5 August 2010

Eric Boyd, Internet2 Deputy CTO

Extending IDC-based Dynamic Circuit Networking Services

Internet2 and Dynamic Circuit Networking

- Internet2 has been working with partners on dynamic circuit networking for many years (BRO, HOPI, DCN)
- In 2009, Internet2 launched the ION service
  - ION is a production service managed by the Internet2 NOC on the Internet2 network backbone
  - ION circuits provisioned using a simple and secure web-based interface or IDC signaling
  - ION circuit setup is done via OSCARS, developed with ESnet, and (until recently) DRAGON, developed with MAX and USC ISI East
  - ION Monitoring is built atop perfSONAR-PS, developed with ESnet, FNAL, SLAC
  - ION speaks the IDC protocol, developed with ESnet and GÉANT
- Internet2 ION service transitioned from dedicated network built atop Ciena CoreDirectors to Juniper MX960s (complete as of July 6th, 2010)
  - MPLS transport more efficient use of resources
  - Bandwidth reserved for circuit instantiation is available for use by other users when circuit owner not utilizing circuit for transfer
  - Opportunity to provide circuits that can burst above their requested commit rate, if sufficient headroom available

Dynamic Circuit Networking - Research

- Internet2 is engaged in standardization effort within the Open Grid Forum (OGF)
  - NSI WG (signal circuit creation)
  - NML WG (topology support of circuits and monitoring)
  - NMC WG (performance middleware)
  - NM WG (measurement definition)
- Internet2 is engaged in GLIF experimental networks
  - perfSONAR Demos
  - Automated GOLE Demos (working with FENIUS effort)
- Internet2, in consultation with partners, plans to evolve working IDC-based service instances over time in keeping with emerging consensus on standards and best practices within the OGF and GLIF

Dynamic Circuit Networking – Service Extension

- Internet2 is engaged in a common service definition of a dynamic circuit networking service with DICE partners
  - Internet2 ION Service (built with OSCARS over Juniper MX-960s and perfSONAR-ps)
  - ESnet Science Data Network (SDN) (built with OSCARS over Juniper MX-960s and perfSONAR-PS)
  - USLHCNet (built with OSCARS/DRAGON over Ciena Core Directors, perfSONAR-ps and MonaLISA)
  - GÉANT (built with AutoBAHN and perfSONAR-MDM, March 2011)
- Internet2 ION Network connects with other IDC instances (e.g. JGN2, LEARN, Caltech)
- Internet2 plans to deploy and operate a separate IDC instance at MAN LAN in 2011
  - Built with OSCARS/DRAGON over Ciena Core Director, perfSONAR-ps)

Dynamic Circuit Networking – Service Extension

- Internet2 would like to continue to extend dynamic circuit networking services (specifically service definitions based on the IDC protocol)
  - US regional networks
  - US campuses
  - US-funded international links
- Three NSF awards of interest (total $3.5 million)
  - IRIS: develop perfSONAR-ps based performance monitoring package for use by IRNC PRONET awardees (~$1 million)
  - DyGIR: develop OSCARS (and DRAGON) based dynamic circuit networking package for use by IRNC PRONET awardees (~$750k)
  - DYNES: develop OSCARS (DRAGON) and perfSONAR-ps based distributed virtual instrument for LHC community and others (~$1.75 million)

IRIS and DyGIR

- PI: Eric Boyd; Co-PI: Ana Preston
  - eboyd@internet2.edu
  - apreston@internet2.edu
- Division of Responsibilities:
  - Eric: Execution of IRIS and DyGIR proposals, direct engagement with other PIs on deliverables
  - Ana: Overall coordination, Internet2 engagement with IRNC projects and partners
- IRIS: International Research Instrumentation System
  - Support performance monitoring via perfSONAR-PS
  - The “performance / perfSONAR” one
- DyGIR: Dynamic Gateways for International Research
  - Support automatic provisioning of circuits via OSCARS / DRAGON
  - The “dynamic circuits / DCN / IDC / ION/ OSCARS” one

IRIS

- Objective: provide the infrastructure necessary for the identification, diagnosis and eventual correction of network performance problems for paths traversing IRNC links
- IRIS will provide a software framework to enable performance monitoring services on IRNC links
  - Produce a set of easy-to-install, tailored software packages of the perfSONAR-PS software suite for use by the IRNC link operators
  - Develop new functionality specific to international exchange points
  - Work with the IRNC link awardees to help them deploy the software

DyGIR

- Objective: enable researchers to reserve dedicated bandwidth over IRNC links
  - Allow for more distributed collaboration
- DyGIR will provide a prototype software framework to enable dynamic circuit services for a pair of IRNC links
  - Produce a set of easy-to-install, tailored software packages for the OSCARS software suite
  - Prototype new functionality specific to international exchange points
  - Work with 2+ of the IRNC link awardees to help them deploy the prototype software

IRIS and DyGIR: General Approach

- Goals:
  - Create 2 software packages
  - Enable successful IRNC PRONET awardees to effectively meet the monitoring and dynamic circuits support requirements of the NSF solicitation
  - Partner with interested IRNC PRONET awardees in order to support their success
- "All network measurement activities must address interoperability with emerging best practice international research and education network performance monitoring, or explain in detail why this is not feasible. This should include the deployment of compatible bandwidth and latency measurement services, as well as compatible measurement archives, and details about access to these archives by relevant parties. Additional services such as monitoring for flow analysis are also encouraged if feasible. These services may be part of the security plan for the project and should be addressed as such."
- "Hybrid network services have matured to the point that they are considered for the time period supported in this solicitation to be production services; thus, dynamic circuit networking capabilities in combination with shared IP services must be supported. Specifically, proposals should address interoperability with emerging production dynamic switching network services, or explain in detail why such interoperability is not feasible."

Software Infrastructure

- Build on software developed in part by Internet2
  - perfSONAR-PS: distributed monitoring framework
  - OWAMP: one-way latency measurement tool
  - BWCTL: throughput measurement tool
  - OSCARS: implements InterDomain Controller protocol
- What's missing
  - IRNC PRONET Awardee requirements
    - Needs ongoing discussion
    - Example: Better support in perfSONAR-PS for passive measurements
    - Example: Better support in OSCARS for exchange points
  - Improved packaging support for quick and easy installation and operation

IRIS Milestones

- Requirements gathering from IRNC PRONET Awardees
- Publish IRNC Measurement Best Common Practice
- Development of features required for link operators and exchange points.
- Creation of easy-to-install software packages to enable IRNC awardees to make monitoring services available.
- Software package deployment assistance for IRNC awardees.
- Regular releases of updated IRIS packages.
- Community Outreach

DyGIR Milestones

- Requirements gathering from IRNC PRONET Awardees
- Release of a hardware recommendation that is known to work across a wide variety of networks.
- Development of features required for link operators and exchange points.
- Creation of easy-to-install software packages to enable IRNC awardees to make dynamic circuit services available.
- Software package deployment assistance for IRNC awardees.
- Regular releases of updated DyGIR packages.
- Community Outreach

DYNES Background

- Internet2 approached by 3 universities (Caltech, University of Michigan, Vanderbilt) and asked to lead a proposal on behalf of LHC and R&E community (July 2009)
  - MRI-R2 solicitation capped the per university submissions at 3
- Issue was referred to advisory councils; recommendation to proceed from AOAC and RAC (July 2009)
- Internet2 convened a series of community calls (August 2009)
- With strong community support, submitted proposal to Major Research Instrument – Recovery and Reinvestment (MRI-R2) (August 2009)
- Funded (August 2010) @ $1.74 million
- http://www.internet2.edu/dynes
- PI: Eric Boyd; Co-PIs: Harvey Newman, Shawn McKee, Paul Sheldon
  - eboyd@internet2.edu
  - Harvey.Newman@cern.ch
  - smckee@umich.edu
  - Paul.Sheldon@vanderbilt.edu

DYNES Summary

- What is it?:
  - A nationwide cyber-instrument spanning ~40 US universities and ~14 Internet2 connectors
    - Extends Internet2's ION service into regional networks and campuses, based on OSCARS implementation of IDC protocol (developed in partnership with ESnet)
- Who is it?
  - A collaborative team including **Internet2**, **Caltech**, **University of Michigan**, and **Vanderbilt University**
  - Community of regional networks and campuses
  - LHC, astrophysics community, OSG, WLCG, other virtual organizations
- What are the goals?
  - Support large, long-distance scientific data flows in the LHC, other leading programs in data intensive science (such as LIGO, Virtual Observatory, and other large scale sky surveys), and the broader scientific community
  - Build a distributed virtual instrument at sites of interest to the LHC but available to R&E community generally

DYNES - Dynamic Network Circuits

- DYNES will deliver the needed capabilities to the LHC, and to the broader scientific community at all the campuses served, by coupling to their analysis systems:
  - Dynamic network circuit provisioning: IDC Controller
  - Data transport: Low Cost IDC-capable Ethernet Switch; FDT Server for high throughput, Low cost storage array where needed (also non-LHC)
  - End-to-end monitoring services
- DYNES does not fund more bandwidth, but provides access to Internet2's dynamic circuit network ("ION"), plus the standard mechanisms, tools and equipment needed
  - To build circuits with bandwidth guarantees across multiple network domains, across the U.S. and to Europe
    - In a manageable way, with fair-sharing
    - Will require scheduling services at some stage
  - To build a community with high throughput capability using standardized, common methods

INTERNET 2

DYNES Community Support

- Internet2 received a total of 60 Letters of Collaboration
  - 44 Universities (some duplicates)
  - 14 Regional Networks
  - 1 Virtual Organization
  - 1 Federal Lab
  - Letter format specified by solicitation
    - By signing below I acknowledge that I am listed as a collaborator and/or instrument user on this MRI proposal, entitled “DYNES”, with Eric Boyd as the Principal Investigator. I agree to perform the tasks assigned to me, as described in the proposal, and I commit to provide or make available the resources therein designated to me.
  - Proposal narrative text explained scope of commitment

DYNES System Description

AIM: extend hybrid & dynamic capabilities to campus & regional networks.

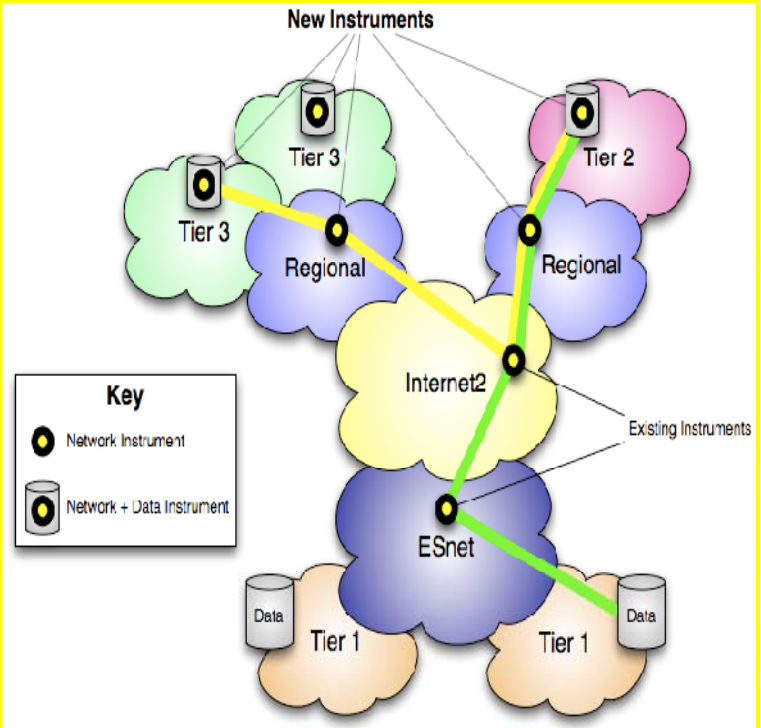
- A DYNES instrument must provide two basic capabilities at the Tier 2S, Tier3s and regional networks:

1. Network resource allocation such as bandwidth to ensure performance of the transfer
2. Monitoring of the network and data transfer performance

- All networks in the path require the ability to allocate network resources and monitor the transfer. This capability currently exists on backbone networks such as Internet2 and ESnet, but is not widespread at the campus and regional level.

- <u>In addition Tier 2 & 3 sites require:</u>

3. Hardware at the end sites capable of making optimal use of the available network resources

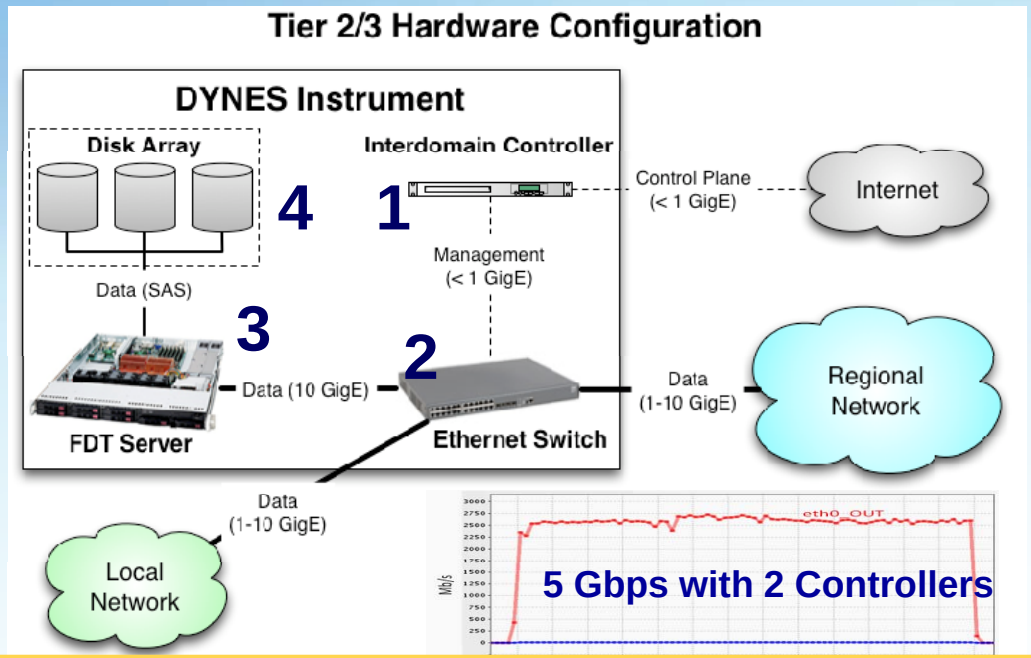


Two typical transfers that DYNES supports: one between a Tier 2 and a Tier 3 and another between a Tier 2 and a Tier 1 site. The clouds represent the network domains involved in such a transfer.

DYNES: Tier2 and Tier3 - Instrument Design

- Each DYNES (sub-)instrument at a Tier2 or Tier3 site consists of the following hardware, where each item has been carefully chosen to combine
  low cost & high performance:

1. An Inter-domain Controller (IDC)
2. An Ethernet switch
3. A Fast Data Transfer (FDT) server. Sites with 10GE throughput capability will have a dual-port Myricom 10GE network interface
   in the server.
4. An attached disk array with a Serial Attached SCSI (SAS) controller capable of several hundred MBytes/sec to local storage.



The Fast Data Transfer (FDT) server connects to the disk array via the SAS controller and runs FDT software developed by Caltech. FDT is an asynchronous multithreaded system that automatically adjusts I/O and network buffers to achieve maximum network utilization. The disk array stores datasets to be transferred among the sites in some cases. The FDT server serves as an aggregator/ throughput optimizer in this case, feeding smooth flows over the networks directly to the Tier2 or Tier3 clusters. The IDC server handles the allocation of network resources on the switch, inter-actions with other DYNES instruments related to network pro-visioning, and network performance monitoring. The IDC creates virtual LANs (VLANs) as needed.

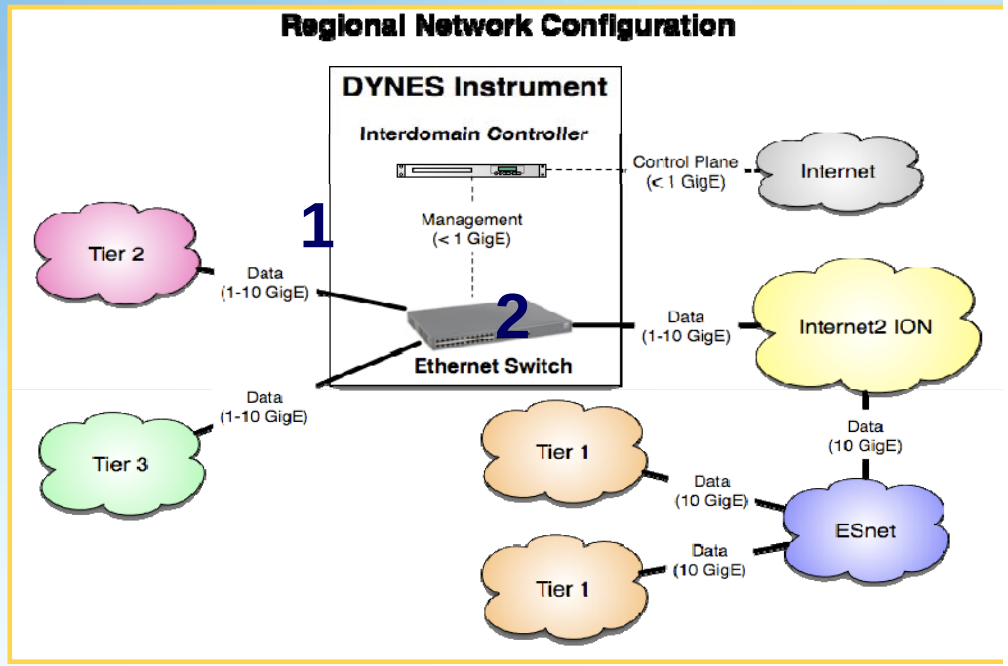
DYNES: Regional Network - Instrument Design

- Regional networks require

1. An Ethernet switch
2. An Inter-domain Controller (IDC), as shown.

The configuration of the IDC consists of OSCARS, DRAGON, and perfSONAR just as in the Tier2/ Tier 3 cases. This allows the regional network to provision resources on-demand through interaction with the other instruments

A regional network does not require a disk array or FDT server because they are providing transport for the Tier 2 and Tier 3 data transfers, not initiating them.



At the network level, each regional connects the incoming campus connection to the Ethernet switch provided. Optionally, if a regional network already has a qualified switch compatible with the dynamic software that they prefer, they may use that instead, or in addition to the provided equipment. The Ethernet switch provides a VLAN dynamically allocated by OSCARS & DRAGON. The VLAN has quality of service (QoS) parameters set to guarantee the bandwidth requirements of the connection as defined in the VLAN. These parameters are determined by the original circuit request from the researcher / application. through this VLAN, the regional provides transit between the campus IDCs connected in the same region or to the global IDC

DYNES Site Selection Process

- All sites will need to apply
- External review panel
  - Composition and process still TBD
  - 6+ members
    - CMS overall representative
    - CMS Tier 3 representative
    - ATLAS overall representative
    - ATLAS Tier 3 representative
    - 2 Non-LHC high performance networking community reps
    - Plus ex-officio members: PI and co-Pis
- Community feedback encouraged on site selection process

DYNES Roadmap

- ICO and Council notification
- Community announcement
  - Letter, website
  - 2 Community Calls
- Set up external review panel (October, 2010)
- CFP (November 1, 2010)
- Calls with individual regionals and with individual campuses (plus associated regional) upon request
- DYNES BoF (November 1, 2010, FMM)
- Site selections announced (January, 2011)
- Distributed virtual Instrument building begins in 2011

Aggregate Goal of IRIS, DyGIR and DYNES

By 2013 …

- Integrated end-to-end performance monitoring is a “normal part of networking”
  - Preserve local autonomy and enable diversity of networking practices
  - Beginning to see monitoring directly supported on networking devices
- Integrated end-to-end automatic provisioning of circuits is a “normal part of networking”
  - Preserve local autonomy and enable a diversity of networking technologies and practices

Questions

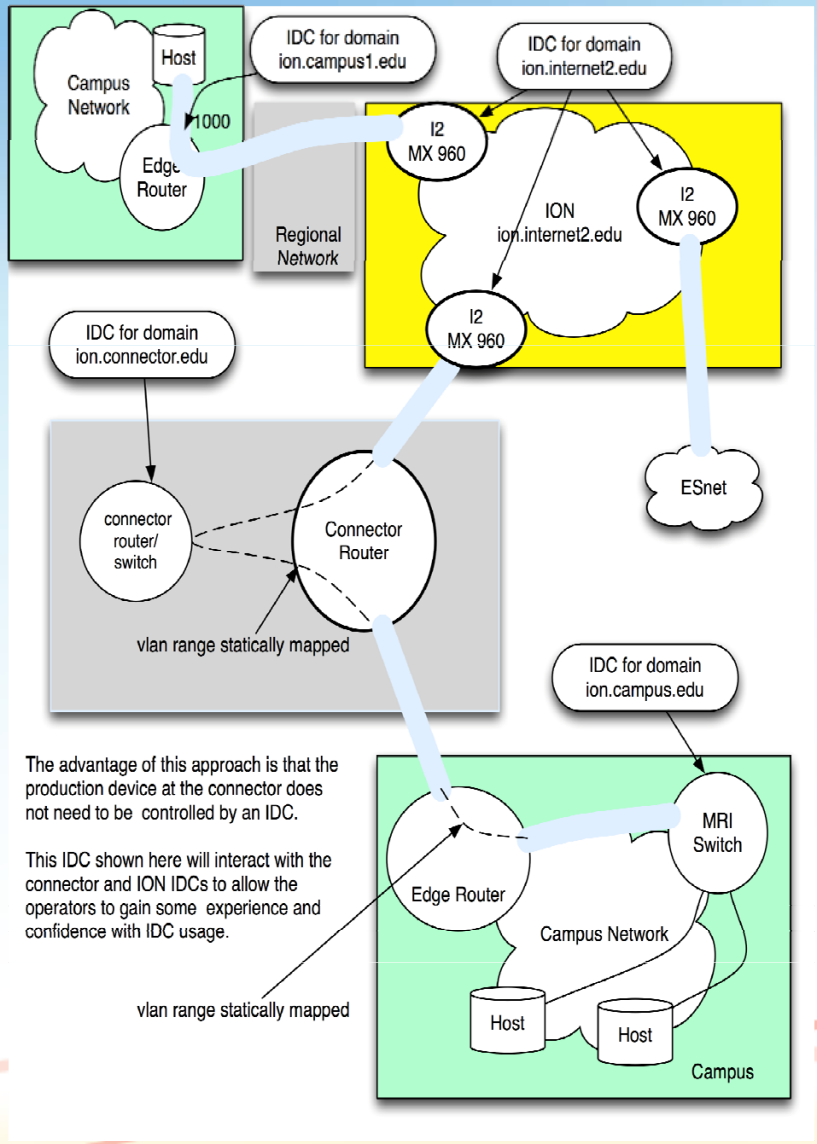
- Eric Boyd
  - eboyd@internet2.edu
- DYNES
  - dynes-questions@internet2.edu

13 October 2010

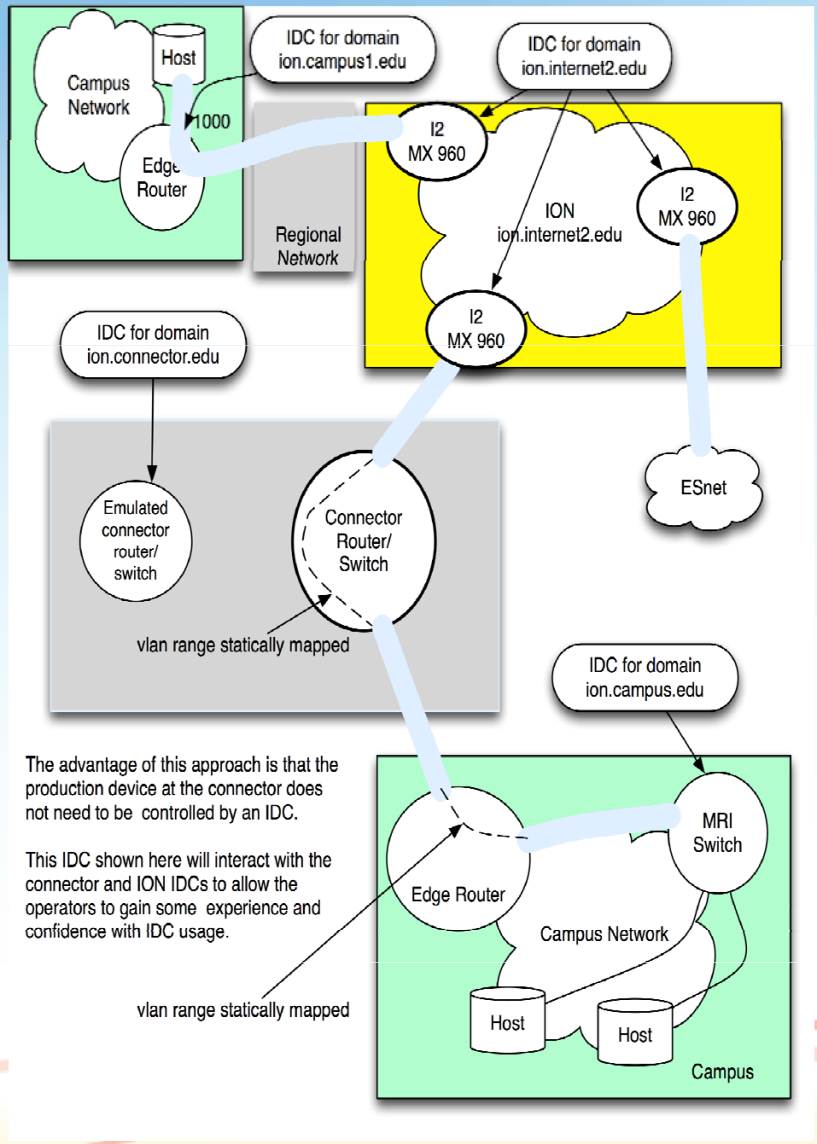
Eric Boyd, Internet2 Deputy CTO

Extending IDC-based Dynamic Circuit Networking Services

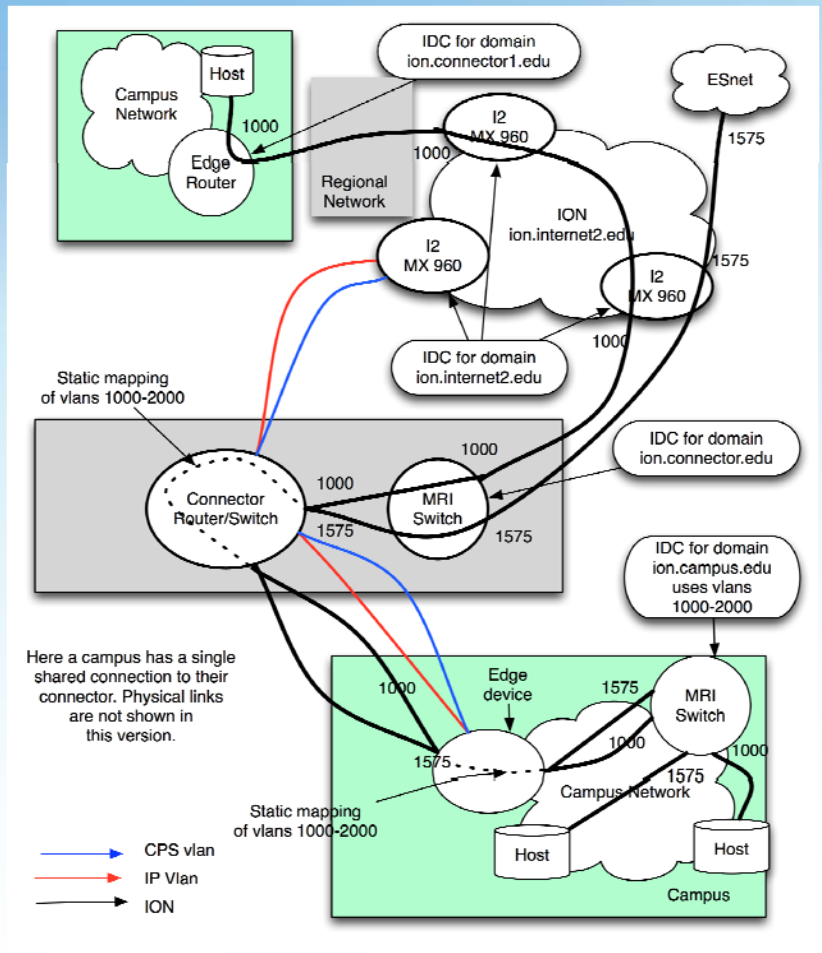
DYNES Minimum Connection Scenario A



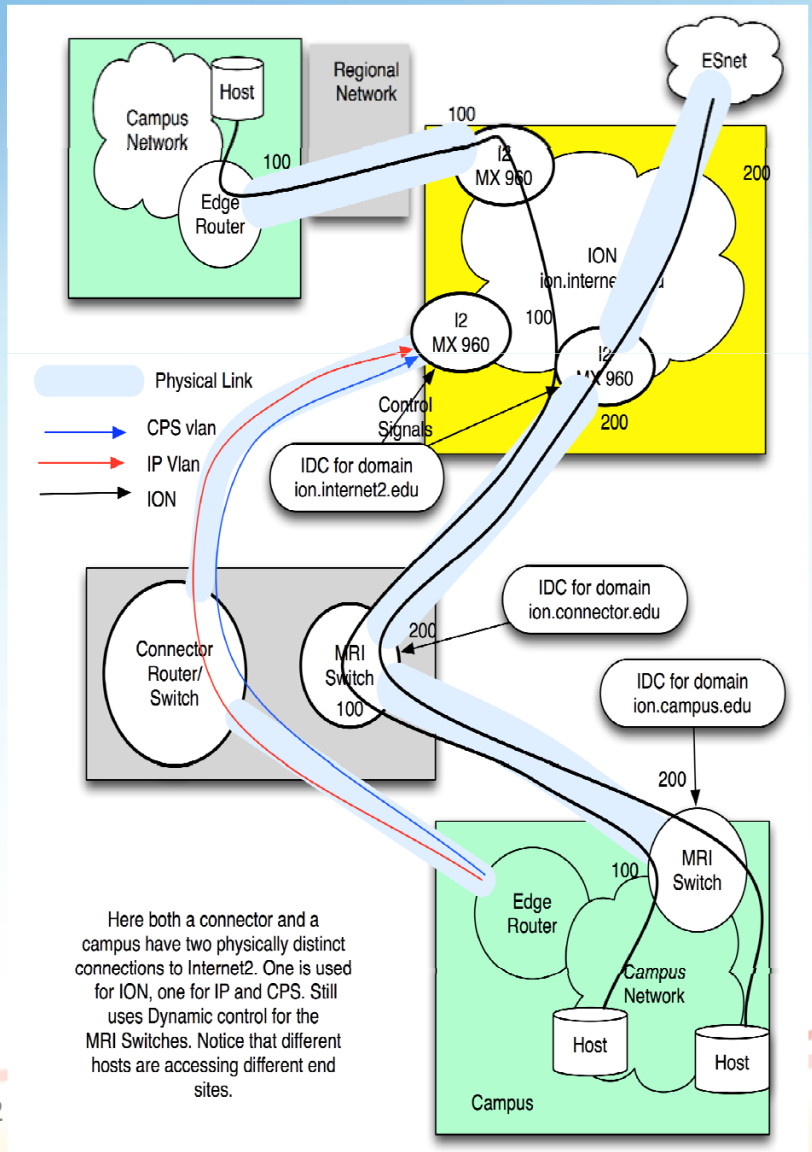
DYNES Minimum Connection Scenario B



DYNES Single Campus Link Scenario



DYNES Dual Campus Link Scenario



DYNES Full Dynamic Control Scenario