The DRAGON Project

Dynamic Resource Allocation over GMPLS Optical Networks

Presented to IPOP 2005

February 21, 2005

Tokyo, JP

National Science Foundation

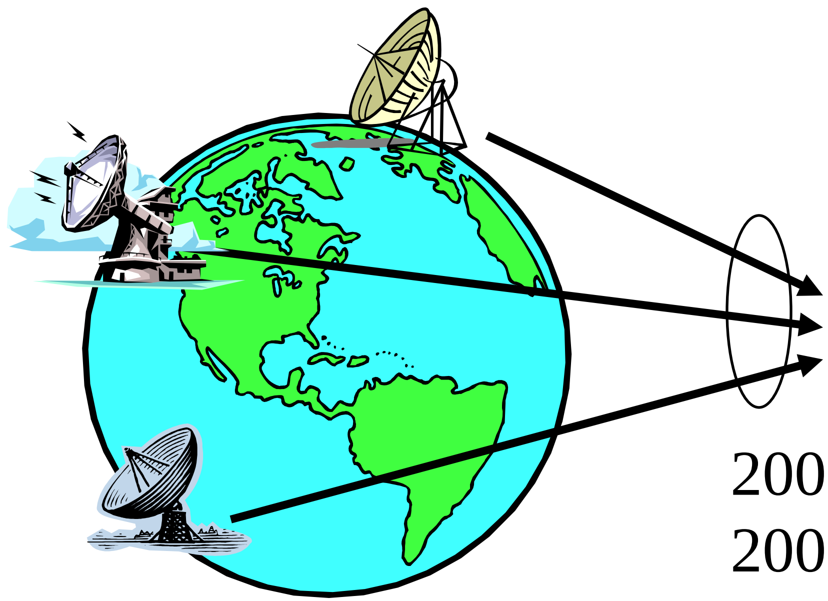
- **Jerry Sobieski**
  Mid-Atlantic Crossroads (MAX)
- **Tom Lehman**
  University of Southern California
  Information Sciences Institute (USC ISI)
- **Bijan Jabbari**
  George Mason University (GMU)
- **Don Riley**
  University of Maryland (UMD)

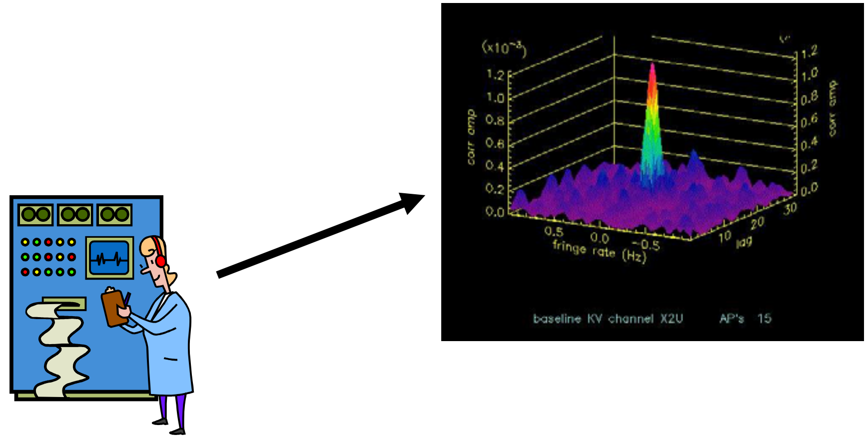
Why Develop Dynamic "Light Path" Networks?

- **There exists a new set of emerging "E-Science" applications that require new network capabilities:**
  - Global science
    - Global climate modeling, life sciences, radio astronomy, high energy physics,…
  - Global teams of scientists
  - Globally distributed tools
    - Sensors, radio telescopes, high energy physics facilities, remote field telemetry, computational/storage/visualization resources, …
- **It is difficult (or impossible) for existing networks to meet the needs of these applications**
  - Best-effort IP networks exhibit unpredictable performance for this small user community with high end specialized applications.
  - Current circuit based services are static and slow (on the order of weeks or months) to provision. These applications themselves are very dynamic in terms of topology

E-science Example: E-VLBI

- **Electronic Very Long Baseline Interferometry (eVLBI)**





2005 ~= 512mbs -> 1gbs
2006 > 10 gbs to 20+ gbs
2008 > 20 gbs to 40+ gbs

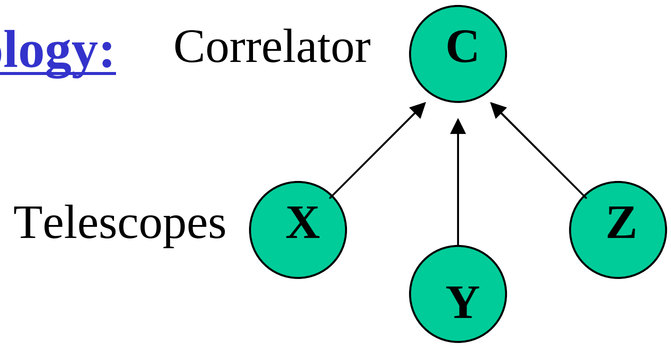
2004 = 128 mbs
2006 > 2 gbs
2008 > 4+ gbs

E-VLBI Requirements

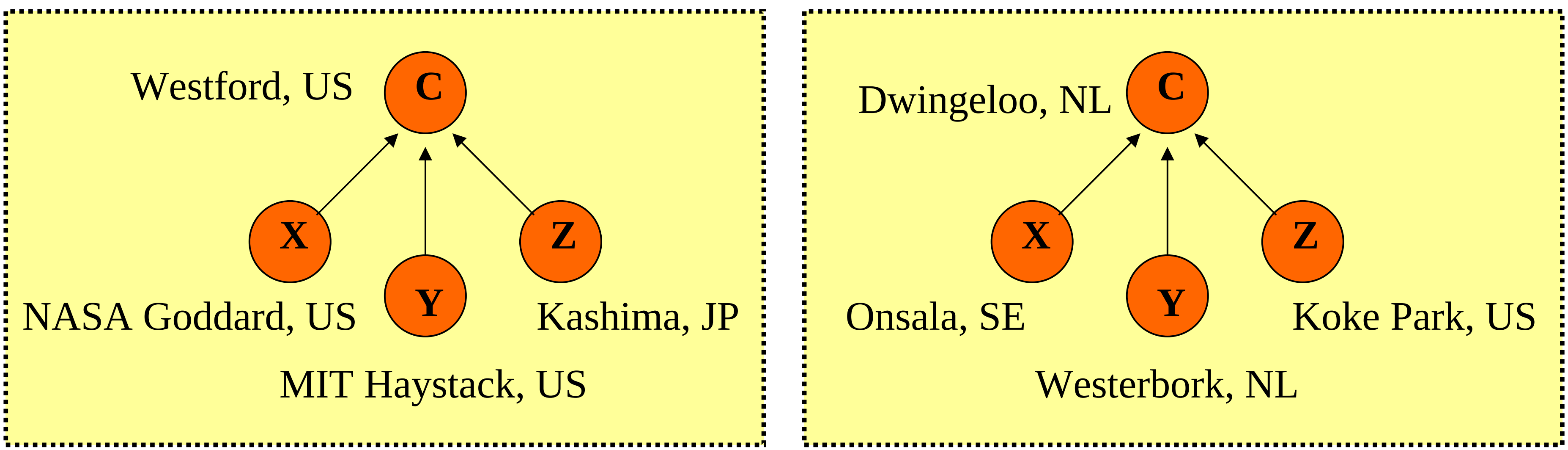
- **From 2 to as many as 20+ radio telescopes distributed around the world**
  - Each generating 500mbs real time streams (this year)
  - Converging on a single computational cluster
- **Realtime correlation requires network resources that are:**
  - Deterministic in terms of loss, latency, and jitter
  - Repeatable
  - Scheduleable – must be coordinated with other resources such as the availability of the telescope itself.
  - Rapidly provisioned, under one minute to establish the topology
- **The network is an integral part of the application itself –**
  - An application specific topology (AST) must be instantiated “en masse” to run the application – I.e. all network resources must be provisioned as a whole
  - This physical AST will vary with the location of available/required nodal resources

The “Application Specific Topology”

Logical e-VLBI Topology:

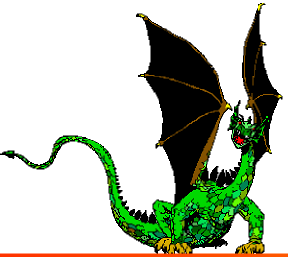


Physical Instantiations of the Application Specific Topology



What are the generic network requirements of these Super-Apps?

- **<u>Dedicated Network Resources</u>**
  - These applications want/need their own network resources, They do not care to "play fair" with other traffic.
- **<u>Deterministic Network Performance</u>**
  - Network performance must be consistent, predictable, and repeatable
- **<u>Reservable and Schedulable</u>**
  - The network must insure that the resources will be available when needed, and for as long as needed.
- **<u>Very High Performance</u>**
  - These applications require resources that often exceed current IP backbones.
- **<u>Dynamically Provisioned</u>**
  - The topologies, the performance requirements, the priorities, and purpose of the applications are not static and will vary from one run to the next.
- **<u>Application Specific Topologies</u>**
  - All resources must be allocated and provisioned as a whole.



So, what is the DRAGON Project?

- **Dynamic Resource Allocation over GMPLS Optical Networks**
  - DRAGON is a four year project funded by the US National Science Foundation (NSF)
  - Testbed deployed in the Washington DC metro area
- **Purpose:**
  - To develop/integrate network hardware and software technologies that can support dynamic, deterministic “light path” services.
  - To demonstrate these “light paths” services with real applications and over real network(s)

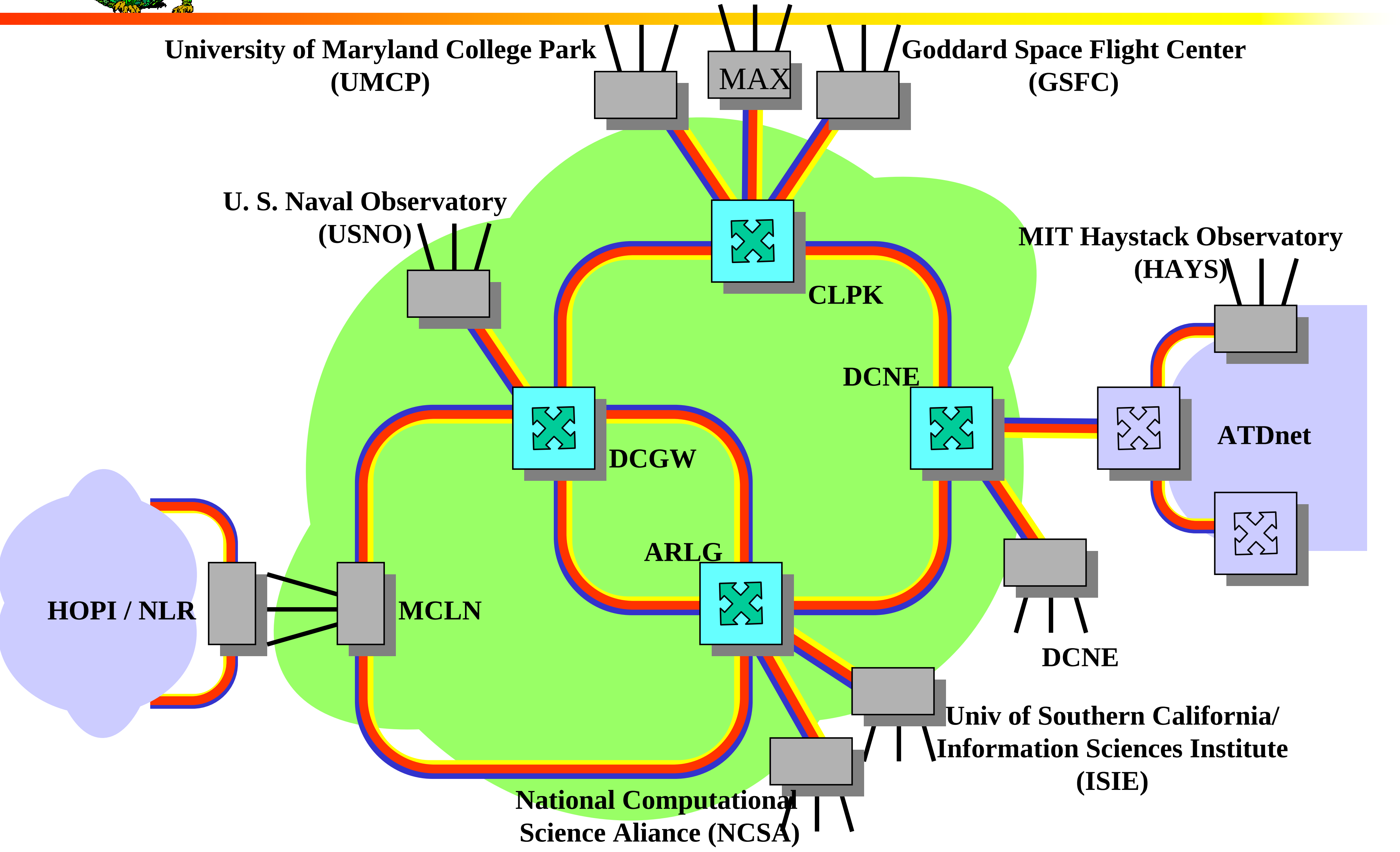
DRAGON Participants

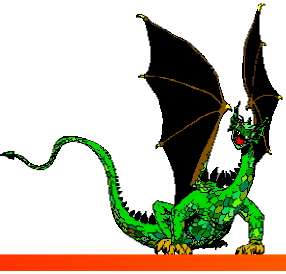
- **Mid-Atlantic Crossroads (MAX)**
- **USC/ Information Sciences Institute (ISI-East)**
- **George Mason University (GMU)**
- **MIT Haystack Observatory**
- **NASA Goddard Space Flight Center (GSFC)**
- **University of Maryland (UMCP)**
- **Movaz Networks (commercial partner)**
- **NCSA ACCESS**
- **US Naval Observatory**

Project Features and Objectives

- **All-Optical metro area network**
  - Reduce OEO in the core, allow alien Waves in.
- **GMPLS protocols for dynamic provisioning**
  - 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, D.C. Metropolitan Area

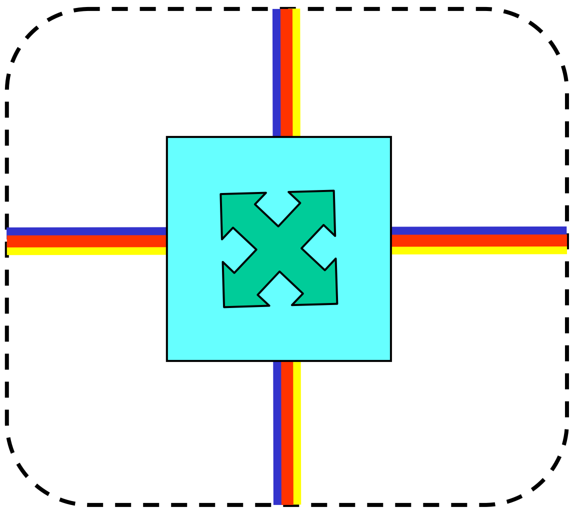


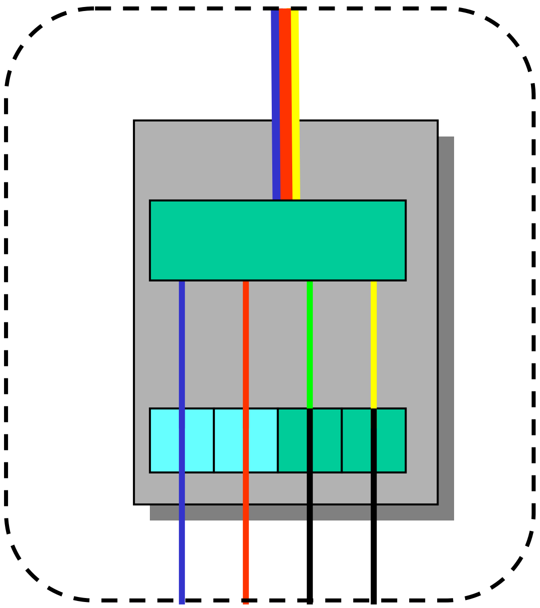


DRAGON Photonic Architecture

- **Principles:**
  - Standard practice of OEO engineering at every node is unnecessary in metro/regional networks
    - Allow the user/client to define the transport
  - Core switching nodes should be all-optical:
    - Any wavelength on any port to any other port
    - Framing agnostic
  - OEO is provisioned *only* as a service function at core nodes:
    - To provide wavelength translation to avoid wavelength blocking conditions
    - To provide regeneration iff network diameter and service specs require it, and only on a request specific basis.
  - OEO transponders are used at edge only for ITU translation
    - External ITU wavelength signaling and sourcing is encouraged
  - All waves are dynamically allocated using GMPLS protocols
    - Extensions for CSPF path computation and inter-domain are new

DRAGON Generic Architectural Cells





Core Wavelength Switching Primitive Cell

-All waves are C-Band ITU compliant on 100 Ghz ITU spacing
-Any wave can be individually switched from any input port to any output port
-Each port goes to either a) another core switching cell, or b) an edge cell
-Other wavelengths outside the C-Band are extinguished on entry and are not progressed thru the switch.
-The switching cell can block any/all input waves on any input port
-The switch is not sensitive to the content, framing of any data plane wave.

Edge Service Introduction and Validation Cell

-Client interfaces provide wavelength conversion to ITU grid lambdas
-External wavelength interfaces verify conformance of customer provisioned waves to network constraints
-Can also be used at core nodes to provide wavelength translation

Commercial Partner: Movaz Networks

- **Private sector partner for the DRAGON project**
  - Provide state of the art optical transport and switching technology
  - Major participant in IETF standards process
  - Software development group located in McLean Va (i.e. within MAX)
  - Demonstrated GMPLS conformance

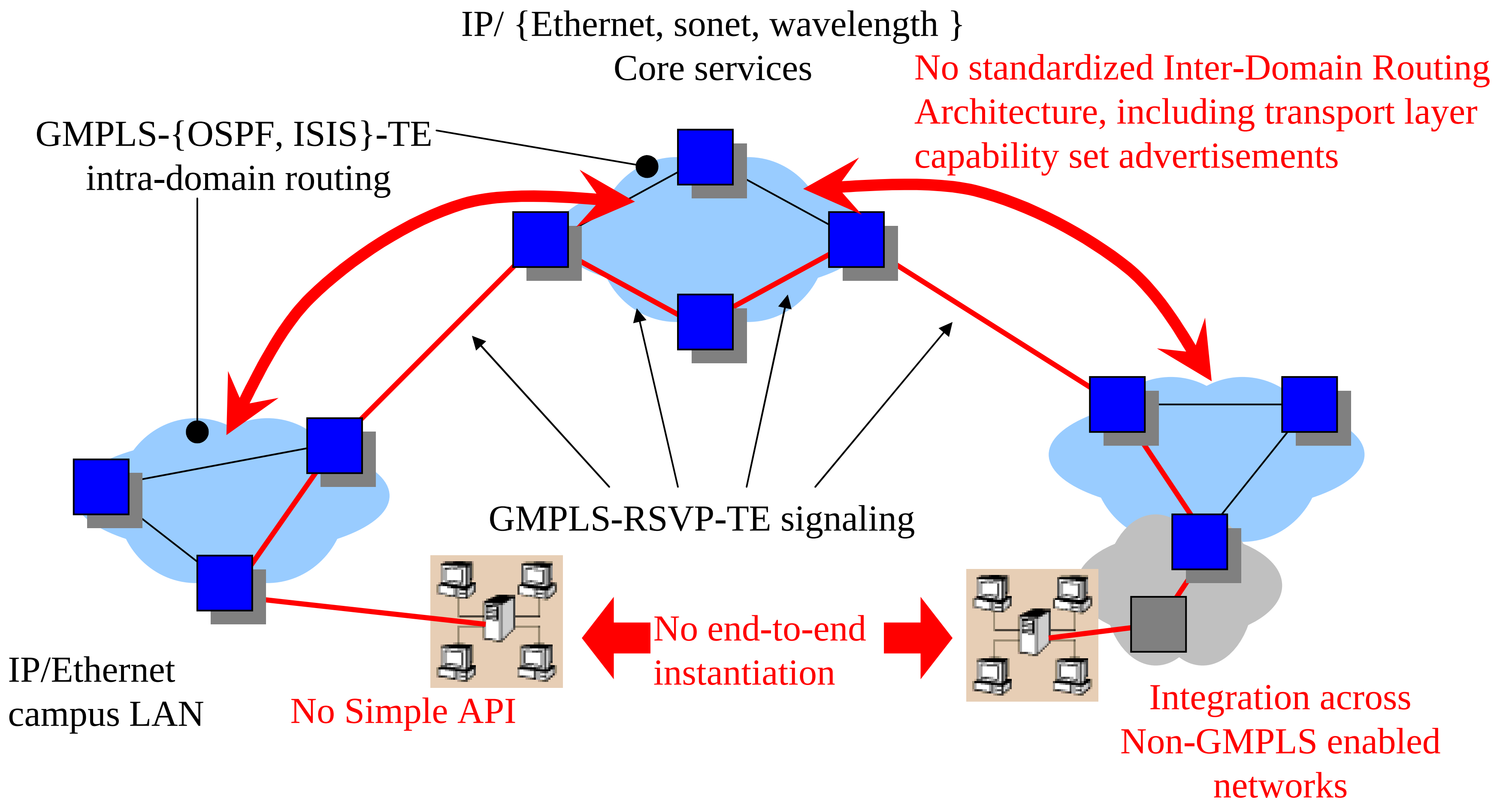
- MEMS-based switching fabric
- 400 x 400 wavelength switching, scalable to 1000s x 1000s
- 9.23"x7.47"x3.28" in size
- Integrated multiplexing and demultiplexing, eliminating the cost and challenge of complex fiber management

Movaz iWSS prototype switch installed at the University of Maryland

Movaz Partnership: New Technology

- **Movaz and DRAGON will be deploying early versions of new technology such as:**
  – Reconfigurable OADMs
  – Alien wavelength conditioning
  – Tunable wavelength transponders
  – 40 gigabit wavelengths
  – Possibly other digital encoding formats such as RZ, DPSK, etc.
- **The development and deployment plans of selected technologies are part of the annual review cycle**

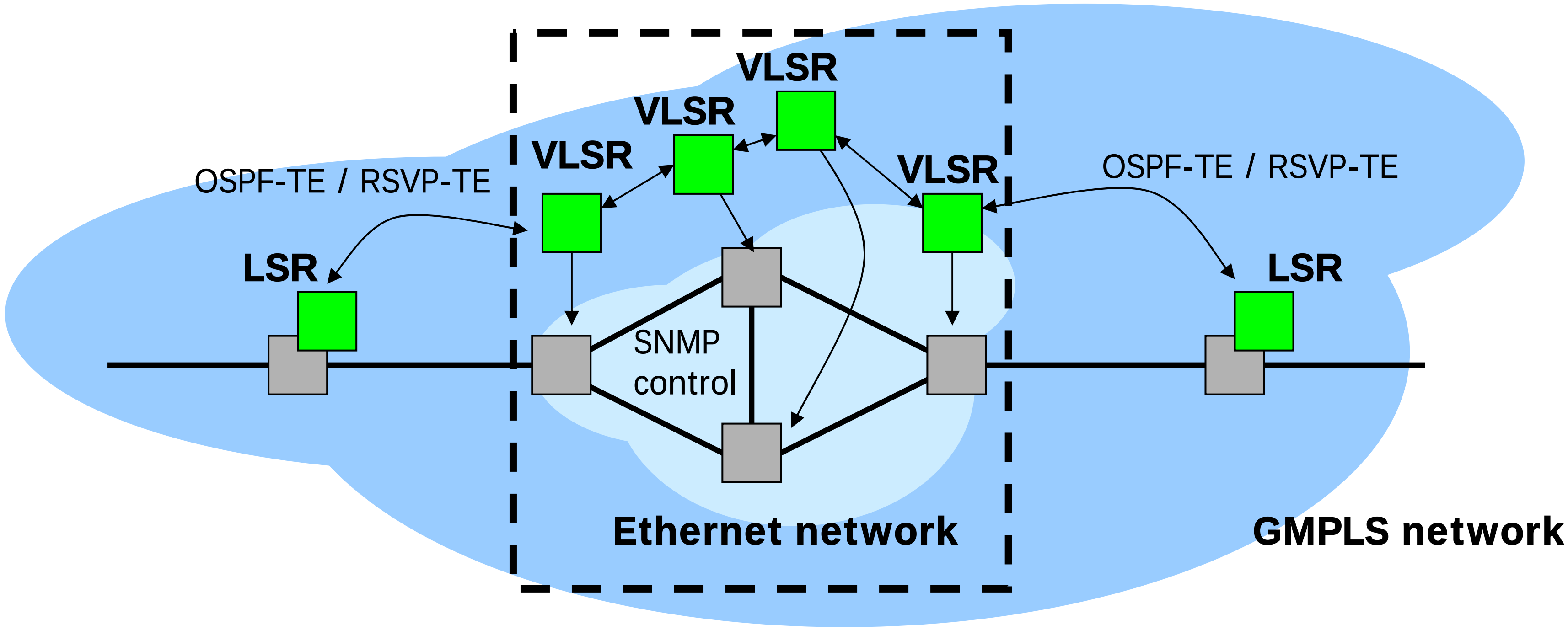
End to End GMPLS Transport
What is missing?



Virtual Label Switched Router: VLSR

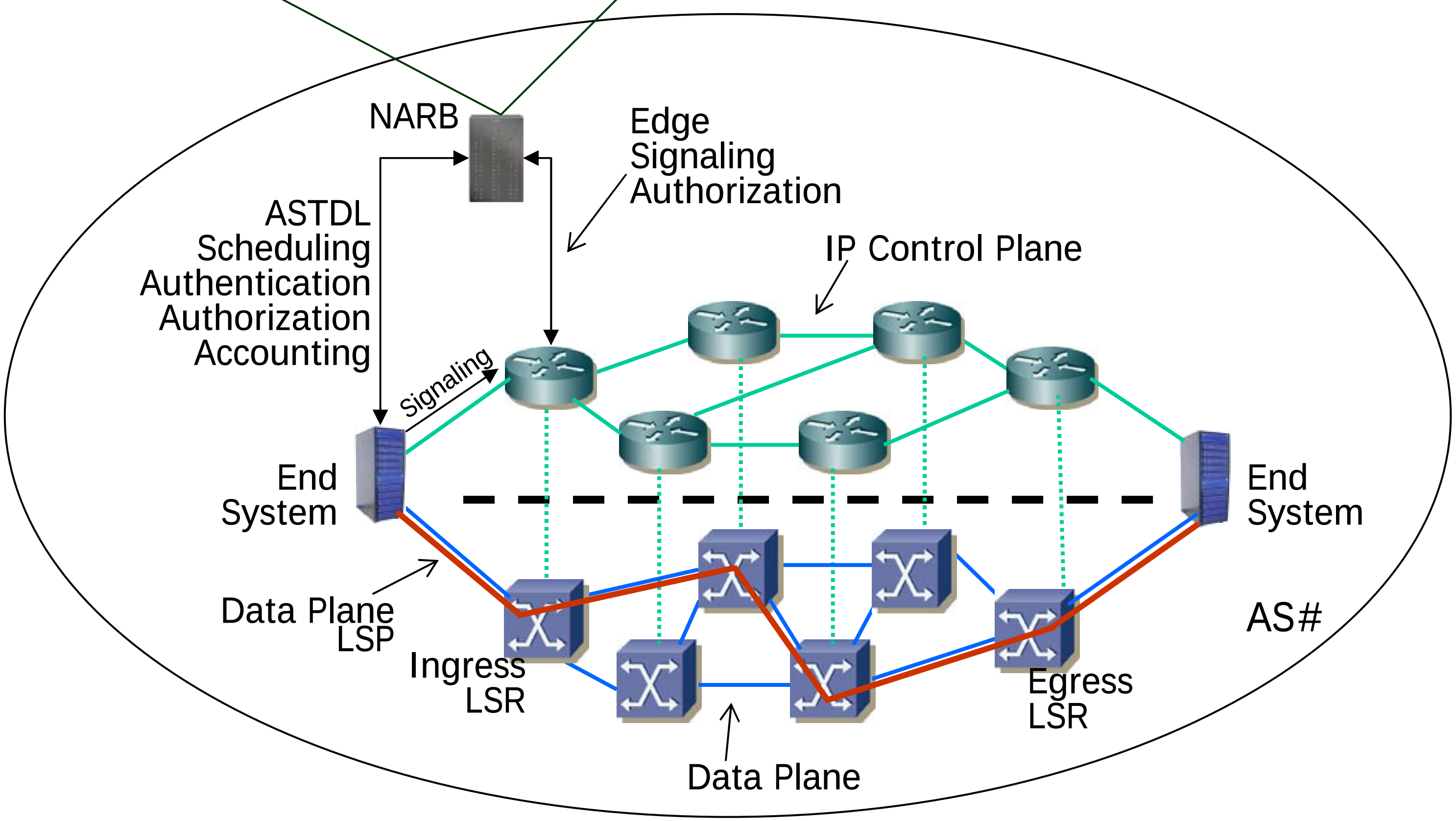
- **Many networks consist of switching components that do not speak GMPLS, e.g. current ethernet switches, fiber switches, etc**
- **Contiguous sets of such components can be abstracted into a Virtual Label Switched Router**
- **The VLSR implements Open Source versions of GMPLS-OSPF-TE and GMPLS-RSVP-TE and runs on a Unix based PC/workstation**
  - Zebra OSPF extended to GMPLS
  - KOM-RSVP likewise
- **The VLSR translates GMPLS protocol events into generic pseudo-commands for the covered switches.**
  - The pseudo commands are tailored to each specific vendor/architecture using SNMP, TL1, CLI, or a similar protocol.
- **The VLSR can abstract and present a non-trivial internal topology as a “black box” to an external peering entity.**

VLSR Abstraction



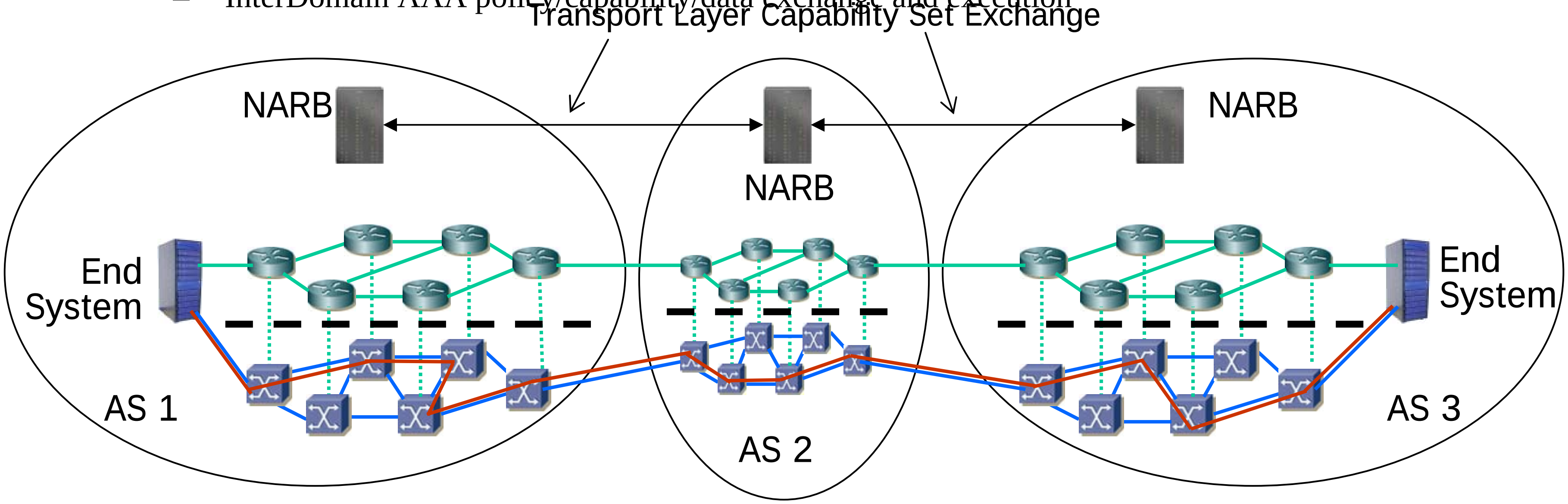
Network Aware Resource Broker (NARB) Functions – IntraDomain

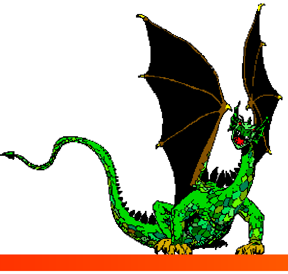
- IGP Listener
- Path Computation
- Scheduling
- Edge Signaling Authentication
- Edge Signaling Enforcement
- ASTDL Induced Topology Computations
- Authorization (flexible policy based)
- Authentication
- Accounting



Network Aware Resource Broker (NARB) Functions - InterDomain

- **InterDomain NARB must do all IntraDomain functions plus:**
  - EGP Listener
  - Exchange of InterDomain transport layer capability sets
  - InterDomain path calculation
  - InterDomain AAA policy/capability/data exchange and execution

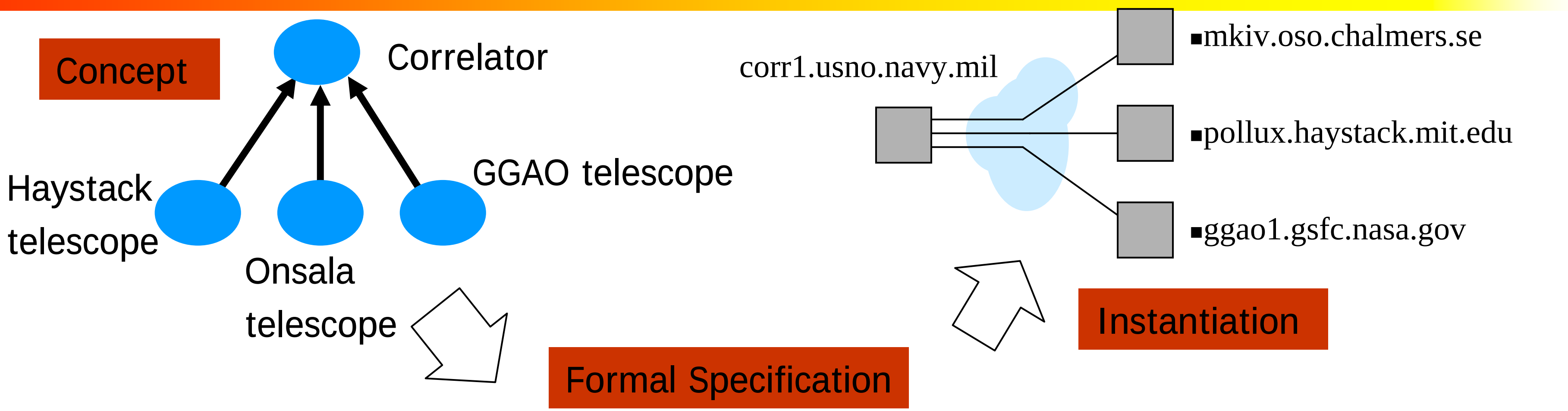




Application Specific Topology Description Language: ASTDL

- **ASTDL is a formalized definition language that describes complex topologies**
  - By formally defining the application’s network requirements, service validation and performance verification can be performed (“wizard gap” issues)
  - Formal definition allows advanced scheduling – which must still be integrated with non-network resources such as computational clusters, instruments, sensor nets, storage arrays, visualization theatres…
- **ASTDL includes run time libraries to instantiate the topology and link in the other resources of the application**
  - Application topologies consist of multiple LSPs that must be instantiated as a set.
  - Resource availability must be dependable and predictable, i.e. resources must be reservable in advance for utilization at some later time

Application Specific Topology Description Language - ASTDL

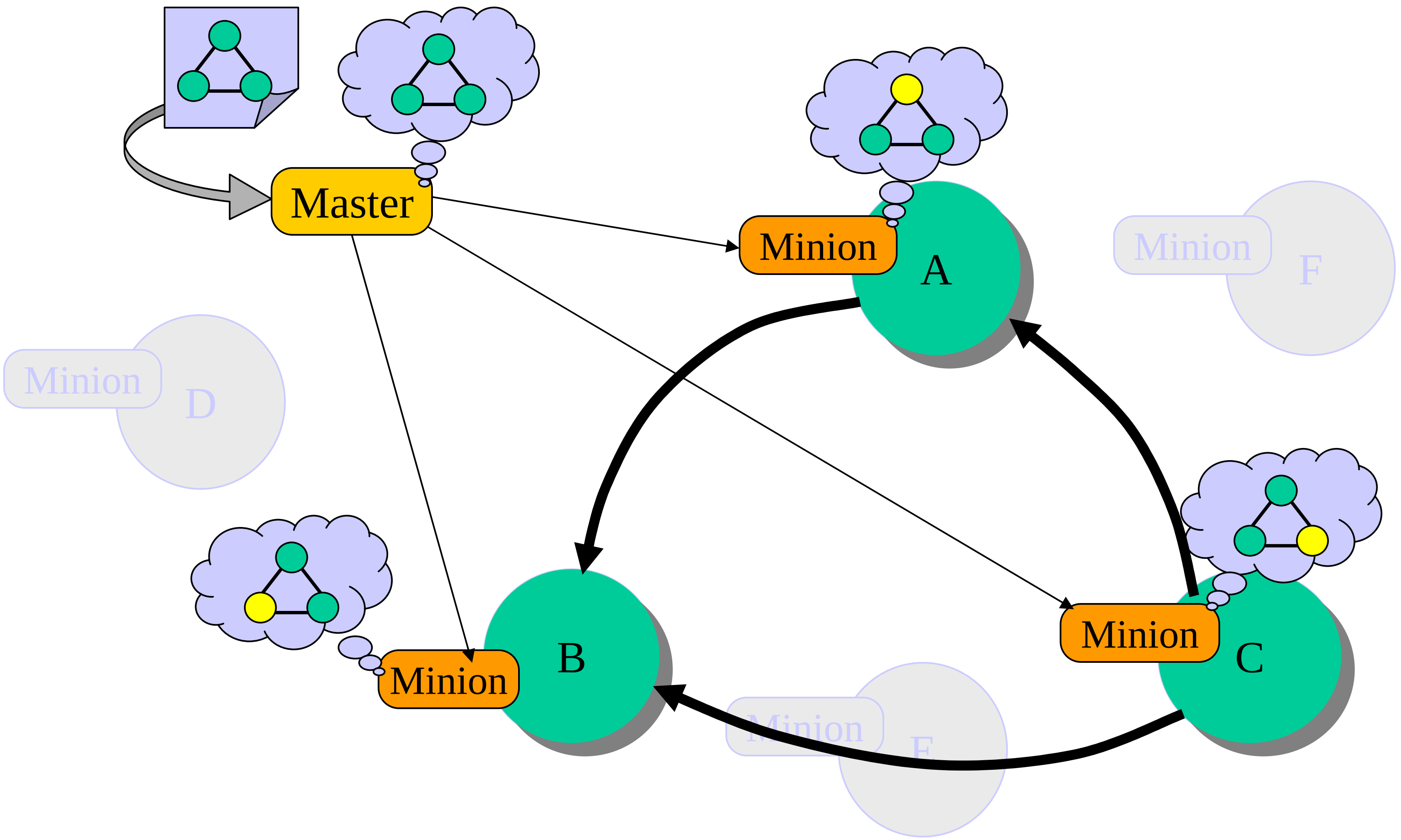


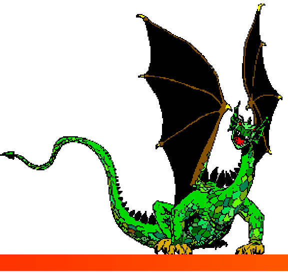

```
Datalink:= { Framing=Ethernet; bandwidth=1g;
                 SourceAddress=%1::vlbid;   DestinationAddress=%2; }
Topo_vlbi_200406 := {
        Correlator:=corr1.usno.navy.mil::vlbid;         // USNO
        DataLink( mkiv.oso.chalmers.se, Correlator );   // OSO Sweden
        DataLink( pollux.haystack.mit.edu, Correlator );// MIT Haystack
        DataLink( ggao1.gsfc.nasa.gov, Correlator );    // NASA Goddard
        }
```

C++ Code invocation example:

```
eVLBI = new ASTDL::Topo( "Topo_vlbi_200406");  // Get the topology definition
Stat = eVLBI.Create();                          // Make it so!
```

The AST Process





ASTDL Driver Example

```
#include "class_Topo.h++"
#define DRAGON_TCP_PORT 5555;
//
// User prime mover "ast_master" for AST miniond
//
using namespace std;
using namespace ASTDL;

int main(int argc, char *argv[])
{
  int stat;
  Topo *topo;
  topo = new Topo(argv[1]);
  if(topo == NULL) exit(1);
  stat = topo->Resolve();
  if(stat != 0) exit (2);
  stat = topo->Instantiate();
  if(stat != 0)
    { cout << "Error stat=" << stat << endl; exit(3);}
  stat = topo->UserHandoff();
  stat = topo->Release();
  exit(0)
}
```

Read topology definition source and create the Topo object

Resolve hostnames and other service specific data

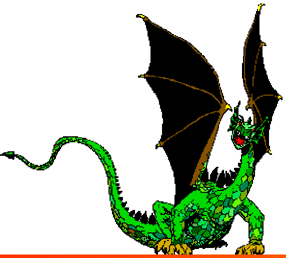
Establish all the connections

exec() the user and pass off the connections.

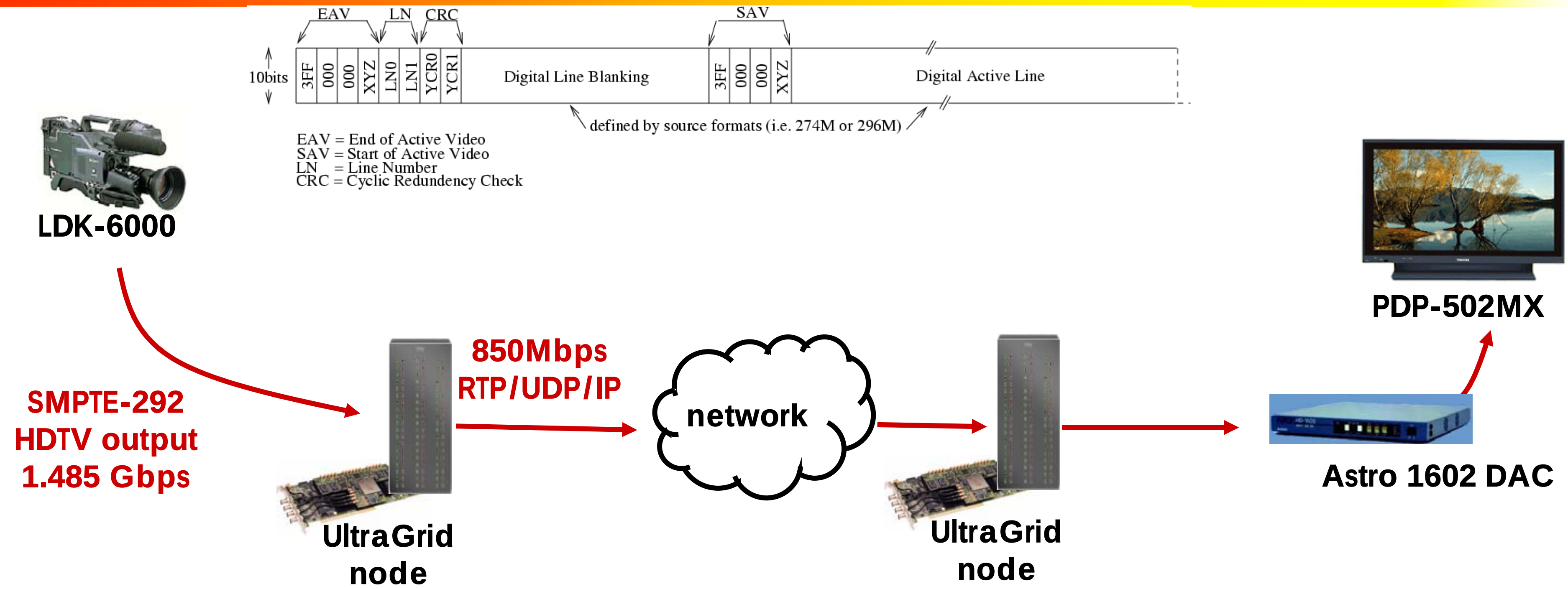
All done. Tear it down.

High Definition Collaboration and Visual Area Networking (HD-CVAN)

- **Dragon dynamic resource reservation will be used to instantiate an application specific topology**
  - Video directly from HDTV cameras and 3D visualization clusters will be natively distributed across network
- **Integration of 3D visualization remote viewing and steering into HD collaboration environments**
- **HD-CVAN Collaborators**
  - UMD VPL
  - NASA GSFC (VAL and SVS)
  - USC/ISI (UltraGrid Multimedia Laboratory)
  - NCSA ACCESS

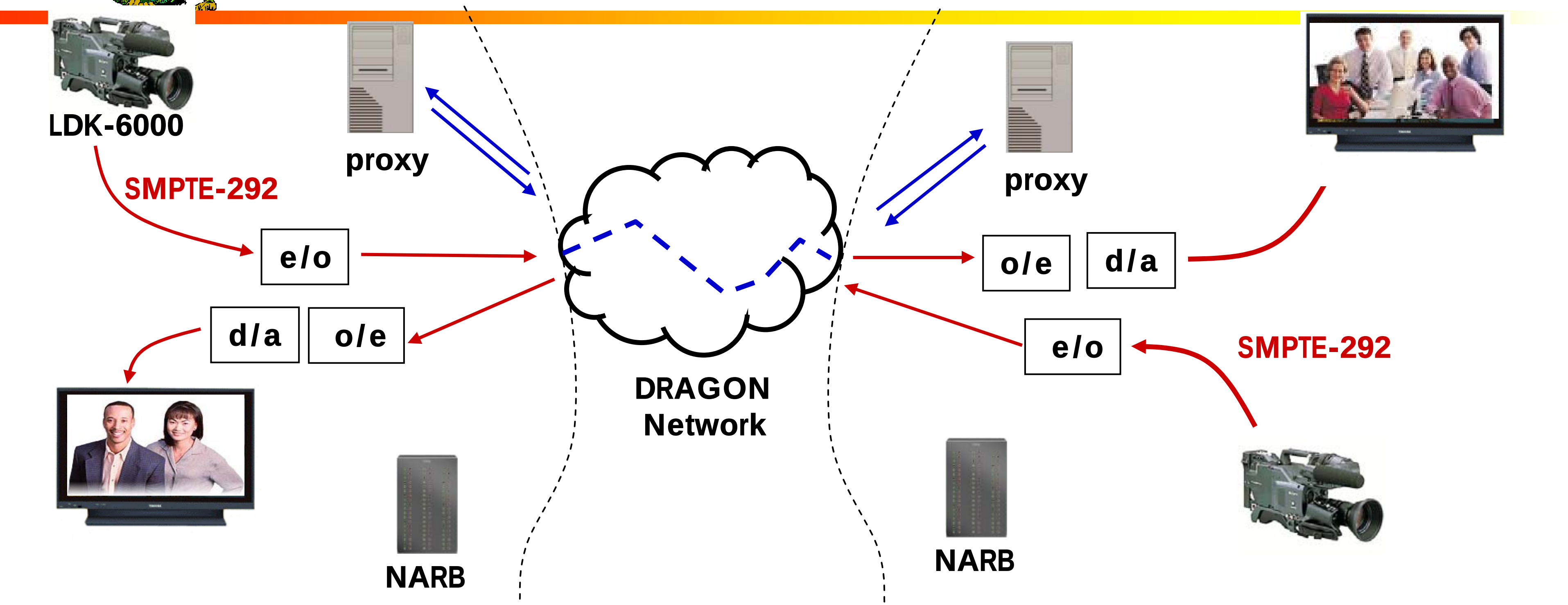


Uncompressed HDTV-over-IP Current Method



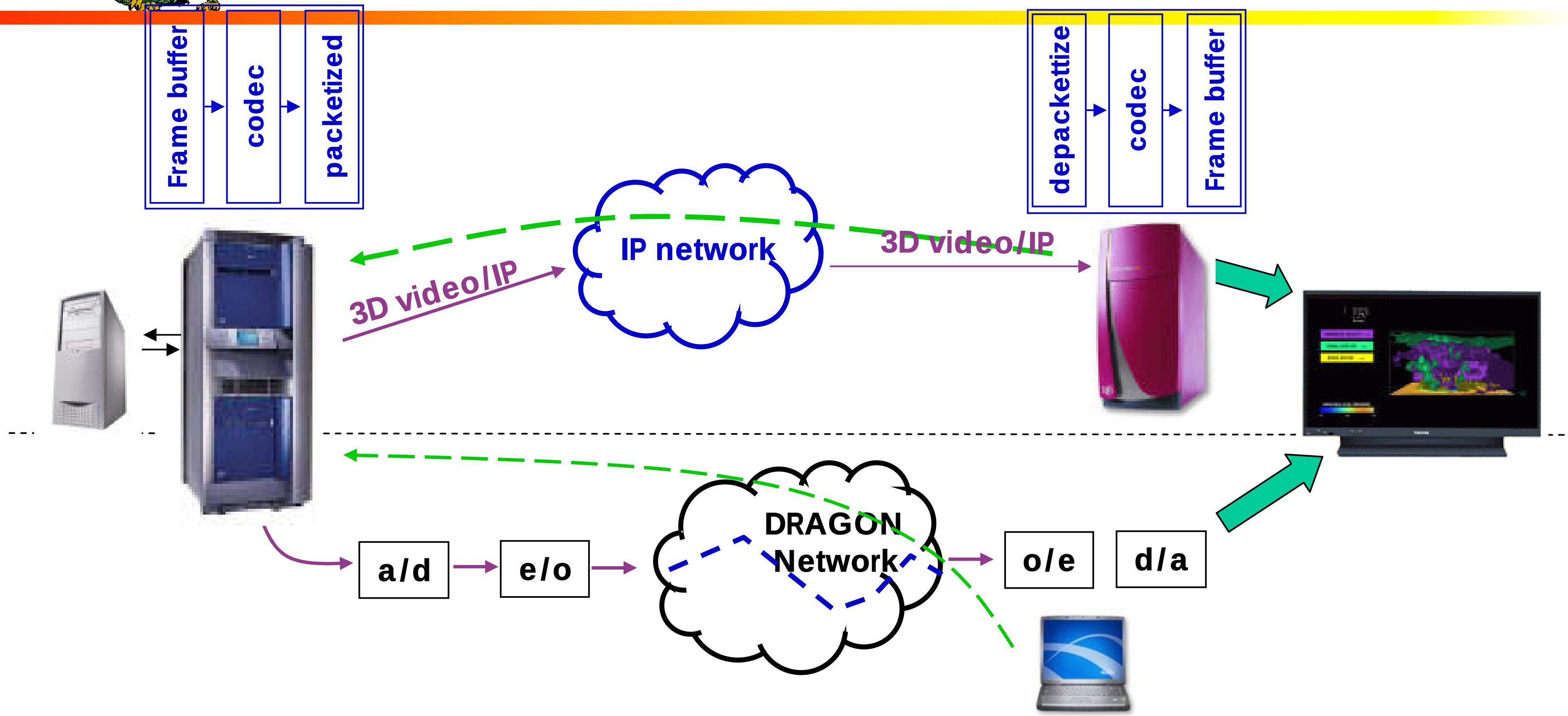
- Not truly HDTV --> color is subsampled to 8bits
- Performance is at the mercy of best-effort IP network
- UltraGrid processing introduces some latency

Low latency High Definition Collaboration DRAGON Enabled



- **End-to-end native SMPTE 292M transport**
- **Media devices are directly integrated into the DRAGON environment via proxy hosts**
  - Register the media device (camera, display, …)
  - Sink and source signaling protocols
  - Provide Authentication, authorization and accounting.

Low Latency Visual Area Networking



- Directly share output of visualization systems across high performance networks.
- DRAGON allows elimination of latencies associated with IP transport.

Status to Date

- **Wavelength Selective Switch staged at UMD**
  - Installed and operational Spring 2004, second phase expansion to happen Summer ‘05
  - GMPLS control plane operational Spring ‘04
- **Initial VLSR functionality demonstrated**
  - Successful interop tests with across Movaz, Juniper, Sycamore, Ethernet switches. More to come…
  - VLSR being used in other testbeds: CHEETAH, HOPI and USN are interested…
- **Initial NARB demonstrated at SuperComputing 2004 in Pittsburgh**
  - Plans being made to extend DRAGON to NLR and begin inter-domain experiments with Omninet, HOPI, USN, and others.
  - Currently planning optical GMPLS peerings with ATDnet(WDM) and HOPI (Ethernet)
- **HD-CVAN UltraGrid node, using DRAGON technologies, demo’d at SC04**
  - Will be tested in DRAGON network over Spring 05

For more information…

- **Web:** **dragon.maxgigapop.net**
- **Contact:** Jerry Sobieski
jerrys@maxgigapop.net
+1-301-314-6662

Thank You!