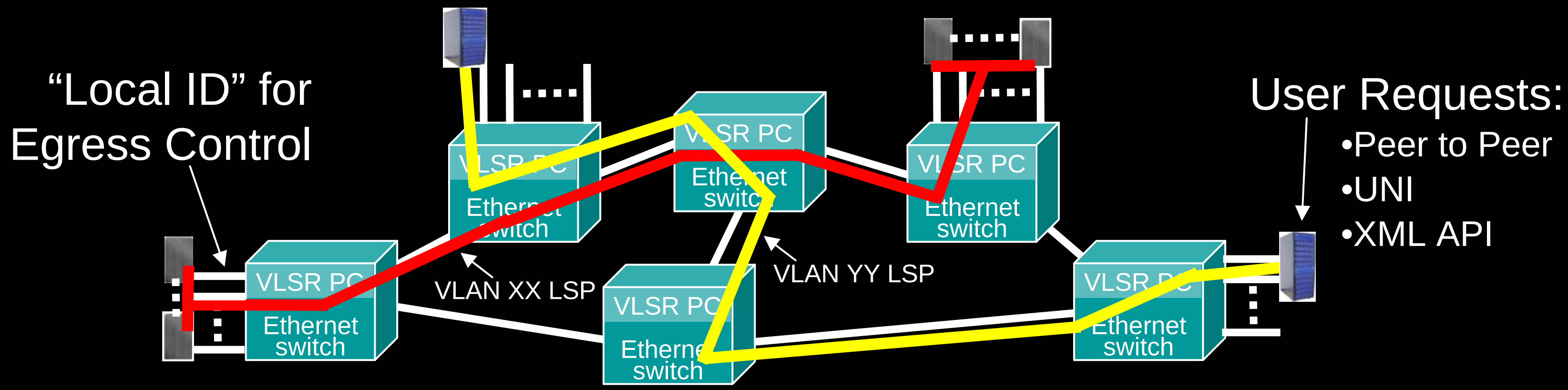


- Full explicit path is obtained before signaling.
- Other supported schemes include Centralized path computation and Forward Per-Domain (FPD) path computation.

DRAGON CSPF Path Computation Heuristics

- A breadth first search based CSPF heuristic in deployment
  - Takes flexible combination of various constraints, such as bandwidth, switch cap., wavelength, VLAN tag and add-on policy constraints.
  - Supports multi-region networks using configurable region-crossing criteria
  - Reliable results; probably time-consuming in large networks (~30ms in the 12-node HOPI+DRAGON network)
- Other heuristics under research; one is based on a channel-graph model in combination with K-shortest path routing.

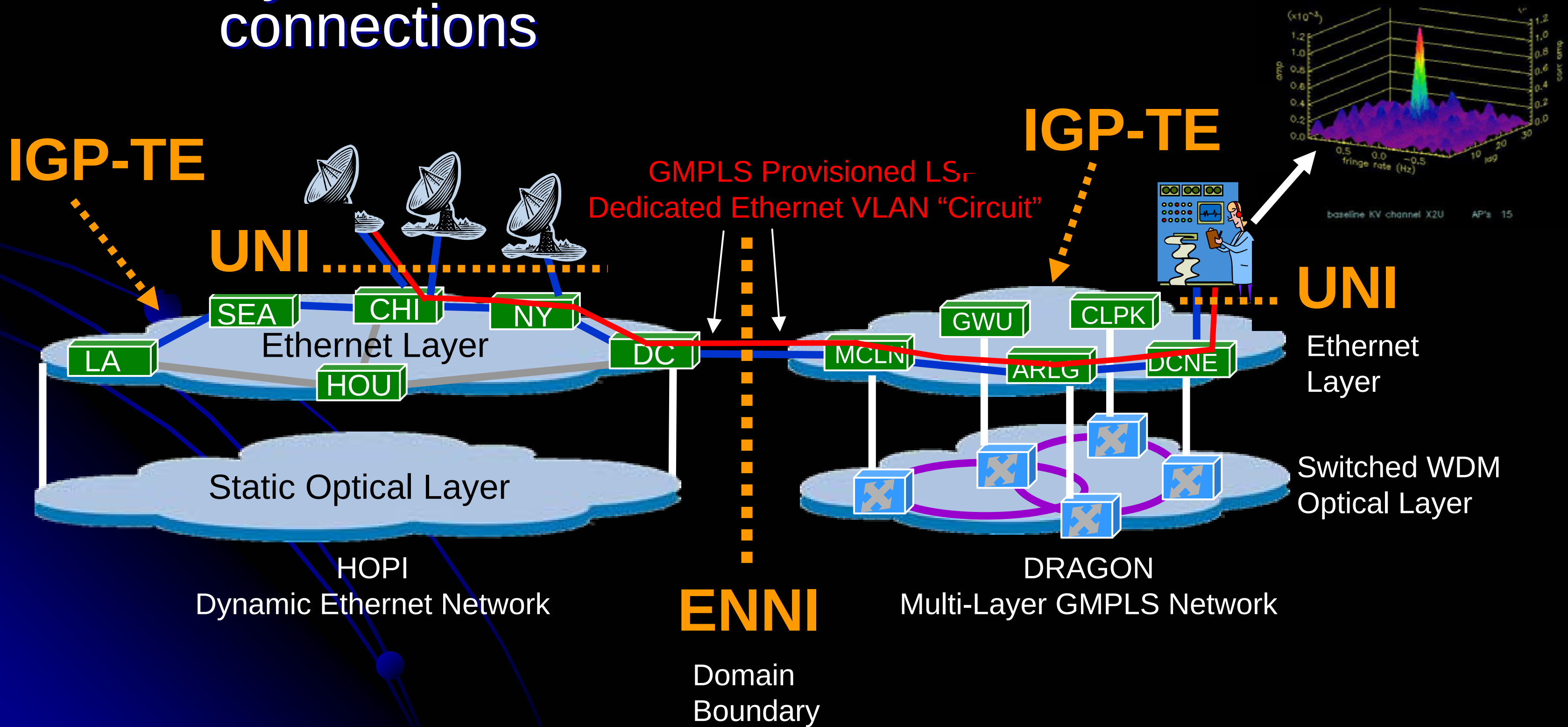
GMPLS Provisioned Ethernet Services



- Multiple Ethernet Provisioning Options
- Point to Point Ethernet VLAN based LSPs
- Ethernet switch (vendor specific) features applied to guarantee LSP bandwidth in increments of 100 Mbit/s
- Edge connection flexibility provided by use of "Local ID" feature which allows flexible combinations of one port, multiple ports, tagged ports, and untagged ports to be glued on to end of LSP. Can be dynamically adjusted.
- Users can request services via Peer to Peer GMPLS, UNI style GMPLS, or via an XML application interface
- Ethernet VLAN space is "flat" across provisioned space. Constrained based path computation utilized to find available VLAN Tags.
- VLAN tags treated in a similar manner to wavelengths

DRAGON/HOPI Control Plane Provisioning Environment

- GMPLS Multi-layer, Multi-Domain
- Ethernet Service Provisioning
- Dynamic dedicated VLAN based connections



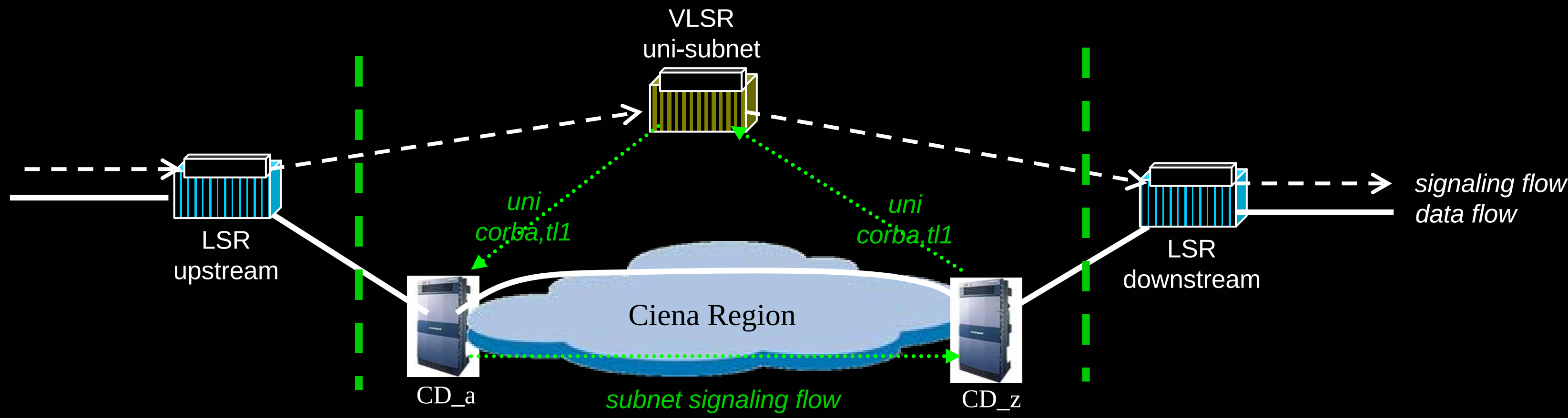
What is the HOPI Service?

- Physical Connection:
  - 1 or 10 Gigabit Ethernet
- Circuit Service:
  - Point to Point Ethernet VLAN Circuit
  - Tagged or Untagged VLANs available
  - Bandwidth provisioning available in 100 Mbps increments
- How do Clients Request?
  - Client must specify [VLAN ID|ANY ID|Untagged], SRC Address, DST Address, Bandwidth
  - Request mechanism options are GMPLS Peer Mode, GMPLS UNI Mode, Web Services, phone call, email
  - Application Specific Topology is a user specific instantiation of multiple individual circuits
- What is the definition of a Client?
  - Anyone who connects to an ethernet port on an HOPI Force 10 Switch; could be RONS, GIgaPops, other wide area networks, end systems

What is the Internet2 DCS Service?

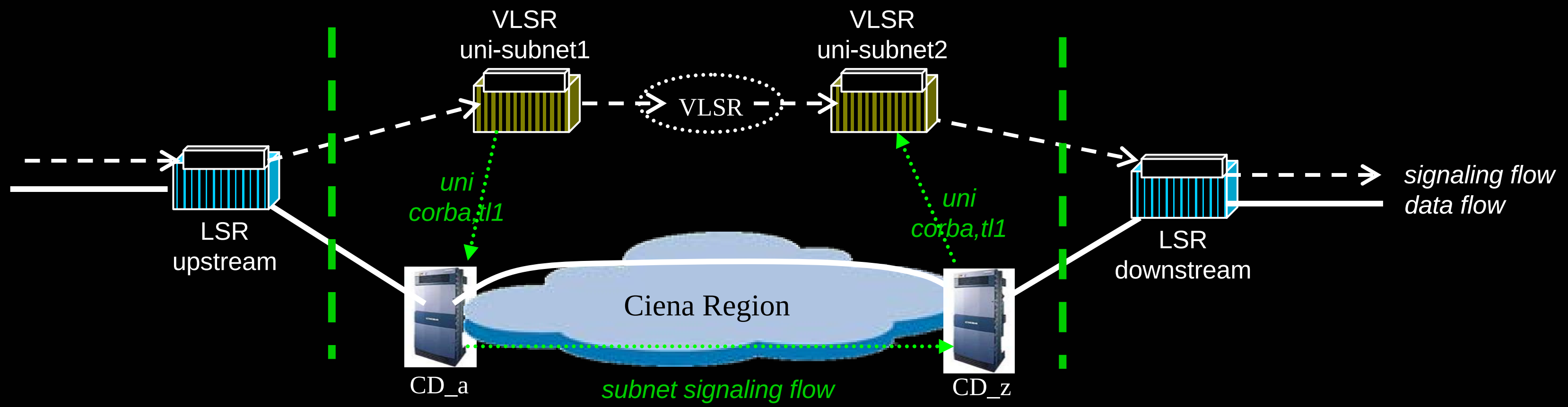
- Physical Connection:
  - 1 or 10 Gigabit Ethernet
  - OC192 SONET
- Circuit Service:
  - Point to Point Ethernet VLAN Circuit
  - Point to Point Ethernet Framed SONET Circuit
  - Point to Point SONET Circuit
  - Bandwidth provisioning available in 50 Mbps increments
- How do Clients Request?
  - Client must specify [VLAN ID|ANY ID|Untagged], SRC Address, DST Address, Bandwidth
  - Request mechanism options are GMPLS Peer Mode, GMPLS UNI Mode, Web Services, phone call, email
  - Application Specific Topology is a user specific instantiation of multiple individual circuits
- What is the definition of a Client?
  - Anyone who connects to an ethernet or SONET port on an Ciena Core Director; could be RONS, GIgaPops, other wide area networks, end systems

Integration Core Director Domain into the End-to-End Signaling



- Signaling is performed in contiguous mode.
  - Single RSVP signaling session (**main session**) for end-to-end circuit.
  - Subnet path is created via a separate RSVP-UNI session (**subnet session**), similar to using SNMP/CLI to create VLAN on an Ethernet switch.
- The simplest case: one VLSR covers the whole UNI subnet.
  - VLSR is both the source and destination UNI clients.
  - This VLSR is control-plane '**home VLSR**' for both CD_a and CD_z.
  - UNI client is implemented as embedded module using KOM-RSVP API.

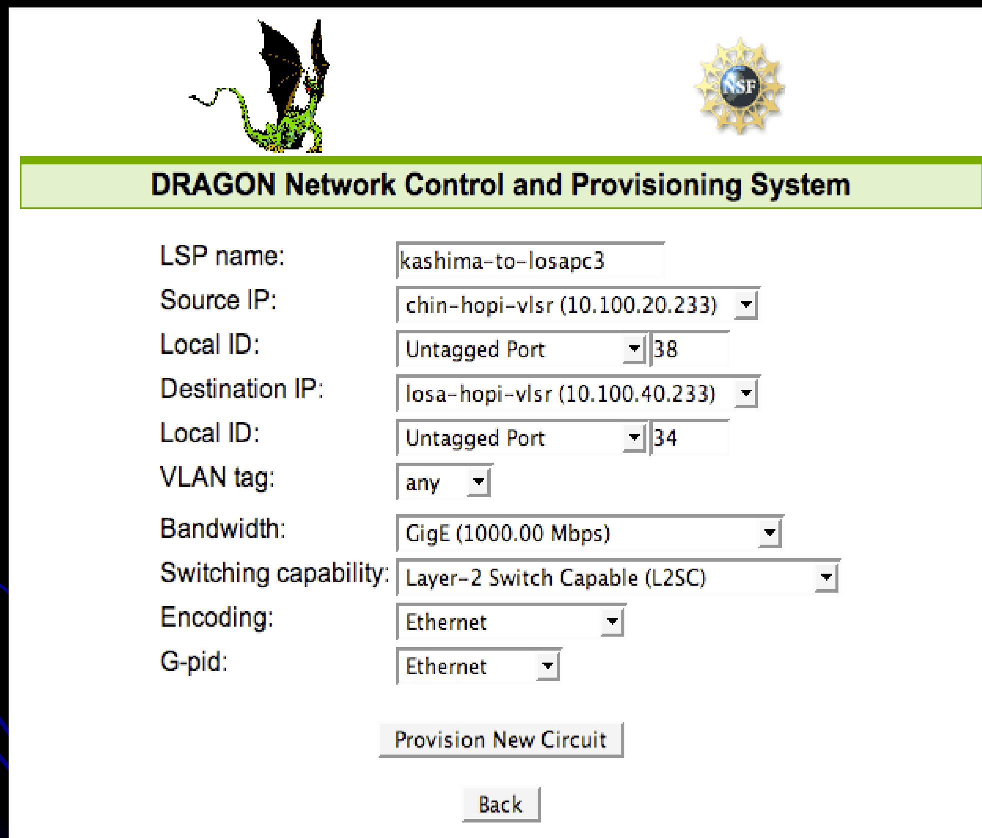
Integration Core Director Domain into the End-to-End Signaling



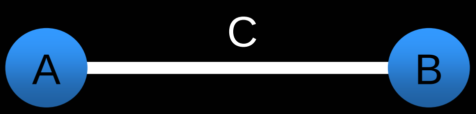
- Other configuration options: Multiple VLSRs divide the Ciena Region.
  - Source and destination UNI clients belong to two separate VLSRs.
  - The two VLSRs collaborate in a subnet RSVP UNI session.
  - CD_a and CD_z have different control-plane 'home VLSRs.'
    - Definition – *Each subnet switch (e.g., CD) has a control-plane '**home VLSR**,' which initiates a subnet signaling session as the source RSVP UNI client.*
  - There could be some intermediate VLSRs between the source and destination home VLSRs in the subnet.
    - *These '**transit VLSRs**' are home for neither CD_a nor CD_z. They pass the main RSVP session messages while not involved in the subnet session.*

DRAGON
Provisioning Web Page

Web Page Interface



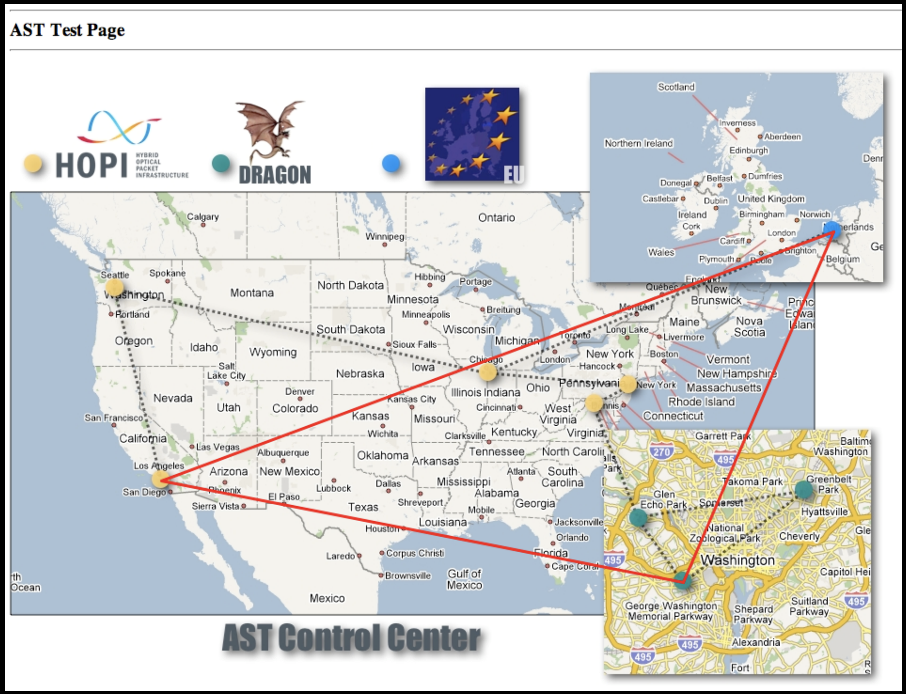
Application Specific Topologies using XML




```
<topology>
   <resource>
        <resource_type> eVLBI.Mark5a          </resource_type>
        <name>    Haystack.muk1               </name>
        <ip_addr> muk1.haystack.mit.edu       </ip_addr>           A
        <te_addr> muk1-ge0.haystack.mit.edu </te_addr>
        <appl>    /usr/local/evlbi_script     </appl>
    </resource>
   <resource>
        <resource_type> eVLBI.Mark5a          </resource_type>    B
        <name>    Westford1                   </name>
        <ip_addr> wstf.haystack.mit.edu       </ip_addr>
        <te_addr> wstf-ge0.haystack.mit.edu </te_addr>
        <appl>    /usr/local/evlbi_script     </appl>
    </resource>
   <resource>
        <resource_type> EtherPipeBasic        </resource_type>
        <src>           Haystack.muk1         </src>               C
        <dest>          Westford.muk1         </dest>
        <datarate>      1 Gbs                 </datarate>
        </resource>
</topology>
```

Application Specific Topologies

- Identify endpoints, control plane sets up topologies
- Set up global multi-link topologies
  - ~30 seconds

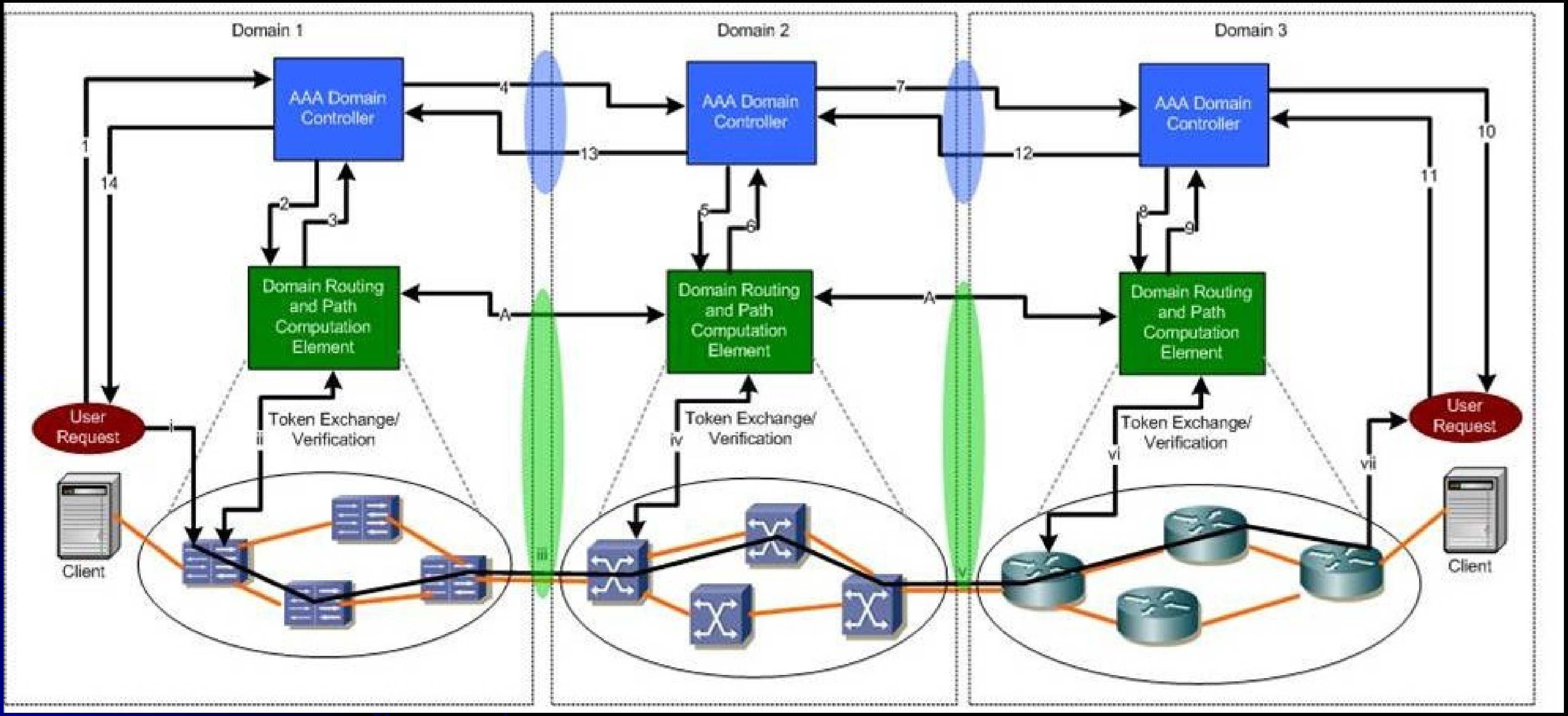


What About Web Services?

- Some domains will prefer a Web Service style interdomain messaging
- The most basic web service needed is (abstracted) topology representation
  - OGF (Open Grid Forum) Network Mark Up Language (NML) Working Group to work on this
    - Network Description Language (NDL)
    - PerfSonar
  - Community needs to agree on a schema(s)

AAA Layer and InterDomain Provisioning

- Combination of AAA Layer on top of Routing, Path Computation Element → Domain Controller
- OSCARS/BRUSCI derived AAA Layer



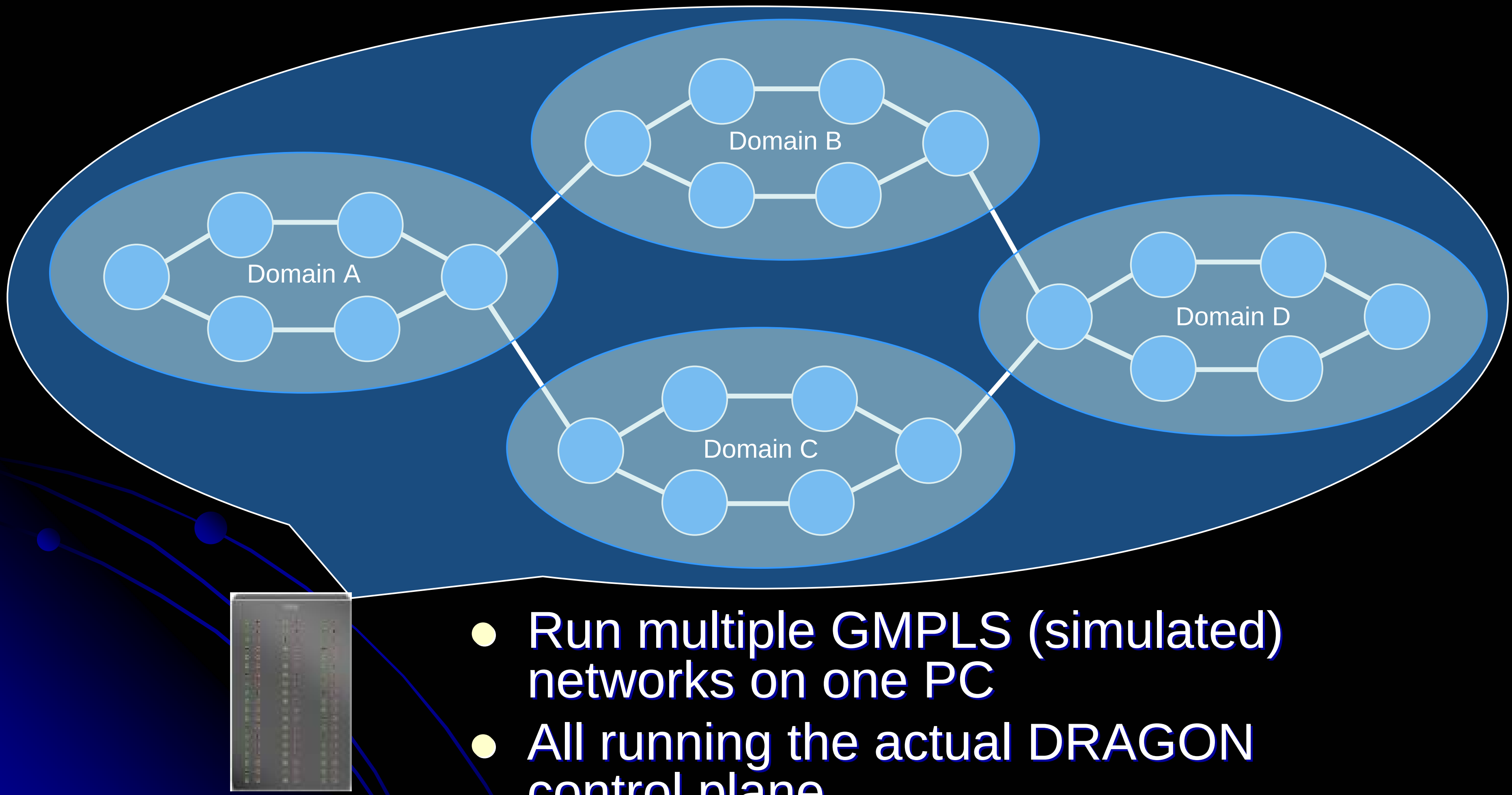
Working with other Groups on Interoperable Control Planes

- ESnet
  - OSCARS based provisioning
- DICE
  - DANTE, Internet2, Canarie, ESnet
  - GEANT JRA3
- University of Amsterdam, SurfNet
- Working to develop agreements on InterDomain Routing, Signaling, Path Computation Techniques and Messaging
  - Web Service Based
  - Protocol Based

Other Communities also working on Dynamic Resource Allocation

- ESNet Science Data Network (SDN) and the OSCARS project
- DANTE/GEANT JRA3 project
- Netherlands SURFnet and collaboration with Nortel on the DRAC project
- University of Amsterdam, Network Description Language
- European Union PHOSPHORUS Project
- G-Lambda project (Japan)
- Enlightened Computing project

DRAGON plus User Mode Linux

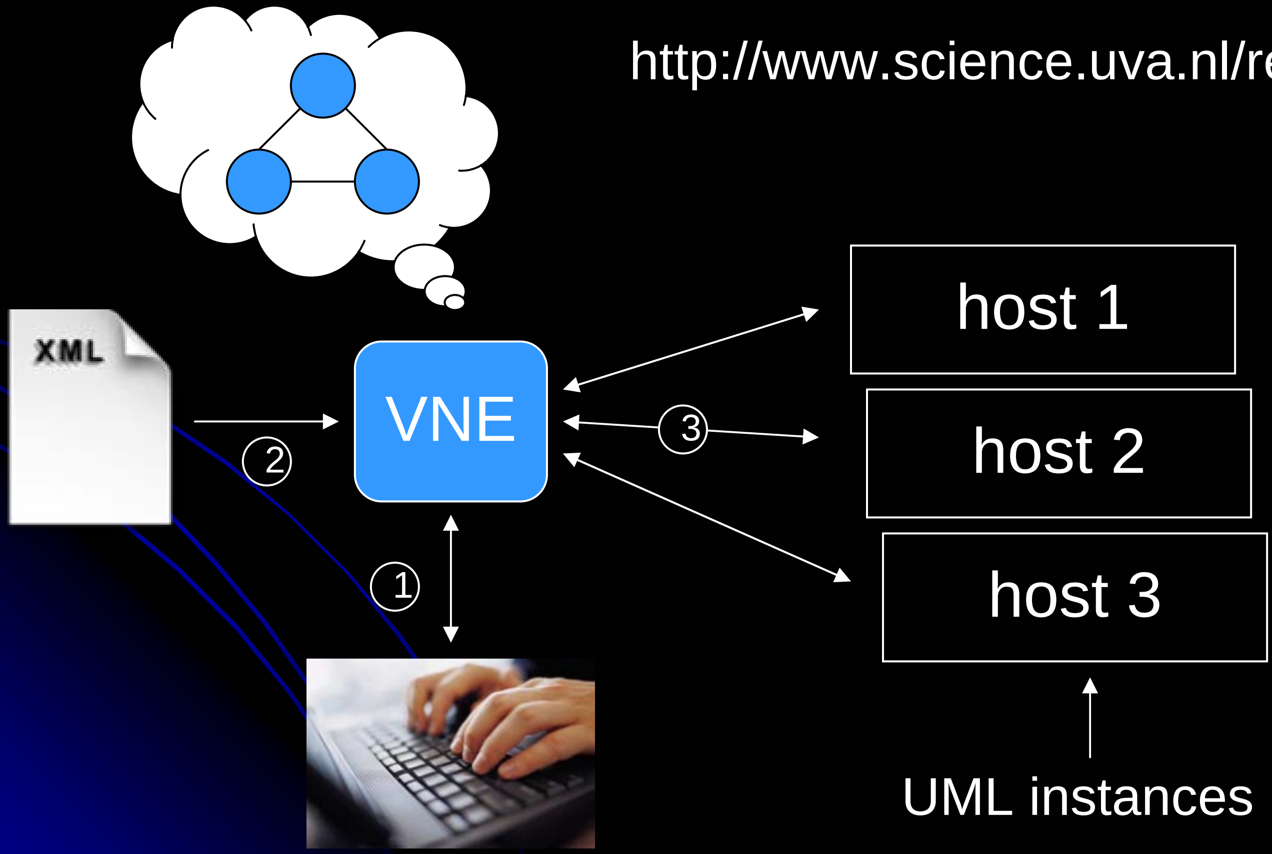


- Run multiple GMPLS (simulated) networks on one PC
- All running the actual DRAGON control plane
- Useful for learning, scalability tests, regression tests, etc.

Virtual Network Experiments

- python package developed by UvA’s Systems and Networking Engineering research group provides rich set of UML management mechanisms

http://www.science.uva.nl/research/sne/



Dragon plus User Mode Linux

- Contact Chris Tracy at MAX for more information
  - chris@maxgigapop.net
- HOWTO is being assembled now:
  - http://dragon.east.isi.edu
  - Click on “User Mode Linux”
  - Work in progress, comments/questions welcome!
- Current XML file examples include:
  - Single domain w/o domain controller
  - Single domain /w domain controller
  - Inter-domain with multiple domain controllers
  - UNI provisioning examples

Thank You

Questions/Comments?:

Tom Lehman
tlehman@isi.edu

DRAGON Web Sites
dragon.maxgigapop.net
dragon.east.isi.edu