Outline

• DRAGON Project Overview
• Research and Education Networks Requirements and Overview
• DRAGON Control Plane Architecture/Design
• Relationship to Standards Activities
• Open Issues and Future Directions
DRAGON
Single-Slide Overview

• Principal Investigators
  – Jerry Sobieski - Mid-Atlantic Crossroads (MAX)
  – Tom Lehman - USC/ Information Sciences Institute (ISI East)
  – Bijan Jabbari - George Mason University (GMU)
  – Don Riley – University of Maryland

• Commercial Partner – MOVAZ Networks

• NSF Funded program
  – Testbed deployed in the Washington DC region

• GMPLS based control plane
  – Dynamic provisioning across heterogeneous network technologies
    • Fiber (FSC), Lambda (LSC), SONET (TDM), Ethernet (L2SC), Packet (LSC)
    • Multi-layer Traffic Engineering
  – Interdomain Provisioning (routing, path computation, signaling)
  – Authentication, Authorization, Accounting (AAA)
  – Scheduling

• http://dragon.maxgigapop.net
The DRAGON Testbed
Washington, DC, USA metro region
The Emerging Global R&E Application

Emerging large-scale, globally distributed applications require more sophisticated network services than have previously been available:

- **Dedicated network resources**
  - An application needn’t worry about its impact on other network users, or vice versa

- **Deterministic performance**
  - Repeatable and predictable from day to day / year to year
  - Very high performance…
  - Multi-Gbs flows, low latency/loss, minimal jitter, global reach

- **Reservable and schedulable in advance**
  - Particularly in conjunction with availability of non-network resources (e.g. radio telescopes, computational clusters, etc.)

- **Flexible and dynamic**
  - Able to acquire these dedicated network resources on short notice from many potential service/resource providers

- **E-science services growth**
  - Many applications: physics, astronomy, climate change, biology, others
  - Petabyte demands, high growth projections (50-200%)
  - Globally distributed applications and collaborations
Very Long Baseline Interferometry
“E-VLBI”

Radio Telescopes
2005 = 512 Mbs
2007 = 2 Gbs
2009 > 4+ Gbs

Aggregated streams at correlator:
2005 > 2 Gbs
2007 ~ 10 Gbs to 20+ Gbs
2009 > 20 Gbs to 40+ Gbs
A "VLBI" Application Specific Network

- Telescopes connect to intermediate realtime storage/spooling facilities
  - These storage facilities may be a) at the telescope, b) at the correlator, or c) somewhere else logistically useful.
R&E “Hybrid” Networks

Multi-Service, Multi-Layer, Multi-Domain

- One “infrastructure” which provides basic IP routed service as well services at lower layer
  - i.e., connectionless and connection oriented services
- Services could be point to point circuits or application specific layer2 multipoint broadcast domains
- Interoperable architectures & control planes needed
- Re-use/adapt existing protocol standards, networks
- Integration challenges (control, data, management planes)
- Multi-layer adaptations “horizontal” for multi-domain
- Multi-layer adaptations “vertically” for traffic grooming
Heterogeneous Network Technologies
Complex End to End Paths

“horizontal” multi-layer adaptations for multi-domain
Multi-Layer GMPLS Networks

“vertical” multi-layer adaptations for traffic grooming, multiple services, multiple “virtual” networks
Research Networks & Testbeds
Some Examples

Energy Sciences Net

Ultra Science Net

National Lambda Rail

DRAGON

Internet 2 / HOPI / Abilene

Pan-European
Multi-Layer Infrastructures

Application Layers

- Multi-media (VoIP, HDTV)
- Storage, data archive, mirroring, peer-peer
- E-science, grid, virtualization
- Virtual reality, data fusion / visualization

Diversified “Cyber-Infrastructures”

- Layer 3: IPv4, IPv6, MPLS
- Layer 2: Ethernet, ATM
- Layer 1.5: SONET, GFP, VCAT, LCAS
- Layer 1: DWDM

Networks:
- ESNet + OSCARS
- Abilene + BRUW
- Ultra Science Net
- CHEETAH
- HOPI
- DRAGON
Multi-Layer / Multi-Domain Focus

**Scale Services Across Layers**

**Unified Inter-Layer Architecture**

- **Resource Discovery**
  - Hierarchical routing
  - Multi-layer database
  - Legacy domain (proxy)
  - Temporal state

- **Path Comp, Scheduling**
  - Dist'd / centralized
  - Domain controllers
  - Path composition
  - Adv. scheduling

- **Signaling & Recovery**
  - Multi-layer LSP: Stitching, merging
  - Multi-layer recovery
  - Signaling extensions

- **Security, AAA**
  - Encryption
  - Integrity
  - Client validation

---

**Need R&D, new standards, vendor support**
Internet 2’s Hybrid Optical Packet Infrastructure (HOPI)

- Footprint across United States
- Ethernet Switches on top of National Lambda Rail (NLR) infrastructure
- Ten Gigabit/s backbone
- Multiple One and Ten Gigabit connections at sites for users, regional networks, and Abilene connection
- networks.internet2.edu/hopi
DRAGON/HOPI Control Plane Provisioning Environment

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

- Live demonstration at Internet2 Spring Member Meeting (April 2006, Washington DC)
  - See [www.internet2.edu](http://www.internet2.edu) for webcast of “HOPI update” presentation.
  (See me after this talk for personal live demo (😊))
- Set up global multi-link topologies
  - ~30 seconds
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
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.
Three Policy Dimensions in GMPLS Service Provisioning

- **Resource dimension**
  - Link availability, bandwidth capability & resource interdependence
  - TE constraints, e.g. switching cap.

- **AAA policy dimension**
  - User privileges
  - App. specific requirements (SLA)
  - Administration policies

- **Time schedule dimension**

  - Integrate and translate network resource states and policies into shared control plane intelligence.
  - Synergize AAA policy decision with TE based provisioning decision, resulting in fast, precise and simplified control process.
3 Dimensional (3D) Resource Computation Model

Resource states, time schedule and AAA policies are exchanged among control-plane entities in both intradomain and interdomain scopes.

Three dimensions of constraints are used in joint to compute which resource to allocate and generate policy decisions.

Actual service provisioning: resource allocation and policy enforcement.

GMPLS routing, path computation

GMPLS signaling
RCE is the power horse in GMPLS control-plane to perform the computation intensive resource management & policy decision tasks.

RCE can be used as a standalone server or as an integrated NARB module.
3D Constraint Based Path Computation

Data source (raw link states from intra- and inter-domain flooding) and 3D constraints

Snapshot of topology reduced by policy filters

Constraint based path computation algorithm - CSPF heuristics
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.
AAA Based Provisioning

- AAA Policy TE Link TLV
- Allows a AAA information to be included as part of path computation
- Path Computation understanding/interpretation of rules very simple
- Much work needed in this area
### Time Based Provisioning

<table>
<thead>
<tr>
<th>0</th>
<th>8</th>
<th>16</th>
<th>24</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong> = TBD</td>
<td><strong>Length</strong> = N*5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resv 1 – Start time</td>
<td>Resv 1 -</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td>Resv 2 – Start time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resv 2 - Duration</td>
<td>Resv 3 ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeated N times (N ≤ 40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Schedule TE Link TLV**
- **Allows a time constraint to be included as part of path computation**
GMPLS Provisioned Ethernet Services

“Local ID” for Egress Control

User Requests:
- Peer to Peer
- UNI
- XML API

- 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
Ethernet VLAN based Provisioning

- Local ID defines the VLAN tag/edge port mapping
  - Several options: tagged, untagged, single port, port groups, automatic
  - Local ID definitions can be adjusted dynamically
- OSPF
  - configure vlans on each interface
  - advertise out in IfSwCap Descriptor TLV inside a TE Link LSA
  - update vlans availability and bandwidth in response to provisioning
    - similar to the existing ifswcap-specific-psc and ifswcap-specific-tdm
- RSVP ERO
  - proprietary Unnumbered Interface ID Subobjects (UnNumIfID) used to encode VLAN information in ERO
  - 32-bit UnNumbered Interface ID: type(1byte):value(24bits, vlan tag info)
- NARB/RCE
  - listen to OSPF
  - path computation with bandwidth and vlan constraints
    - create EROs with UnNumIfID objects
- Driven by need to provision across HOPI (10 gigabit interfaces)
GMPLS Interdomain Routing and Signaling Solution – OIF, DRAGON

• OIF
  – Mapping IETF GMPLS protocols to ITU ASON architecture
  – Domain/sub-network boundaries drawn by E-NNI, I-NNI and UNI
  – E-NNI enforces interoperation while allowing for flexibility of vendor implementation in each domain (GMPLS or non-GMPLS)

• DRAGON
  – An open-source implementation of the GMPLS architecture
  – Extended interdomain routing and signaling capability
    • focus on Ethernet VLAN based provisioning, multi-domain, and multi-level
  – Used as platform for research and testing
GMPLS Interdomain Routing and Signaling Solution – OIF, DRAGON

• Similar in overall concept in terms of
  – use of hierarchical link state (OSPF derived) for routing
  – RSVP for signaling
• Many differences in the details
• Domain/Routing Controllers
  – OIF OSPF daemons are called Routing Controllers (RC); RC ID = Router ID
    • One or more RC in each routing domain as routing speakers for the domain
  – DRAGON has the Network Area resource Broker (NARB) as RC, which has no corresponding router and operates a dedicated instance of OSPF in a separate address space
  – Both have adjacency via IP tunnels and control communications via separate tunnel addresses
  – OIF introduces Local/Remote Node ID sub-TLV for separation of data plane from control pane (each RC can correspond to multiple routers (nodes)) and Hierarchy List sub-TLV to add vertical hierarchies to TE topology.
• Connection End Points
  – OIF UNI uses TNA w/ Node ID addresses, which introduces Reachable TNA Opaque LSA and Node ID sub-TLV into OSPF-TE advertisement
  – DRAGON uses edge router loopback IP with Local-ID, which introduces Local-ID to end users but does not add anything into the OSPF-TE
• The plan is for DRAGON be become standards compliant as they mature (with hopefully interoperation with other domains providing specific requirements)
Standards Tracking

Multi-Layer / Multi-Domain Activities

IETF WG’s
Architectures, protocols, L1 VPN

OIF Networking WG’s
UNI, NNI specifications

ITU-T SG-15, SG-13 WG
Architectures, L1 VPN

Liasion Activities
User Network Interface (UNI) 2.0
- Multi-vendor interoperable client provisioning
  Automated end-pt & service discovery, signaling (parameters)
- Improved resiliency, control security, Eth support (IETF, ITU-T inputs)
- UNI-N side supports multi-layer call/connections (VCAT)

Network to Node Interface (Internal – NNI, External - NNI)
- Decouple intra & inter-domain mechanisms (protocols, algorithms)
- Signaling protocol: parameter negotiation, protection/diversity
- Hierarchical routing: topology / resource discovery
- Generally lacks provisions for advance scheduling

IEC Supercomm interoperability trials
- Interim UNI 1.0 (2001): End-pt discovery, setup/teardown, full $\lambda$ rates
- UNI 2.0, E-NNI 1.0 (2005):
  13 vendors, 7 service providers (focus on EoS services)
Automatically-Switched Optical Network (SG - 15, G.8080)
- Multi-level hierarchical link-state routing (G.7715.x):
  Horizontal (areas), vertical (leaders), inter-level state exchange
- Distd call / connection management (G.7713.x, SN controllers):
  Recently addressing protection/restoration, no crankback yet

Layer 1 VPN (SG - 13)
- Close liason w. IETF (routing area) on suitability of IETF protocols

Other liason activities to evolve “ASON compliant” protocols
- Signaling:
  IETF RSVP-TE drafts for ASON, OIF UNI 2.0 & NNI 1.0 alignment
- Link-state routing:
  - Reqs RFC 4258, OSPF-TE and IS-IS drafts for ASON (G.7715.1)
  - OIF NNI 1.0 routing
Internet Engineering Taskforce

CCAMP working group (GMPLS)

- GMPLS control for SONET/SDH (RFC 4257)
- GFP/LCAS interface discovery (OSPF-TE, RSVP-TE implications)
- Multi-layer/multi-region (MRN) networks drafts:
  Interface switching capability (ISC), unified TE database
- Drafts on multi-domain routing (OSPF-TE, O-BGP), no temporal state
- Other drafts on multi-domain/AS signaling & recovery:
  Crankback, inter-AS exclude routes, etc

Path computation element (PCE) working group (TE)

- Path composition for TE-LSP paths:
  Centralized / distributed, loose-domain / hop-by-hop
- Inter-area / AS / layer considerations (virtual topology management)
- New PCEP signaling protocol, possibly one for PCE discovery
- No PCE considerations for advance scheduling
- Various requirements drafts (2004-5), no RFC yet
IETF Multi-Layer Network

Overview

- Networks w. multiple domains, nodes w. multiple layers
- Run single GMPLS instance (routing, signaling):
  - Multiple links in TE database (TED) w. FA-LSP, ISC
  - Node-internal links for multi-layer nodes
- Path-computation can use ISC to qualify links
- Virtual network topology (VNT) via TE links @ lower layers
- Inter-domain aspects not addressed in drafts
Layer 1 VPN working group

- "Infrastructure virtualization": DWDM lighpath, SONET circuit
- Basic and enhanced modes: signaling only vs. dist^d signaling & routing
- Drafts on BGP & OSPF PE discovery (opaque LSA), single AS focus for now
- Proposal to extend RSVP-TE signaling (per VPN instances)
- Framework draft (near last call), no RFC yet
IETF L1 VPN Service Models

Differing Levels of CE-PE Functionality / Exchange

Physical topology

Customer site A (VPN#1)

Provider network (GMPLS transport network)

Customer site B (VPN#1)

Overlay extension

Virtual Link

Virtual Node

Per VPN Peer
Accelerate “carrier-class” Ethernet

- **Service focus**, layered network decomposition:
  - Applications, Eth services, metro Eth network (MEN)
  - Agnostic to MEN technology (SONET, DWDM, MPLS)
  - UNI spec for client-MEN boundary (UNI-C, UNI-N), NNI

---

**E-Line Service, MEF 4 (2004)**

- Point-to-point (unicast) Ethernet VC (EVC)
- Service attributes (at UNI):
  - Interfaces, BW profiles, service performance, frame delivery, service multiplexing, L2 control tunneling/discard, etc
- UNI multiplexing (EVPL service)

---

**E-LAN Service, MEF 4 (2004)**

- Multipoint-to-multipoint (broadcast) EVC
  - Best-effort or QoS between UNI’s
- Similar service attributes
- Support address learning over UNI
Summary

• We have a need today for a GMPLS control plane which can operate in a multi-domain, multi-level environment with features for AAA and scheduling.
• Current implementation is being utilized to provide services to science users on multiple experimental testbeds.
• Full compliance with ASON/GMPLS standards is an important goal.
• Major challenges in the areas of higher fidelity incorporation of AAA and scheduling information, interoperation with other control planes, robust multi-level, multi-vendor operation.
Additional Information

• Dynamic Resource Allocation via GMPLS Optical Networks (DRAGON) software and documentation
  – http://dragon.east.isi.edu

• Tom Lehman, tlehman@isi.edu