The DYNES Instrument: A Description and Overview

Jason Zurawski[1], Robert Ball[2], Artur Barczyk[3], Mathew Binkley[4], Jeff Boote[1], Eric Boyd[1], Aaron Brown[1], Robert Brown[4], Tom Lehman[5], Shawn McKee[2], Benjeman Meekhof[2], Azher Mughal[3], Harvey Newman[3], Sandor Rozsa[3], Paul Sheldon[4], Alan Tackett[4], Ramiro Voicu[3], Stephen Wolff[1], and Xi Yang[5]

[1] Internet2
[2] University of Michigan
[3] California Institute of Technology
[4] Vanderbilt University
[5] ISI East

E-mail: zurawski@internet2.edu, ball@umich.edu, artur.barczyk@cern.ch, mathew.binkley@vanderbilt.edu, boote@internet2.edu, eboyd@internet2.edu, aaron@internet2.edu, bobby.brown@vanderbilt.edu, tlehman@east.isi.edu, smckee@umich.edu, bmeekhof@umich.edu, azher.mughal@cern.ch, Harvey.Newman@cern.ch, sandor.gyula.rozsa@cern.ch, paul.sheldon@vanderbilt.edu, alan.tackett@vanderbilt.edu, ramiro.voicu@cern.ch, swolff@internet2.edu, xyang@east.isi.edu

Abstract. Scientific innovation continues to increase requirements for the computing and networking infrastructures of the world. Collaborative partners, instrumentation, storage, and processing facilities are often geographically and topologically separated, as is the case with LHC virtual organizations. These separations challenge the technology used to interconnect available resources, often delivered by Research and Education (R&E) networking providers, and leads to complications in the overall process of end-to-end data management.

Capacity and traffic management are key concerns of R&E network operators; a delicate balance is required to serve both long-lived, high capacity network flows, as well as more traditional end-user activities. The advent of dynamic circuit services, a technology that enables the creation of variable duration, guaranteed bandwidth networking channels, allows for the efficient use of common network infrastructures. These gains are seen particularly in locations where overall capacity is scarce compared to the (sustained peak) needs of user communities. Related efforts, including those of the LHCOPN [3] operations group and the emerging LHCONE [4] project, may take advantage of available resources by designating specific network activities as a "high priority", allowing reservation of dedicated bandwidth or optimizing for deadline scheduling and predicable delivery patterns.

This paper presents the DYNES instrument, an NSF funded cyberinfrastructure project designed to facilitate end-to-end dynamic circuit services [2]. This combination of hardware and software innovation is being deployed across R&E networks in the United States at selected end-sites located on University Campuses. DYNES is peering with international efforts in other countries using similar solutions, and is increasing the reach of this emerging technology. This global data movement solution could be integrated into computing paradigms such as cloud and grid computing platforms, and through the use of APIs can be integrated into existing data movement software.

1. Introduction

International scientific innovation, including physics, astronomy, and biology, continue to increase requirements for the computing and networking infrastructures of the world. Collaborative partners, instrumentation, storage, and analysis tools are often geographically separated. Activities that were once localized to a given facility are now deeply reliant on access to high capacity networking for information exchange; the technical requirements for different disciplines stand to grow over the coming years [5]:

- **Collaboration Size:** the addition of more researchers, research facilities, along with increases the pool of users and resources available to process scientific data sets
- **Location of Collaborators:** scientific activity has scaled globally, particularly into non-traditional and remote regions of the world, forcing the construction of supporting infrastructure
- **Data Collection Rates:** upgrades to the basic instruments of science, e.g. colliders, detectors, telescopes, and genome sequencers, are producing finer grained observations that directly translate into increases in the amount of data to process and store
- **Experimental Expectations:** The time expectation to analyze raw information in search of meaningful results is decreasing, thus pushing technology to be available and responsive to user demands

These realities challenge the technology used to interconnect available resources, often delivered by Research and Education (R&E) networking providers, and leads to complications in the overall activity of end-to-end data management. Capacity and traffic management are key concerns of this community; a delicate balance is required to serve both long-lived, high capacity network flows, as well as more traditional end-user activities. Keeping these well stated factors in mind, R&E networking needs to find a way to support scientific demands. Many innovations have emerged, spanning hardware technologies, protocols, software, and services; all of which were directly targeted to domain researchers.

The advent of dynamic circuit services, a technology that enables the creation of variable duration, guaranteed bandwidth networking channels, allows for the efficient use of existing common network infrastructures. These gains are seen particularly in locations where overall capacity is limited compared to the (sustained peak) needs of user communities. Related efforts, including those of the LHCOPN [3] operations group and the emerging LHCONE [4] project, may take advantage of available resources by designating specific network activities as a "high priority", allowing reservation of dedicated bandwidth or optimizing for deadline scheduling and predicable delivery patterns. These advancements, when combined into the cohesive network design strategy informally named the "science DMZ", has gained popularity in the Department of Energy research community as well as within general campus IT infrastructure [6].

In addition to overall network design, the adoption of advanced network services has had a pivotal role in this new paradigm. Monitoring software, such as the perfSONAR framework, can now reliably be used to pinpoint performance trouble spots, or predict potential bottlenecks with the aide of historical measurements [7]. Hybrid networks, including ESnet SDN, GÉANT AutoBAHN, and the Internet2 ION service, offer facilities that enable provisioning of an end-to-end circuits (i.e. a protected Layer 2 VLAN), thus allowing applications to avoid overly congested Layer 3 paths and preventing overuse of general purpose infrastructure [8, 9, 10].

To date, many of the aforementioned technologies were targeted toward the largest component of the worldwide R&E networking infrastructure: Backbone and Regional providers. These far reaching networks, designed to offer large amounts of available capacity, may become congested in certain geographical areas that experience high demand [11, 12]. The ability to create a protected path, separate from the general-purpose IP traffic and using existing network resources (i.e. not requiring new investment), remains highly desirable for bulk data movement applications; particularly those relying on transmission control protocol (TCP) and operating on high round trip time (RTT)/high bandwidth paths, including FDT [13]. This new “circuit” technology, while designed with the WAN in mind, has been slow to migrate to smaller scales where users could take full advantage of this new functionality.

Simple steps to alter the network architecture or enable innovative networking technologies, could lead to significant gains in productivity for scientific users and traffic management ability for network operations staff at institutions of all sizes. The DYnamic NEtwork System (DYNES) is an effort designed to assist in these two key areas by providing efficient hardware, and accompanying advanced network services, to address the needs of the R&E community [2]. Funded by the NSF through the MRI program, DYNES is developing a nationwide “cyber-instrument”; enabling dynamic circuit capability, fast data transfer, and network monitoring at numerous U.S. based universities and regional networks.

This paper will proceed as follows: Section 2 will discuss some of the agile networking considerations for data intensive science. Section 3 introduce the DYNES solution, focusing on the hardware and software interactions of this emerging framework. Section 4 will discuss the current deployment footprint of DYNES. We will conclude and present future work in Section 5.

2. Networking Design for Scientific Use

A delicate balance is required to serve both long-lived, high capacity network flows, as well as more traditional end-user activities. Traditional local area network (LAN) designs favor the important notions of “protection” and “availability”; campus environments will place emphasis on deploying firewalls to protect the client machines from malicious attacks, and packet shaping technology to silently reduce the overall consumption of bandwidth. Keeping these well stated factors in mind, R&E networking needs to find a way to support scientific demands. Many innovations have emerged, spanning hardware technologies, protocols, software, and services; all of which were directly targeted to domain researchers.

To facilitate the bandwidth demands of growing scientific communities, network design has evolved to address two key, yet diametrically opposed, areas:

- **Enterprise requirements**, including the general population of desktops, laptops, and mobile devices on a given network
- **Science requirements**, which encompass the data centers and instruments that have a direct role in the collection, storage, and processing of data sets

Figure 1 shows a simplified campus LAN, highlighting the “complete protection” afforded by a single device positioned on the border. This design affords the same treatment to all users — students in dormitories surfing the web as well as the supercomputing center with sophisticated instrumentation. While effective in protecting users and network resources, data intensive science and remote collaboration may suffer unexpected consequences due to the overall architectural considerations.

Applications designed to move large amounts of data are often based on TCP. FDT, a tool used as the basis for data movement in the LHC Virtual Organization (VO), utilizes multiple streams of TCP traffic to transmit data. While this is an effective way to consume more of the available resources, it may be unfair to other users, and simply creates an “arms race” for use of the network.

Firewalls and packet shaping devices have a profound effect on TCP performance, particularly as the RTT between locations increases [14]. These devices often have small memory buffers, which hampers the ability to handle a single large flow, not to mention multiple flows of various sizes. Architectural changes, including the concept of a “science DMZ”, offer a vast improvement

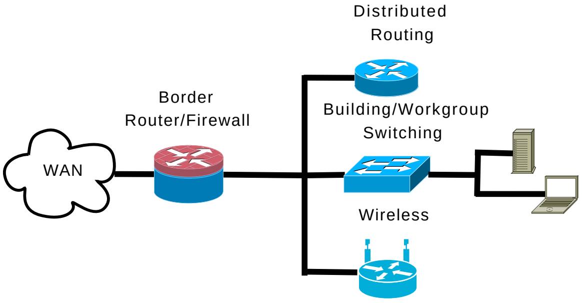

Figure 1. Traditional Campus Network

to data movement activities by reducing the number of disruptive devices on the local portion of the path, and allow storage and processing a more direct access method to the WAN.

Enterprise protection requirements can still be met, as demonstrated in Figure 2. Additionally, the use of access control lists on routing devices and per-host and per-service security settings can offer similar functionality to a traditional firewall device at a much lower overhead.

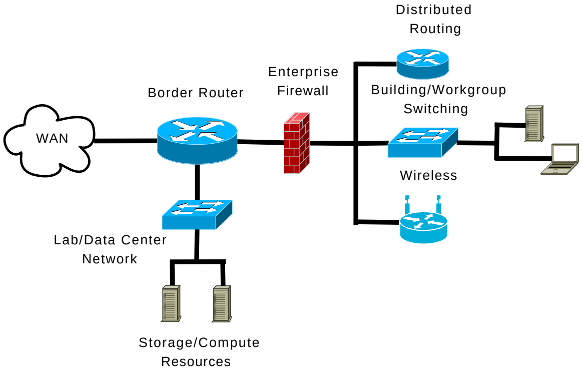

Figure 2. Network Featuring a Science DMZ

These simple architectural measures have had a profound effect on functionality of the network for research use. Additional support, provided by advanced software and services, can further improve operational concerns on R&E networks.

3. DYNES Specifics

The effective design of networks to support scientific activities addresses many key issues, namely by encouraging the development of a physical separation of traffic classes and broader hardware management strategies. These steps alone will often lead to performance gains in a local setting; scientific innovation, however, is no longer a locally based concern. Global collaborations rely on the existence of worldwide R&E networking infrastructure as well as the assurance of end-to-end capacity and performance. This broad view applies to all networks - from the largest providers to the smallest consumers. Backbone networks are designed to transit traffic from regional aggregation points, each serving countless facilities. These far reaching networks, designed to offer large amounts of available capacity, may become congested in certain geographical areas experiencing high demand. International links, funded by parties including the DOE and NSF, link continents and countries to enable distributed activities [15, 16]. The sum of available networking resources globally remains high, and continues to grow. Effective use of the available resources continues to be a struggle that is a actively being addressed in the network research community.

The ability to create a protected path, separate from the general-purpose IP traffic and using existing network resources, remains highly desirable for bulk data movement applications; particularly those relying on the TCP and operating on high RTT and high bandwidth paths. This new "circuit" based technology, called as such due to the use of end-to-end Virtual Local Area Neworks (VLANs), was designed with the WAN in mind and has been slow to migrate to smaller scales. The DYnamic NEtwork System (DYNES) addresses this technology gap by providing hardware and software solutions to regional and campus networks — the overarching goal being to extend the technology already available on many backbone networks. This NSF sponsored project is developing and deploying a nationwide "cyber-instrument", designed to span approximately 40 US universities and 11 regional networks. DYNES was awarded to a collaborative team including Internet2, the California Institute of Technology, the University of Michigan, and Vanderbilt University in 2010. The DYNES team will partner with the LHC and astrophysics communities, the OSG, and Worldwide LHC Computing Grid (WLCG) to deliver these capabilities to the LHC experiment as well as others such as LIGO, and the SDSS/LSST astronomy programs, broadening existing Grid computing systems by promoting the network to a reliable, high performance, actively managed component [17, 18, 19, 20, 21].

By integrating existing and emerging protocols, software for dynamic circuit provisioning and scheduling, in-depth end-to-end network path and end-system monitoring, and higher level services for management on a national scale, DYNES will allocate and schedule channels with bandwidth guarantees to several classes of prioritized data flows with known bandwidth requirements, and to the largest high priority data flows, enabling scientists to utilize and share network resources effectively. DYNES is designed to support many data transfers which require aggregate network throughput between sites of 1-20 Gbps, rising to the 40-100 Gbps range as the underlying network technology is upgraded. This capacity will enhance researchers' ability to distribute, process, access, and collaboratively analyze 1 to 100 TB datasets at university-based Tier2 and Tier3 centers now, and PB-scale datasets in the future.

DYNES is based on a "hybrid" packet and circuit architecture composed of Internet2's ION service and extensions over regional and state networks to US campuses. It will connect with transoceanic (IRNC, USLHCNet), European (GÉANT), Asian (SINET3) and Latin American (RNP and ANSP) R&E networks through the aid of related efforts including IRNC DyGIR [22]. DYNES will build on existing key open source software components that have already been individually field-tested and hardened in part by the PIs: DCN Software Suite (OSCARS/DRAGON), perfSONAR, UltraLight Linux kernel, and FDT.

3.1. DYNES Hardware

The DYNES framework requires 3 key pieces of hardware:

- A network device (e.g. a switch) that can be dynamically controlled
- A controller to integrate with the OSCARS control plane
- A data movement server, capable of storing large amounts of data and utilizing dual network connections

Figure 3 shows these components in a block diagram. The OSCARS software, described in Section 3.2, functions as the software "glue", linking together the networks participating in dynamic control, and allowing the creation of dynamic end-to-end circuits. FDT, described in Section 3.3, functions end to end — allowing the data movement servers provided by DYNES (as well as other existing compute and storage resources) to stage and migrate information over short term circuits. perfSONAR, described in Section 3.4 is available to monitor the status and health of the network participants.

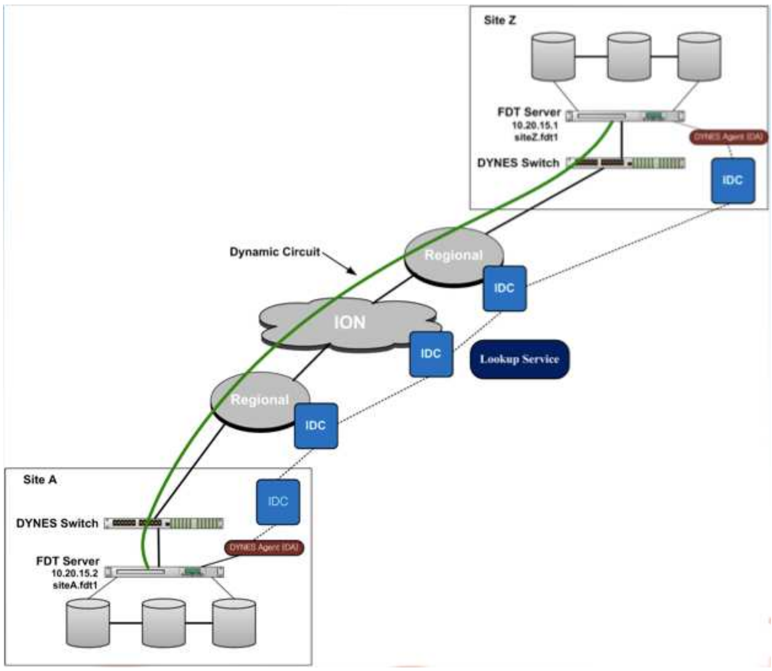

Figure 3. DYNES Hardware

3.2. OSCARS

The On-Demand Secure Circuits and Advance Reservation System, or OSCARS, provides multi-domain, high-bandwidth virtual circuits that guarantee end-to-end network data transfer performance [24]. Originally a research concept, OSCARS has grown into a robust delivery mechanism in the form of ESnet Science and Data Network (SDN) and the Internet2 ION Service. OSCARS virtual circuits carry fifty percent of ESnets annual 60 petabytes of traffic. This segmentation of network resources has had a profound effect on research innovation — users are free to worry about science instead of becoming experts at network design.

The dynamically provisioned, multi-domain, guaranteed bandwidth circuits provided by OSCARS extend Layer 2 VLAN concepts across a complete end-to-end path. Muti-domain circuit operation is accomplished through a series of protocols designed to negotiate on various resources. Each domain maintains local control over all components, and has the ability to set policy regarding use; this is shown in Figure 4. OSCARS servers as a first generation of the concept of "Software Defined Networking", of which the OpenFlow project has emerged. OpenFlow attempts to standardize APIs for network switching device. By creating a standard way to interact with network devices, innovative applications can make intelligent choices for data movement; similar to the overlay networks that are being created by tools such as OSCARS currently.

The concept of virtual circuits integrates cleanly with existing networking solutions, and often co-exists on the same physical infrastructure. Thus it becomes possible to make a conscious decision to switch a target machine's connect between a dynamically provisioned circuit or the Layer 3 infrastructure with simple networking commands. APIs and web-based toolkits are also available to ease adoption further; applications can be modified to use either functionality [13]. Prior demonstrations of this technology have shown use cases both as a way to manually dedicate resources in a traffic management capacity, but also dynamically to support applications including data transfer and high definition video [26, 27].

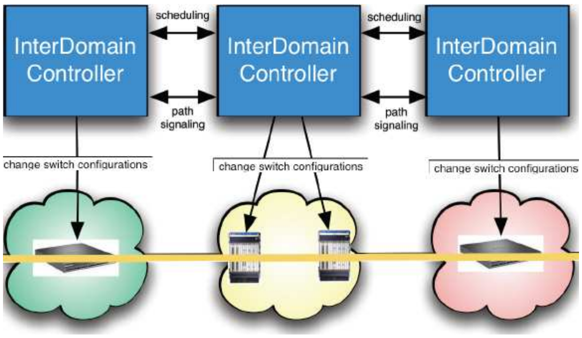

Figure 4. OSCARS in Action: Controlling Switch Operation

3.3. FDT

FDT is an application designed to enable efficient data transfers; performance emphasis is placed upon the interaction between the hosts, their 'internal' components (e.g. disk, bus, processor, and network card) and the wide area network. FDT is written in the JAVA programming language, which has enabled deployment on many major operating system and architectural platforms.

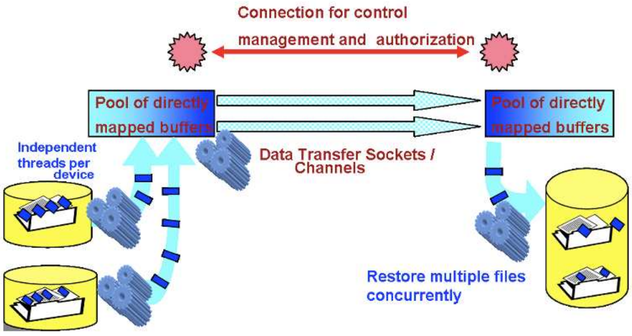
Connection for control
management and authorization
Pool of directly mapped buffers
Pool of directly mapped buffers
Independent threads per device
Data Transfer Sockets / Channels
Restore multiple files concurrently

Figure 5. FDT Operation

Figure 5 demonstrates the typical use case for FDT. FDT is based on an asynchronous, flexible multi-threaded system and includes the following features:

- Streams a complete dataset (e.g. a list of files) continuously, using a managed pool of buffers through one or more TCP sockets on the host machine
- Uses independent threads to read and write on each physical device
- Can transfer data in parallel using multiple TCP streams
- Optimizes the interaction between network and disk I/O

FDT has embraced the concept of dynamic circuit networking by integrating with the OSCARS control framework through the use of programmatic APIs. This mode of operation allows the FDT agent on a host to transmit data using traditional Layer3 infrastructure, or invoke an on-demand circuit to perform a point to point transfer for short periods of time. FDT is capable of requesting bandwidth, and managing the nuances of the circuit, freeing the user to ignore details about the underlying network.

3.4. perfSONAR

The PERFormance Service Oriented Network monitoring Architecture, or perfSONAR, is a federated network monitoring framework, that facilitate end-to-end sharing of performance measurement data. This component based suite of tools decouples the tasks of measuring network performance from mechanisms used to store data, share and visualize results, and authenticate user permissions [28]. perfSONAR has been widely adopted by the R&E education community, and is used to expose active and passive network measurements on networks of all sizes, as well as for resources managed at end sites including scientific VOs

Measurement data, provided through a mechanism such as perfSONAR, is vital to network aware applications including remote collaboration tools and software designed for the task of data movement. Knowledge of current performance, as well as historical trends, can be used to making routing decisions, invoke dynamic connectivity, or schedule future data movement activities [29].

4. Deployment Status

DYNES is an NSF sponsored project, under the "Major Research Instrumentation" (**MRI**) program, and is tasked with developing and deploying the components of a nationwide "cyber-instrument". Original plans had designed DYNES to span approximately 40 US universities and 11 regional networks. Figure 6 shows the current deployment on a scale model of the United States. It is important to note that certain regional networks are participating in "static" fashion, e.g. they are functional for creating circuits but not in a static operational method. Certain VLANs are available for use and are pre-configured.

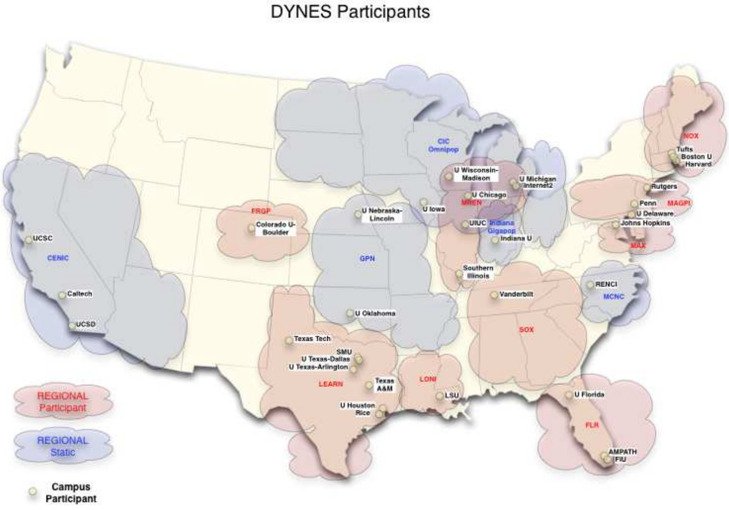

Figure 6. Current DYNES Deployment

A logical diagram, shown in Figure 7, shows the current DYNES locations as well as other reachable facilities and networks. Locations on the ESnet SDN, and extending into continental Europe and South America, can be reached using the available technology.

The DYNES project is funded through August of 2013. Remaining tasks for the project include, but are not limited to:

- Increasing the total number of deployed regional networks and end sites
- Working with application communities to adopt Layer 2 networking paradigms
- Encouraging the deployment of network architectural changes, including the *Science DMZ*

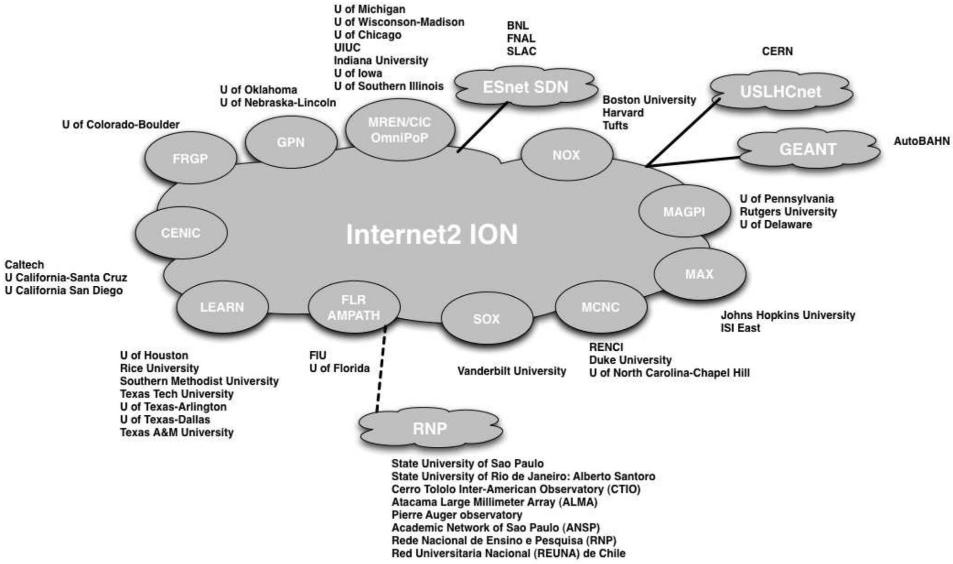

Figure 7. Current DYNES Logical Topology

Completion of funding does not end the usefulness of DYNES. It is expected that the underlying Layer 2 technologies will continue to advance, and may be based on the aforementioned "Software Defined Networking" technologies, including *OpenFlow*. Future hardware upgrades will enable this on exiting DYNES deployments, along with software projects devoted to enabling these technologies into future applications.

5. Conclusion

The technological realities of scientific collaboration are trending toward increased use of network resources. The R&E networking community has responded with several solutions that work in concert:

- Increased capacity on the backbone, regional and campus networks
- Suggestions to design dual-use network architectures to serve scientific users as well as the general population
- Innovative technologies to ensure end-to-end performance for high capacity network use cases

DYNES, an infrastructure designed to deliver a hardware and software based solution to end users and regional networks, is in the process of deploying a nationwide instrument to address the needs of the scientific community. This infrastructure integrates with similar efforts already in place on backbone networks, and will inter-operate with similar efforts designed to address international connectivity. Recent activities to demonstrate this capability have shown that this technology is ready for production networks, and capable of meeting the needs of scientists.

6. Acknowledgements

The authors would like to thank staff members from the California Institute of Technology, Internet2, the University of Michigan, and Vanderbilt University for their assistance in the operation and construction of DYNES. The DYNES project is also grateful for the contributions of the FDT, OSCARS, and perfSONAR-PS projects — the underlying software set that controls DYNES. Lastly, the authors would like to acknowledge Dell Computers for assisting the DYNES team in our evaluation and purchase of hardware for this effort. These contributions have been crucial to the overall success in designing and implementing the DYNES framework.

The authors would also like to acknowledge funding bodies that support this ongoing work. The DYNES project is supported by the NSF grant **0958998**. Related work for the DISUN, Ultralight, and PlaNetS projects is sponsored by NSF grants **PHY-0533280**, **PHY-0427110**, and **PHY-0622423**. The Department of Energy (DOE) sponsors US LHCNet under grant **DE-FG02-08ER41559**.

References

[1] IOP Publishing is to grateful Mark A Caprio, Center for Theoretical Physics, Yale University, for permission to include the `iopart-num` BibTEXpackage (version 2.0, December 21, 2006) with this documentation. Updates and new releases of `iopart-num` can be found on `www.ctan.org` (CTAN).

[2] MRI-R2 Consortium: Development of Dynamic Network System (DYNES). `http://www.internet2.edu/ion/dynes.html`.

[3] LHC Optical Private Network. `http://lhcopn.web.cern.ch/lhcopn/`.

[4] LHC Open Network Environment. `http://lhcone.net/`.

[5] ESnet Network and Science Requirement Workshops Reports. `http://www.es.net/about/science-requirements/reports/`.

[6] Science DMZ. `http://fasterdata.es.net/fasterdata/science-dmz/`.

[7] perfSONAR-PS. `http://psps.perfsonar.net`.

[8] ESnet SDN. `http://www.es.net/hypertext/network.html`.

[9] GEANT2 AutoBAHN. `http://www.geant2.net/server/show/ConWebDoc.2544`.

[10] Internet2 ION. `http://www.internet2.edu/ion/`.

[11] Internet2 ION Case Study: Large Hadron Collider (LHC). Case study, Internet2, September 2009. http://www.internet2.edu/pubs/200909-CS-ION-LHC.pdf.

[12] I. Monga, C. Guok, W. Johnston, and B. Tierney. Hybrid networks: Lessons learned and future challenges based on esnet4 experience. *IEEE Communications*, pages 114–121, May 2011.

[13] Fast Data Transfer (FDT). `http://monalisa.cern.ch/FDT`.

[14] M. Mathis. Pushing up Performance for Everyone. Presentation, Internet2, December 1999. http://staff.psc.edu/mathis/papers/.

[15] USLHCNet. `http://lhcnet.caltech.edu/`.

[16] International Research Network Connections. `http://irnclinks.net/`.

[17] Laser Interferometer Gravitational Wave Observatory. `http://www.ligo.caltech.edu/`.

[18] Large Synoptic Survey Telescope. `http://www.lsst.org`.

[19] Sloan Digital Sky Survey. `http://www.sdss.org/`.

[20] Open Science Grid. `http://www.opensciencegrid.org/`.

[21] Worldwide LHC Computing Grid. `http://lcg.web.cern.ch/lcg/`.

[22] National Science Foundation International Research Network Connections (IRNC) Awards. Case study, Internet2, October 2010. http://www.internet2.edu/pubs/201010-IS-IRNC.pdf.

[23] DYNES Hardware. `http://www.internet2.edu/ion/hardware.html`.

[24] ESnet OSCARS. `http://www.es.net/services/virtual-circuits-oscars/`.

[25] OpenFlow - Enabling Innovation in Your Network. `http://www.openflow.org/`.

[26] A. Hutanu, J. Ge, C. Toole, R. Paruchuri, A. Yates, and G. Allen. Distributed Visualization Using Optical Networks: Demonstration at Super-computing 2008. Technical report, LSU CCT, October 2008. http://www.cct.lsu.edu/CCT-TR/CCT-TR-2008-10.

[27] N. Charbonneaum, V. Vokkarane, C. Guok, and I. Monga. Advance reservation frameworks in hybrid ip-wdm networks. *IEEE Communications*, pages 132–139, May 2011.

[28] A. Hanemann, J. Boote, E. Boyd, J. Durand, L. Kudarimoti, R. Lapacz, M. Swany, S. Trocha, and J. Zurawski. Perfsonar: A service oriented architecture for multi-domain network monitoring. In *Third International Conference on Service Oriented Computing - ICSOC 2005, LNCS 3826, Springer Verlag*, pages 241–254, Amsterdam, The Netherlands, December 2005.

[29] R. Wolski. Dynamically forecasting network performance using the network weather service. *Journal of Cluster Computing*, pages 119–132, January 1998.