

IRNC ProNet: DYNESTAR: A Dynamic Network System for Transatlantic Research

Project Summary

Intellectual Merit: Caltech proposes, together with CERN, US LHCNet, Internet2, NLR, ESnet, GLORIAD and major regional and state network partners throughout the US, the NRENs and associated projects such as Geant3 in Europe and partners in Latin America (RNP, ANSP), to deploy and develop DYNESTAR, the first fully resilient, dynamic high performance transatlantic hybrid network for data intensive science. DYNESTAR will interconnect the GLIF Open Lambda Exchanges across the Atlantic, including StarLight/TransLight in the US, and CERNLight, Netherlight, Northernlight, CzechLight, UKLight, MoscowLight and the planned MarseillesLight, as well as other exchange points overseas. By integrating emerging standard protocols for dynamic circuit provisioning, network path and end-system monitoring, and services for circuit management and scheduling from UltraLight and partners such as the proposed FENRIR project, DYNESTAR will allocate channels with bandwidth guarantees to prioritized data flows, enabling thousands of scientists to utilize network resources more effectively.

A focus of DYNESTAR will be to support the long-range multi-terabyte scientific data flows generated by the LHC program, in particular the university-based Tier2 centers, complementing DOE's US LHCNet whose main focus is Tier0-Tier1 and Tier1-Tier1 operations. DYNESTAR will also serve other leading programs such as LIGO, the Virtual Observatory, large scale sky surveys, the Earth System Grid and the broader scientific community through the use of standard (IDC-based) dynamic circuits.

DYNESTAR will build on key open source software components that have already been deployed, field-tested and hardened in part by our team: the Internet2 DCN Software Suite (OSCARS/DRAGON), perfSONAR, the UltraLight Linux kernel and high throughput applications such as FDT. Within the hybrid DYNESTAR architecture, existing applications such as LambdaStation and Terapaths will direct flows from selected end-systems towards the dynamic circuits, on a per-flow basis. The system will also support standalone services and applications targeted at individuals and small groups, to enhance the capabilities of scientists in many data intensive fields on campuses across the U.S. and Europe.

We are requesting funds from the NSF IRNC program for transatlantic links that are the key to success of DYNESTAR: initially two OC-192 links in 2010 that will interoperate with the existing and emerging infrastructures to create a resilient mesh of approximately 8 transatlantic and 8 continental OC-192 links. As DYNESTAR and its partners progress to 40 Gbps and 100 Gbps next-generation networks, the NSF-funded part of DYNESTAR will reach 100Gbps of transatlantic capacity by 2014, with an anticipated contribution on the same scale from our partner networks. This bandwidth will be complemented by US LHCNet for Tier0 and Tier1 operations, and 100G links in the US and Europe provisioned by the R&E network partners mentioned above by that time.

DYNESTAR intends to collaborate with the IRNC:Exp project FENRIR, that will build on the DRAGON results to develop a generally applicable dynamic cyber-resource provisioning, control and management service. Acknowledging its importance and potential direct benefits to DYNESTAR and the broad community served, DYNESTAR will dedicate 1 Gbps between the US and Europe as part of the global FENRIR test-bed, along with the other FENRIR partners.

Broad Impact: In addition to dynamic end-to-end circuits, DYNESTAR will apply its resilient Layer 1 mesh-restored connections to support peerings at Layer3 across the Atlantic with an exceptional availability: in excess of 99.9% based on US LHCNet experience. DYNESTAR will work with Internet2 and NLR, ESnet to the DOE HEP labs, and the NRENs and GEANT3 in Europe via these peerings, as an integral part of a robust, global architectural solution, to provide the bandwidth and data transfer services needed by hundreds of physics groups at Tier2 and Tier3 sites. Through adoption of the GLIF open lightpath exchange policies, it will work with TransLight/StarLight to provide the broadest reach possible, linking R&E networks and research teams in many fields of science in the US, Europe, Asia and Africa.

DYNESTAR will deliver these capabilities to the broader scientific community by coupling dynamic circuits, high throughput data transport and end-to-end monitoring to the grid-based analysis systems. Leading examples are ATLAS, CMS and the other LHC experiments, where DYNESTAR will greatly improve the performance of data analysis. DYNESTAR will amplify the capabilities of Open Science Grid by providing the highly reliable network services essential for efficient grid operations.

Project Description

A. Introduction

Intellectual Merit: Caltech proposes, together with CERN, US LHCNet, Internet2, NLR, ESnet, GLORIAD, and major regional and state network partners throughout the US, the NRENs and associated projects such as Geant3 in Europe and partners in Latin America (RNP, ANSP), to deploy and develop DYNESTAR, the first fully resilient, dynamic high performance transatlantic hybrid network for data intensive science. DYNESTAR will interconnect the GLIF Open Lambda Exchanges across the Atlantic, including StarLight/TransLight in the US, and CERNLight, Netherlight, Northernlight, CzechLight, UKLight, MoscowLight and the planned MarseillesLight, as well as other exchange points overseas. By integrating emerging standard protocols for dynamic circuit provisioning, network path and end-system monitoring, and services for circuit management and scheduling from UltraLight and partners such as the proposed FENRIR project, DYNESTAR will allocate channels with bandwidth guarantees to prioritized data flows, enabling thousands of scientists to utilize network resources more effectively.

A focus of DYNESTAR will be to support the long-range multi-terabyte scientific data flows generated by the LHC program, in particular the university-based Tier2 centers, complementing DOE's US LHCNet whose main focus is Tier0-Tier1 and Tier1-Tier1 operations. DYNESTAR will also serve other leading programs such as LIGO, the Virtual Observatory, large scale sky surveys, the Earth System Grid and the broader scientific community through the use of standard (IDC-based) dynamic circuits.

DYNESTAR will build on key open source software components that have already been deployed, field-tested and hardened in part by our team: the Internet2 DCN Software Suite (OSCARS/DRAGON), perfSONAR, the UltraLight Linux kernel and high throughput applications such as FDT. Within the hybrid DYNESTAR architecture, existing applications such as LambdaStation and Terapaths will direct flows from selected end-systems towards the dynamic circuits, on a per-flow basis. The system will also support standalone services and applications targeted at individuals and small groups, to enhance the capabilities of scientists in many data intensive fields on campuses across the U.S. and Europe.

We are requesting funds from the NSF IRNC program for transatlantic links that are the key to success of DYNESTAR: initially two OC-192 links in 2010 that will interoperate with the existing and emerging infrastructures to create a resilient mesh of approximately 8 transatlantic and 8 continental OC-192 links. As DYNESTAR and its partners progress to 40 Gbps and 100 Gbps next-generation networks, the NSF-funded part of DYNESTAR will reach 100Gbps of transatlantic capacity by 2014, with an anticipated contribution on the same scale from our partner networks. This bandwidth will be complemented by US LHCNet for Tier0 and Tier1 operations, and 100G links in the US and Europe provisioned by the R&E network partners mentioned above by that time.

DYNESTAR intends to collaborate with the IRNC:Exp project FENRIR [4], that will build on the DRAGON results to develop a generally applicable dynamic cyber-resource provisioning, control and management service. Acknowledging its importance and potential direct benefits to DYNESTAR and the broad community served, DYNESTAR will dedicate 1 Gbps between the US and Europe as part of the global FENRIR test-bed, along with the other FENRIR partners.

Broad Impact: In addition to dynamic end-to-end circuits, DYNESTAR will apply its resilient Layer 1 mesh-restored connections to support peerings at Layer3 across the Atlantic with an exceptional availability: in excess of 99.9% based on US LHCNet experience. DYNESTAR will work with Internet2 and NLR, ESnet to the DOE HEP labs, and the NRENs and GEANT3 in Europe via these peerings, as an integral part of a robust, global architectural solution, to provide the bandwidth and data transfer services needed by hundreds of physics groups at Tier2 and Tier3 sites. Through adoption of the GLIF open lightpath exchange policies, it will work with TransLight/StarLight to provide the broadest reach possible, linking R&E networks and research teams in many fields of science in the US, Europe, Asia and Africa.

DYNESTAR will deliver these capabilities to the broader scientific community by coupling dynamic circuits, high throughput data transport and end-to-end monitoring to the grid-based analysis systems. Leading examples are ATLAS [1], CMS [2] and the other LHC experiments, where DYNESTAR will greatly improve the performance of data analysis. DYNESTAR will amplify the capabilities of Open

Science Grid [3] by providing the highly reliable network services essential for efficient grid operations.

B. Organization Description

DYNESTAR will provide, through collaboration with its partners TransLight/StarLight, SURFnet and NORDUnet, the transatlantic bandwidth and circuit-oriented services needed to bridge the US and European R&E networks. By interconnecting major lightpath exchanges on both continents with high capacity links, it will also extend the reach to other regions of the world. DYNESTAR will leverage, enhance and expand upon, the existing US LHCNet infrastructure, profiting from both cost sharing as well as the operational structure proven to be highly efficient during the past 5 years.

DYNESTAR's proponents have a unique range of capabilities, spanning production networking, dynamic circuit development in the context of DICE and GLIF as well as resilient virtual circuits based on VCAT [6] and LCAS [7] standards across the Atlantic, next generation network development and deployment (including 100G network trials), and the development and deployment of the state of the art in high throughput applications (such as FDT [8], [9], [10] and the UltraLight Linux kernel).

Organizationally, DYNESTAR will integrate into the USLHCNet management and operation. USLHCNet is co-managed by Caltech and CERN, as detailed in Section H. In addition to its work in DICE and GLIF on dynamic circuit software and methods, Caltech's roles include: originating transatlantic networking for high energy physics in 1982; together with CERN and its partners in Europe, providing transatlantic networking for HEP since 1988; developing US LHCNet together with CERN since 1995; developing the state of the art for high-throughput open-source data transfer applications and tools while leading the NSF-funded UltraLight and PlaNetS projects (Fast Data Transfer and the UltraLight kernel); founding along with its US and European partners, the open hybrid circuit- and packet network architecture widely adopted by the major R&E networks in the US and Europe in 2004; deploying, operating and developing the first layer 1 fully resilient circuit-oriented network in production since 2007.

CERN has been engaged in advanced wide area networking for the last 30 years, is the originator of the World Wide Web and the leader of the LHCOPN. As a partner in US LHCNet, DICE and GLIF, it remains at the forefront, together with its US and European partners, of state of the art networks and grid systems for research and education, with a focus on the LHC program.

In the US, US LHCNet collaborates with ESnet and Internet2 as well as the Ultralight/PLaNetS project to provide end-to-end services to the research community, both IP routed as well as circuit switched. DYNESTAR will integrate from day one into this collaboration, extending the common service portfolio to reach a wider range of end sites both in the US, notably TransLight/StarLight, as well as in Europe.

In the scope of the DYNESTAR project, Caltech will partner in Europe with CERN, SURFnet [13], NORDUnet [14] and DANTE [12]. SURFnet, in collaboration with CERN, will contribute to the DYNESTAR project by providing capacity on its dark fiber, lit with 40 and later 100Gbps waves, between the Netherlight and CERNlight exchanges, creating a mesh topology and increasing the overall bandwidth and resilience. DYNESTAR will partner with NORDUnet through capacity exchange and interconnection at Lightpath exchanges on both continents. NORDUnet's new transatlantic circuit between New York and Copenhagen (part of the NSF sponsored TAJ project [16]) to be commissioned in 2009, will also connect to Greenland and Iceland. Through the NorthernLight exchange point in Copenhagen, DYNESTAR will also be able to reach NORDUnet's connected research sites, as well as connect further through RUNNet/GLORIAD to MoscowLight.

DYNESTAR will collaborate with DANTE through peering arrangements with its GEANT3 network, guaranteeing IP connectivity with NRENs not directly present at the Lightpath Exchanges, as well as through dynamic circuit connections with its AutoBAHN system, providing Layer 2 connectivity between end-sites in the US and Europe.

B.1. Path to 40/100Gbps Technology Deployment

In the US, ESnet and Internet2 are collaborating with industrial partners on future 100Gbps based backbone on national footprint. ESnet's Advanced Networking Initiative, funded by DOE's ARRA

program, will start with a prototype wide area network spanning the continent, providing a test bed facility for researchers, and aiming at near-term transfer to production network. Internet2 has issued an RFI to identify potential partners to provide its future 100Gbps backbone.

On the European footprint, efforts are under way for construction of a 100Gbps optical backbone interconnecting open Lightpath exchanges. A first 100G capable dark fibre is being deployed by SURFnet and CERN between the Netherlight exchange point and Geneva, where CERN is preparing a new exchange point, CERNLight. The commissioning of this 100Gbps link between the two sites is planned for September 2009. The fiber will initially be lit by three 40 Gbps channels. 100Gbps links are planned to be introduced in the 2010/2011 timeframe. RENATER, the French NREN, is building a 100Gbps link between CERN and Lyon, with possibility to be extended to a new Lightpath exchange in Marseille giving access to northern Africa and Asia in 2010. DYNESTAR will connect into this infrastructure by initially landing one transatlantic circuit in Amsterdam, and one in Geneva. The topology, including the existing USLHCNet infrastructure is shown in [Figure 4](#) below. The two exchange points (Netherlight and CERNLight) will be interconnected through the SURFnet/CERN dark fibre, providing resilience and robustness of the whole network.

With the above developments in mind, DYNESTAR includes a plan for an upgrade to 100Gbps, once the technology will become available on transatlantic cable systems [17]. We expect this to happen by around 2012/2013. The timeline includes hardware upgrades to 100G capable devices before 2012.

C. Science Impact: Meeting the Needs of Data Intensive Science

C.1. Requirements

High Energy Physics (HEP), and in particular the Large Hadron Collider (LHC) [18] experiments, is a leading example where a revolutionary paradigm shift has occurred within the last 9 years. Instrument-site based computing has given way to globally distributed computing to support the very large collaborative data analysis efforts, and to secure sufficient computing power to analyze the data at the rates generated by the detectors and the subsequent processing steps. All of this depends completely on highly capable, scalable, high speed, highly interconnected, and very reliable networks. The LHC experiments' "Tiered" computing and storage system already encompasses hundreds of sites, each of which hosts from tens of terabytes (Tier3) to hundreds of terabytes (Tier2) to petabytes (Tier1).

Sustained throughputs among the sites at speeds of 1-10 Gbps (and in some cases > 10 Gbps) lasting for days are in production use today by some Tier2 sites as well as the Tier1 sites, particularly in the US. The LHC data volumes and transfer rates are expected to expand by an order of magnitude over the next several years, as higher capacity storage and regional, national and transoceanic network links of 40 and 100 Gbps become available and affordable. The bandwidth use by HEP, which has risen by a factor of 300 - 1000 per decade, is expected to continue its steady exponential climb in the coming years [87]. Transatlantic bandwidth in US LHCNet for example, is expected to reach approximately 280 Gbps by 2014, with similar bandwidths between its points of presence in the US (NYC and Chicago) and Europe (CERN and Amsterdam).

Other fields of data intensive sciences face similar challenges [19] Recent requirements workshops [20] have shown that nuclear physics experiments will require 20-40 Gbps to the BNL and LBNL, 30 Gbps

Year	Production	Experimental	Remarks
2001	0.155	0.622 – 2.5	SONET/SDH
2002	0.622	2.5	SONET/SDH; DWDM; GigE Integr.
2003	2.5	10	DWDM; 1 & 10 GigE Integration
2005	10	2-4×10	λ Switch, λ Provisioning
2007-8	2-4×10	~10×10 (and 100)	1 st Gen. λ Grids
2009-10	6-8×10	~20×10 (and 2×100)	40 Gbps λ Switching
2011-12	~20×10 (or ~2×100)	~10×100	2 nd Gen. λ Grids, Terabit networks
2013-15	~Terabit	~Multi-Terabit	~Fill one fiber

Table 1: Bandwidth Roadmap (in Gbps) for Major HEP Network Links

across the Pacific and 20 Gbps across the Atlantic by 2012, leading to requirements of up to 10 Gbps at major sites such as Vanderbilt and MIT. In a similar time frame, the needs [1] of climatology, the Earth System Grid [21] and a wide range of biological and environmental research areas and subsurface science are expected to be in the 100-250 Gbps range. In fusion energy sciences [22], the per-site bandwidth requirements of presently deployed tokomaks and supercomputer centers are expected to grow to 10-80 Gbps in this time frame, followed by ITER [23] that will require more than 80 Gbps of bandwidth just to keep up with the anticipated flow of real and simulated data (82 Tbytes/hr). The needs per site in basic energy sciences [24] to support remote experimentation, visualization and analysis, at the major synchrotron light sources, the Spallation Neutron Source, and supercomputer studies of combustion, are expected to grow from 4-20 Gbps by 2012, to 10-100 Gbps in the following years.

C.2. Motivating Science Applications

DYNESTAR's resilient infrastructure will deliver the capabilities required to support data intensive science to the scientific community by coupling dynamic circuit provisioning, high throughput data transport and end-to-end monitoring services to their grid-based analysis systems. Leading examples are ATLAS and CMS, where DYNESTAR will help them greatly improve the performance observed in data analysis operations. DYNESTAR will also amplify and broaden the capabilities of the Open Science Grid (OSG), augmenting the OSG software stack with services that build guaranteed bandwidth circuits for high priority data transfer tasks, and monitoring data transfer performance, to optimize the overall throughput among the grid sites. DYNESTAR will use results from the NSF-funded UltraLight [25][26], PLaNetS and DISUN [27] projects along with the DCNSS (see Section D.3) to accomplish these goals.

C.2.1. LHC: Physics and Computing Challenges

High Energy Physics experiments are breaking new ground in understanding the unification of forces, the origin and stability of matter, as well as structures and symmetries that govern the nature of matter in our universe. To improve our understanding of the nature of matter and space-time itself, researchers work to isolate and observe rare events, predicted by a variety of new physics theories that go beyond our current understanding. Even by utilizing highly-processed "analysis object data", the size of an LHC dataset suitable for discovering such rare events is at the terabyte level with similar amounts of Monte Carlo simulated data required for hypothesis testing. Further, assuming a typical LHC data taking period ("Run") of 1-3 hours of stable operations, the size of a canonical RAW dataset is also of order a terabyte. Hence, physicists performing searches for possible new physics discoveries as well as physicists conducting detector calibration and systematic studies, which are vital for establishing the early presence of new physics signals, will frequently request terabyte or larger sized dataset "chunks" for their work. Over time, total HEP data volumes are expected to rise from the multi-Petabyte (10^{15} Byte) to the Exabyte (10^{18} Byte) range within the next 10-15 years, and the corresponding bandwidth requirements on the major links are expected to rise from 10 Gigabit/sec (Gbps) now to the Terabit/sec (Tbps) range during this period, as summarized in the roadmap of major HEP network links [28].

C.2.2. Virtual Observatory: Astrophysics and Computing Challenges.

Astronomy continues to generate data with an exponentially growing rate. There are currently a few petabytes of data on astronomical archives worldwide, with a doubling time of ~1.5 years. This trend will continue in the next decade with several new and massive surveys of the Universe spanning the whole electromagnetic spectrum from γ -rays to X-rays to the ultraviolet, optical and infrared including the Fermi [29], GALEX[32] and JDEM[33], Spitzer[34], Herschel[35], and ground-based surveys. Measurements of the cosmic microwave background and radio spectrum include the WMAP[39] and PLANCK[40] satellites, and a number of ground-based instruments, including eVLA[41], LOFAR[42], ATA[43], and SKA[44] prototype arrays. It is only when these datasets are combined – collating data from several different surveys or matching simulations to observations – that the full scientific potential is realized; the scientific returns from the total will far exceed those from any one individual component. One area of particular growth and interest are synoptic sky surveys, which cover the sky many times looking for moving (e.g., Earth-crossing asteroids) or variable objects (e.g., Supernovae, and other types of cosmic

explosions). Leading examples include the soon-forthcoming Pan-STARRS[47] and SkyMapper[48] surveys (~1 TB/night), and the future LSST[49] (~ 30 TB/night). Real-time detection and follow-up of transient events discovered by these surveys poses special challenges, as astronomers want to be notified of changes in the sky within minutes of a γ -ray burst or supernova, meaning that the data and analysis pipeline must be able to meet these real-time requirements.

C.3. Meeting the Requirements: Hybrid Network Architecture

In order to meet the science requirements, Internet2 and ESnet, along Caltech/US LHCNet, and GEANT2's AutoBAHN project [50] in Europe, have developed a strategy (originated at a meeting in CERN in 2004) based on a dual or hybrid network architecture, where the traditional IP network backbone is paralleled by a second, circuit-oriented core network reserved for large-scale science traffic. Major examples are Internet2's Dynamic Circuit Network (DCN) and ESnet's Science Data Network (SDN). By extending the circuits between US and European end sites across the Atlantic, building on US LHCNet's infrastructure, DYNESTAR will provide:

- 1) ***Increased bandwidth capacity and reliability of network access***, by mutually isolating the large long-lasting flows (on the DCN and/or SDN) and the traditional IP mix of many small flows
- 2) ***Guaranteed bandwidth as a service*** by building a system to automatically schedule and implement virtual circuits traversing the network backbone, and
- 3) ***Improved ability of scientists to access network measurement data*** for all the network segments end-to-end that are critical to their science through the perfSONAR monitoring infrastructure.

The separation of the circuit-oriented network from the IP backbone also is driven by the need to meet different functional, security, and architectural needs:

- 1) The general purpose campus networks must accommodate a huge number of disparate devices, and open access means supporting flows from many unidentified sources, while guarding against improper exploitation. Firewalls are therefore used, which are in general not architected to support very large traffic volumes if there are very large individual flows. The science network, on the other hand, can be limited to accept traffic only from authenticated, authorized sources. Using digital certificates to establish identity obviates the need for firewalls.
- 2) While high capacity traditional layer 3 routers are needed to support the huge number of routes inherent in general purpose internet traffic, supporting the rapidly growing volumes of science data traffic on these routers would be very expensive due to the high cost of their 10 Gbps network interfaces. This is expected to continue in next generation networks, where 100 Gbps router port prices are expected to remain in the several hundred k\$ range for the next few years.

The need for dynamic, scheduled circuits also is clear. Static "nailed-up" circuit architectures cannot scale to meet the demand. In DYNESTAR, each Tier2 site will typically aim for 3-10 Gbps to complete typical dataset transfers of 30 TBytes in a 24 hour day or less. It would be impractical to set up a mesh of "nailed up" circuits of 1 and 10 Gbps to serve the more than 100 Tier2 sites.

D. DYNESTAR System Design

D.1. Functional Description

DYNESTAR will extend hybrid network and dynamic circuit capabilities among campuses and laboratories across the Atlantic, providing several critical functions for effective network use to the Tier2 sites involved, as well as the national and international networks: (1) Network resource allocation with bandwidth guarantees to ensure performance of data transfers, (2) Monitoring of the network and data transfer performance, (3) A request infrastructure, where grid applications can request bandwidth and provide information on the estimated data transfer volume, and policy-based information such as relative priority within a given virtual organization (VO), (4) Adjustment of dynamic circuits in response to link outages or impairment, or shifting workloads. In the transatlantic core, as in done in 24 X 7 operation in USLHCNet today, this will be done in a "hitless" manner at Layer 1 using VCAT and LCAS, while

maintaining non-stop operation and (5) High throughput data transfers also will be enabled through state of the art applications, such as Caltech's Fast Data Transfer **Error! Reference source not found.**

Figure 1 shows a typical example: a 50 Terabyte transfer of the type that DYNESTAR will support between a Tier 2 and a Tier 1 site. Similar transfers occur frequently at both Nebraska and Caltech, as well as UCSD, Michigan and a growing list of other Tier2 sites.

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 and (as above) at a number of Tier2 sites. A companion proposal to this one, for a national Dynamic Network System (DYNES), has

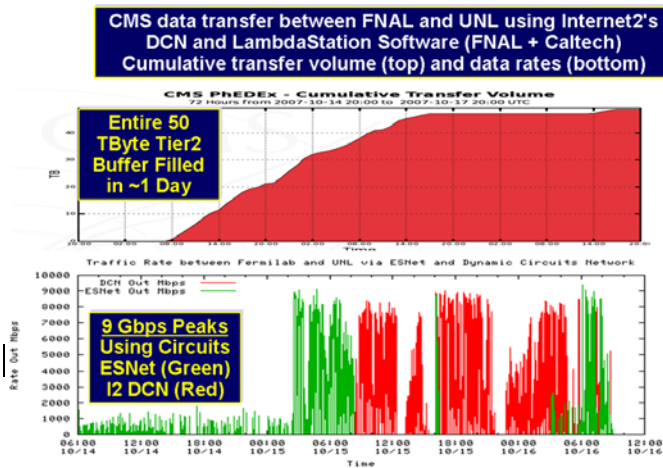


Figure 1: Restoring a 50 TB dataset (FNAL – Nebraska) over Internet2's DCN (Red) and ESnet (Green).

to guarantee end-to-end connectivity between campuses, research sites and data centres.

To do so, DYNESTAR will provide a share of its routed bandwidth to General Purpose Networking, complementing, and collaborating with, TransLight/StarLight, NORDUnet, SURFnet and GEANT3's IP connectivity.

Figure 3 shows the topology at Layer 1 through Layer 3. At Layer 1 and 2, the infrastructure consists of Ciena [51] CoreDirector multiservice switches, interconnected through OC-192 circuits, providing the resilient fabric for Layer 2 and 3 connections. At Layers 2 and 3, Force10 E600 switch-routers are interconnected in full-mesh topology through virtual circuits provided by the CoreDirector switches. By sharing the infrastructure with US LHCNet, DYNESTAR will be integrated with, and will expand upon its highly resilient fabric. All services are virtualized; the virtual circuit connections follow a default path defined during instantiation, but are mesh protected within the whole network. The services DYNESTAR will provide are therefore based on provisioning bandwidth, not on the use of a particular physical link. The key advantage of this field-proven approach (in daily production in US LHCNet) is non-stop operation. In case of a link outage, the connectivity and all services are maintained, with bandwidth automatically re-allocated among the virtual circuits according to policy and priorities.

Link Capacity Adjustment Scheme (LCAS) is well implemented on the CIENA platform, and had been field-tested in operation of US LHCNet. It allows modifying the allocated capacity of a virtual circuit without impact on its operational state. US LHCNet uses this feature for additional resiliency in case of link failures. In such a scenario, where only part of a virtual circuit can be restored on a backup path, the circuit can remain operational at reduced bandwidth, if configured to do so. Given the multiplicity of links already deployed on several geographically diverse paths as well as those planned, and to the use of the US and European networks for intra-continental connectivity, the impact of a cable cut or other disruption of one or even two links will be relatively minor.

The CIENA CoreDirector multiservice switches provide an advanced platform supporting Layer 1 and Layer 2 operation. US LHCNet virtualized network strategy makes best use of all advanced features provided. Virtual Concatenation (VCAT) is used to provide scalable virtual circuits between end points

been submitted to the NSF-MRI R2 program by Internet2, Caltech, Michigan and Vanderbilt. DYNES would extend DYNESTAR's transatlantic capabilities across the Internet2 DCN nationwide, reaching 39 U.S. campuses via 16 state and regional networks.

D.2. Production Network Design and Services

DYNESTAR will provide dynamic circuit services among network domains, as well as IP routed connections. As shown in **Figure 2**, the proposed infrastructure will bind seamlessly into the fabric of GLIF Open Lightpath Exchanges (GOLEs) for maximum reach and collaboration with Research and Education Networks on both continents and beyond. The goal is to provide connectivity not only between the networks, but

with bandwidth between 150Mbps and 9.5Gbps (OC-192) in steps of 150Mbps. Mapping between the Ethernet and TDM domain is done on the advanced Ethernet Service Line Module (ESLM), based on VLAN tag of the incoming frames. VLAN translation is supported on the ESLM, an important feature for the scalability of multi-domain dynamic circuit network. Contrary to QoS implementations in routers and Ethernet switches, the bandwidth segregation is “hard” due to the fixed allocated number of time slots per virtual channel. The ESLM functionality includes also the mapping of multiple VLANs onto a single virtual circuit. “Soft” prioritization of the flows within such a shared virtual circuit is provided through Class of Service (CoS) profiles.

Mesh protection is used as a highly efficient way to provide protection. While traditional protection mechanisms like 1:N and ring protection necessitate the availability of unused resources along the same path as the primary circuits, mesh protection allows for a more flexible scheme. A single transoceanic circuit, along with terrestrial segments is enough to provide protection for the entire network in case of a single link failure. Together with VCAT and LCAS, this protection capacity does not need to be located on a single link, but any free time slot can be used for protection purposes, again adding more flexibility.

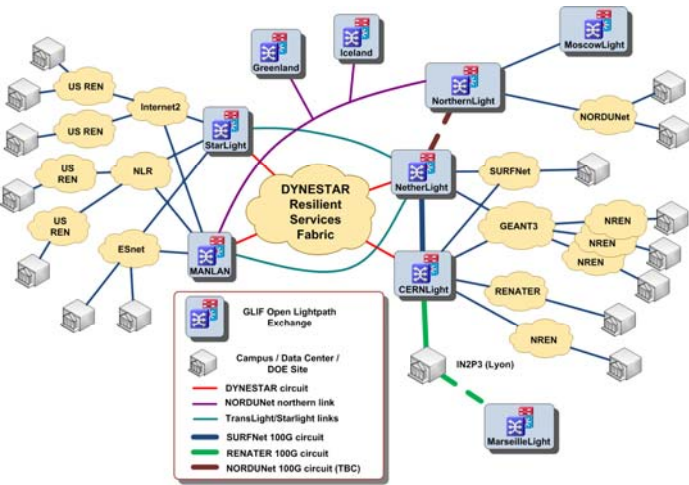


Figure 2: Integration of DYNESTAR and partner networks into the Open Lightpath Exchange infrastructure, nat'l and regional networks

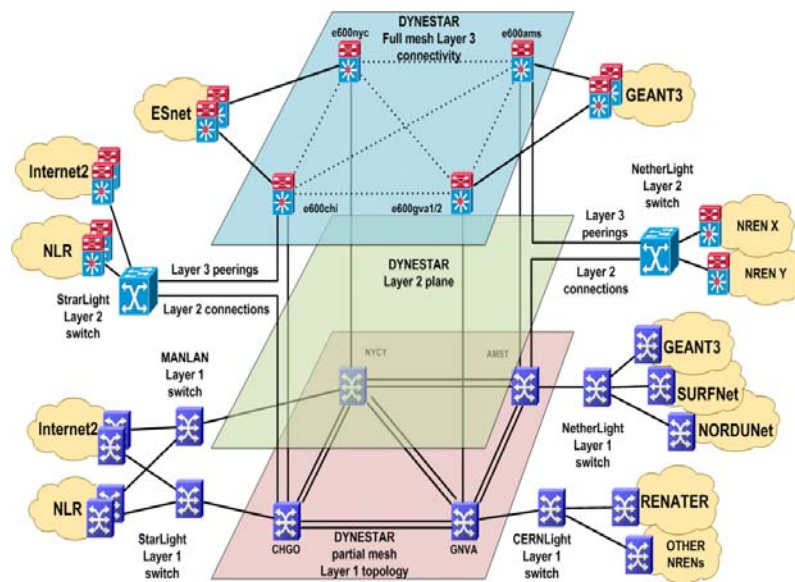


Figure 3: Concept of DYNESTAR/USLHCNet service planes, Layer 1-3. At Layer 2, only the StarLight and Netherlight switches are shown for clarity. Same infrastructure is deployed in all 4 GOLEs.

VCAT/LCAS):

- **protected, guaranteed bandwidth** through mesh-protection
- **unprotected, reduced bandwidth in case of outage** through use of LCAS
- **unprotected, full bandwidth or 0 in case of outage** without LCAS enabled

Using the above described mechanisms, US LHCNet and DYNESTAR are able to provide virtualized services on transatlantic routes. With a partial mesh topology at Layer 1, defined by cost optimization, and a full-mesh virtual topology between the Layer 3 routers, DYNESTAR (like US LHCNet) will provide highly resilient, highly configurable services both at IP and circuit (Layer1 and Layer2) level, as indicated in [Figure 3](#).

Using the state-of-the-art features provided by the current SONET standard [52] as well as mesh protection in the network, we identify the following classes of services, at Layer 1 of the network (unprotected, and mesh-protected services, with or without

We emphasize that these services are already in production, around the clock, in USLHCNet.

At Layer 3, we provide IPv4 and IPv6 services through peerings with our partner networks in Europe and the US. Dedicated capacity for IP peerings will be provided through a set of protected circuits between the DYNESTAR PoPs.

Caltech has constructed and been operating the Ultralight experimental network on US national footprint for the past 5 years. The team has gained expert level knowledge in IPv4, IPv6, multicast routing as well as MPLS. Building on this experience, we will provide and support the corresponding necessary services in DYNESTAR.

The bandwidth of DYNESTAR will begin with two links (NY-GVA and NY-AMS) added to the US LHCNet footprint (as shown in [Figure 4](#)) and will advance annually to meet the expanding needs of the LHC Tier2 sites as well as other needs of the data intensive science community, while taking advantage of ongoing evolution of the cost per unit bandwidth to stay within the project's budget envelope. Based on our long experience with annual RFPs for transatlantic links in US LHCNet, we expect the NSF-funded portion of DYNESTAR to reach a capacity of 100 Gbps by 2014.

D.3. Dynamic Network Allocation: OSCARS and DRAGON

The dynamic allocation of network resources in DYNESTAR is provided by two software packages: (1) On-demand Secure Circuits and Advance Reservation System (OSCARS) [53] and (2) DRAGON [54], as shown in These open-source tools, which are widely deployed and used in production environments today, are combined in a publically available package called the DCN Software Suite (DCNSS) [55] managed by Internet2. OSCARS implements the Inter-domain Controller Protocol (IDCP) [56] developed by DICE [57]. The IDCP is being used as input in standardization efforts at the Open Grid Forum (OGF) [58] and Global Lambda Integrated Facility (GLIF) [59].

At OGF the IDC protocol is being used as input to the Network Service Interface Working Group (NSI-WG) [60]. In GLIF it is being integrated into the GLIF Network Interface (GNI) [11] and heavily leverages topology work from the OGF Network Measurement Working Group (NMWG) [61].

Internet2 has seen approximately 6100 circuits built from January 2008 to July 2009 [63]. 44% of those have been initiated by the ATLAS TeraPaths [64] project and the CMS LambdaStation [65] project, both of which are designed to offload data from local IP networks onto circuit networks when large flows are established. ESnet statistics show that as early as 2007, 77% of the traffic from FermiLab (10 Petabytes that year), including the transfer shown in [Figure 1](#), went over dynamically allocated circuits [66].

The IDCP defines the format of messages that exchange information about local network topology, check resource availability, and coordinate circuit creation between networks. The OSCARS software handles user requests, inter-domain interactions, and the scheduling of resources. When a circuit needs to be created or removed from the network, OSCARS contacts the DRAGON software ([Figure 5](#)). The DRAGON software, which currently supports nearly 20 switch models [67], uses GMPLS [68] to configure switches with VLANs and quality of service (QoS) parameters in the local network. US LHCNet has been one of the first networks to deploy the Internet2 DCN Software Suite (DCNSS) in its production network in early 2008. Since then, the DCN has been used for demos and applications between Ultralight servers at CERN and end sites in the US, such as Caltech which has its own DCNSS instance in production use.

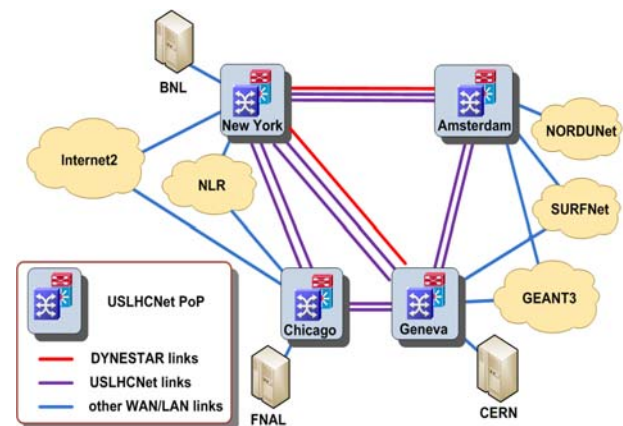


Figure 4: Circuit topology for 2010 operations, adding DYNESTAR links (red) to the USLHCNet footprint.

The US LHCNet deployment of the DCN, shown in [Figure 6](#), includes possible future inter-domain connections with systems in Europe such as AutoBAHN, Phosphorus or DRAC. DYNESTAR will profit directly from this already-operational infrastructure, by binding into the US LHCNet topology. Moreover, by extending the connectivity to our partners' networks, DYNESTAR will provide the bridge between the

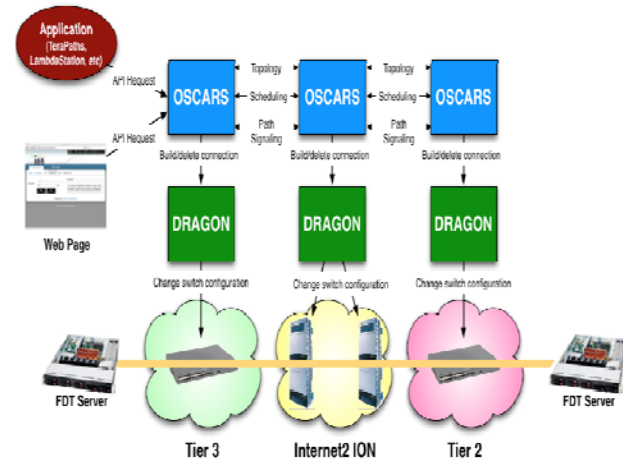


Figure 5: Dynamic circuit architecture based on OSCARS and DRAGON.

developed between Internet2 (DCN), ESnet (OSCARS), DANTE (AutoBAHN) and others, with participation from USLHCNet.

Internet2's DCN, now an operational pilot, is expected to be available as a full service in Summer 2010. In Europe, AutoBAHN is part of the planned service portfolio in GEANT3. In case AutoBAHN is delayed, the best practice (already field-tested) solution is to use static Layer2 extensions from the exchange point at which the DCN arrives, to the end-point at the research facility. Although clearly not scalable, such extensions can be easily pre-configured, and provide a viable short-term alternative.

D.4. perfSONAR Preliminary Results

The perfSONAR project aims to provide infrastructure capable of enabling cross-domain monitoring of networks [69]. Currently most organizations have site-specific methods of performing network monitoring. Multi-domain environments (such as DYNES) require coordination to identify performance bottlenecks or other network issues as they arise. perfSONAR alleviates some of the challenges of these tasks by taking site-specific performance data, and making it available in a standard web services format, using a data format originally defined in the OGF NMWG [61], that is currently being finalized in the OGF Network Measurement and Control (NMC) group [71]. perfSONAR has deployments at 83 sites around the world with the number continuing to grow [72]. Implementations of perfSONAR are made available via toolkits distributed by Internet2 and ESnet, among others. Although perfSONAR has been used

US dynamic circuit networks (DCN and SDN) and their counterparts in Europe. Together with the latter and the US dynamic circuit infrastructure planned by Internet2 together with the regional R&E networks, as in Internet2's proposed NSF MRI-R2 DYNES project, this system will create true end-to-end dynamic circuit connectivity among end-sites on both continents.

DYNESTAR's dynamic circuit services will be based on current standards. The control plane, as developed within the DICE and GLIF frameworks, is already in use by USLHCNet. Interoperability between the DCN software deployed by US LHCNet and the corresponding services in our partner networks [74] has been successfully demonstrated on several occasions, during SC07 and SC08, as well as at the GLIF 2008 workshop.

The common IDC-P protocol is being jointly

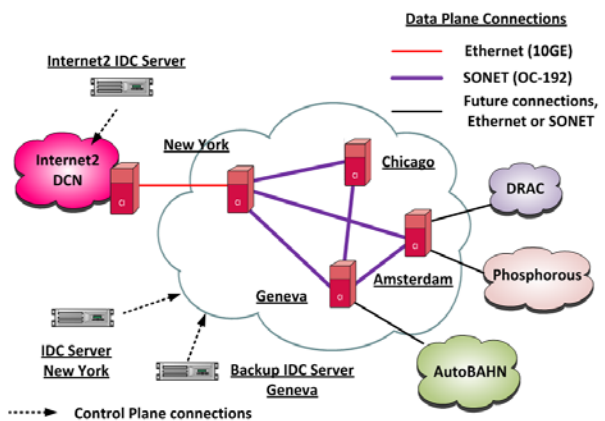


Figure 6: US LHCNet DCN deployment. DYNESTAR will profit directly from this already-operational infrastructure.

mostly to monitor IP networks until now, perfSONAR developers also have worked to support the monitoring of dynamic circuits, and these tools have been used on backbones such as Internet2 and ESnet [73]. This includes weather maps of network status and data utilization on network links.

E. Operations, Monitoring, Quality Assurance and Security Plan

By adding an engineer to the US LHCNet team who will focus on the NSF-funded links, DYNESTAR will profit fully from US LHCNet's cost effectiveness and long-standing experience in transatlantic link management. All DYNESTAR operational, monitoring and security aspects will be fully integrated into the corresponding structures within US LHCNet.

Today, the distributed US LHCNet NOC, with engineers located at CERN and Caltech, manages the LHCOPN Tier 0 – Tier 1 connections between CERN and US Tier 1s (BNL, Fermilab) and several dedicated Tier 1 – Tier 1 connections between the US and European Tier 1s (e.g. Femilab - FZK). The current team includes one lead network engineer, two network engineers and one network engineer/system developer in charge of monitoring system development and support. The NOC has generic contacts (e-mail and phone number) to report faults, forwarded based on the time zone to the appropriate network engineer(s) on duty. During weekends, one engineer is designated on-call and calls are forwarded to his mobile phone.

The USLHCNet NOC monitors the quality, availability, and effectiveness of the services provided. Procedures for fault isolation, timely problem resolution and service assurance are well-defined, and based on 5 years of operational experience in transoceanic networking. USLHCNET NOC has a well documented operational manual, including contingency plan for every possible network failure in its domain. For emergency interventions, USLHCNET has remote hands contracts at each of its POPs. Estimated resolution time is foreseen for every failure scenario. A strict SLA is part of the contracts for transatlantic links for quick problem resolution or re-routing of circuits in case of extended outages.

Service assurance is provided using redundant services and monitoring with automated alarms. OSI Layer1 protection (Mesh protection with VCAT/LCAS) is enabled for high priority network services. There are two dynamic circuits provisioning service instances in the USLHCNet domain. A modular distributed monitoring system provides resiliency in gathering the monitoring data (MonALISA and PerfSONAR), and raises alarms to the NOC engineering team when attention is required.

To track the status of the reported faults, and of planned or emergency maintenance work, the USLHCNet NOC uses a ticketing system. A ticket is opened when the NOC receives a partner NOC's announcements, monitoring system reports, e-mails or calls indicating a problem. The tickets are updated at each iteration during fault isolation and recovery process, and all involved parties (partner NOCs, users) are notified of the progress. After problem resolution, the functionality is checked, the recovery process is documented if needed, and the ticket is closed.

To assure high quality performance monitoring, USLHCNet uses the open source PerfSONAR monitoring system, a toolset widely used by the research community and R&E networks (gridftp, OSG, Internet2, ESnet, GEANT and European NRENs). USLHCNet has deployed one of the first TL1 based PerfSONAR monitoring services (developed by Internet2) that monitors the status of the virtual circuit connections in addition to the physical link status. The collected information is shared with authorized partners like DANTE's E2EMON monitoring system for monitoring of the LHCOPN circuits.

E.1. Fault isolation, Timely problem resolution, Service assurance

Service assurance in DYNESTAR, based on USLHCNet's experience, will be guided by several field-proven concepts: (1) high availability by design, (2) trouble management and fault isolation, (3) performance assurance and monitoring, (3) secured systems and operations, (4) contingency and service recovery guidelines. These assurance guidelines are detailed in the operations manual of USLHCNet.

In addition to the application of state of the art high-throughput methods and tools, DYNESTAR has been designed to provide a high performance network with 99.9+% availability, through the use of multiple links across the Atlantic, network equipment which provides robust fallback at the optical layer in case of link failure, and automatic re-direction of network traffic using redundant network equipment at each of the DYNESTAR points of presence. This target represents an achievable network availability level, based

on 5 years of operations, monitoring and management experience with multiple OC-192 and 10 GE circuits in US LHCNet, as well as the use of resilient virtual circuits on a geographically diverse network footprint with automatic fall back and full agent-based monitoring at Layers 1-3 since 2007 [77].

E.1.1. Fault Isolation

The fault isolation procedure consists of identifying the cause of the outages based on the monitoring system reports, equipment logs, partner NOC's reports and tickets as well as interaction with the carriers' operation centres. Faults detected by the agents of the monitoring system are isolated and recovered. The fault handling procedure contains the following major steps:

- Determine network stack layer at which the problem appears (IP, switched, physical). This information is usually obtained from the type of problem, monitoring system and equipment logs.
- If the fault is believed to be in the carrier domain (link failure), a trouble ticket is opened with the corresponding carrier's NOC through web portal or phone call.
- If the fault is in DYNESTAR's hardware, the faulty module is identified and replaced. USLHCNet has NBD maintenance contracts with its hardware vendors.
- If the fault is related to routing, it is resolved internally (if originating in our domain) or together with partner NOCs in case of external routing issues.
- If the fault presents a major network incident than the contingency plan is followed. A detailed contingency plan is prepared for USLHCNet operations, and will be applied to DYNESTAR.
- After the fault is repaired, the system is tested. A monitoring phase concludes the recovery. The fault and the recovery process are documented for reference and to enhance the problem resolution in the future.

The US LHCNet operations manual contains detailed procedures relating to problem and change management.

E.1.2. Performance

In order to provide high quality services the following performance metrics were identified for DYNESTAR: (1) System and circuit uptime, (2) Mean time between failures/Mean time to recover (MBTF/MTTR), (3) Observed network throughput, (4) Software and hardware failure rate (4) Service response time (5) Call blocking statistics – successful reservation and instantiation statistics, and (6) Failure rate – the number of dropped connections. Based on USLHCNet's experience, while transatlantic circuit availability is around 97-98%, due to path diversity, and the use of mesh protection for primary services, the resulting service uptime is expected to be above 99.9%.

E.2. Security Plan

DYNESTAR project will continue to follow the state of the art security practices in use by USLHCNet, and it will take in consideration aspects including:

1. **Remote Access Control** is addressed using Access Control Lists (ACLs) and Management VLANs. Remote access to network equipment is allowed through ACLs. The same is true for SNMP and TL1 access, where read-only access is allowed to the management/ monitoring stations.
2. **System Update** – Optical, switching and routing network equipment is kept up-to-date based on the recommendations by vendor, and only when a reliable version is available. Software bugs are actively reported to the vendors. Network security advisories e.g. CERT, are actively followed to avoid operational failures and system break-in attempts.
3. **Physical access** – The collocation areas are actively monitored by the providers (StarLight, NYSERNET, SARA and CERN). Activity inside the cage is recorded with strict check-in and check-out procedures. Only a limited number of network engineers has physical access to the network equipment. The collocation facilities provide remote hands either from their own staff or authorized sub-contractors, and any non-USLHCNet personnel (e.g. technicians dispatched by hardware providers) requesting access to our equipment has to be authorized in advance by USLHCNet operations team, and registered with the collocation facility.

4. **Unauthorized Network Functions** – Unused ports and modules are shut down to avoid physical break-in-attempts.
5. **Failure Management** – a link or network equipment failure triggers DYNESTAR's contingency plan. The structure of the contingency plan is described in the following paragraph.

E.3. Contingency Plan

DYNESTAR's contingency plan addresses all the possible failure scenarios and the recovery processes for them. It contains the following failure scenarios, together with estimated recovery time:

1. **Link failure recovery** contains procedures for recovery from extended transatlantic link failure. Circuits and services are rerouted to DYNESTAR's remaining links during the outage. According to SLA (Service Level Agreement) the service provider might be asked to provide alternate path.
2. **Service failure recovery** provides procedures in case the DYNASTAR's main service, the circuit provisioning service, is failing. This recovery process is based on the service resiliency in the circuit provisioning servers.
3. **Equipment failure recovery** procedures for bypassing and replacing faulty equipments.
4. **PoP failure recovery** describes the procedures for reallocating network services of the failed PoP to DYNESTAR's alternate PoPs.

DYNESTAR's contingency plan is based on USLHCNet's, which is a well developed, regularly updated backup plan, tested and proven effective during past failure scenarios.

E.4. Control Plane Integrity

DYNESTAR's control plane consists of management stations (including IDCs) and their control connections with the network devices as well as each other. For circuit services, the control plane is separated from the data plane, as it uses IP connectivity between the nodes. Resiliency in the IP routed connections together with Layer 1 protection makes the control plane highly robust. Incidents in the control plane will not affect the data plane, and incidents in the data plane will not affect the control plane. The control plane respects security best practices: physical security; remote access using Virtual Private Network (VPN) services; encrypted information exchange; up to date hardware and software. In case of a security incident in the control plane, the response plan will go through the following steps:

1. **Identification of the source of the incident** – Security incidents and threat risks are identified using proactive monitoring, users' feedback, and periodical security checks.
2. **Isolation of the source of the incident** – the compromised part of control plane need to be taken out from production system
3. **Recovery process** – The security threat is corrected. Affected components are tested and put back in production
4. **Documentation** – Each security incident is well documented in DYNESTAR's knowledge base
5. **Elaboration of a prevention plan** – Methodologies are defined and tested to avoid the recurrence of the recovered security incident.

F. Usage Policy of International Links

The international links making part of this proposal will interconnect through lightpath exchanges major R&E networks on both sides of the Atlantic and beyond. Particular consideration is given to connections to LHC Tier1, Tier2 and Tier3 sites, as they are expected to be the main users of the services at least in the initial stages.

The Acceptable Use Policy (AUP) will encompass all data intensive research fields requiring high throughput data transfers, such as eVLBI, LIGO [83], nuclear and plasma physics, bioinformatics, etc.

Utilization metrics will be collected on site and virtual organization basis for circuit oriented connections by means of AA mechanisms incorporated in the control plane software. Each request is authenticated and authorized before accepting the circuit reservation. For IP connections, peering policies will enforce rightful utilization, while netflow/sflow metrics will be collected for evaluation purposes.

G. Budget proposal

The DYNESTAR budget takes large advantage of cost sharing with the existing US LHCNet project funded by DOE. In particular colocation, maintenance and especially manpower costs are reduced, and represent a substantial share of the total. The transoceanic OC-192 circuits will connect to new ports on USLHCNet's CIENA optical multiplexers.

Part of the infrastructure costs are contributed by our partners. CERN contributes directly by acquiring the necessary CIENA hardware in Geneva, as well as bearing the colocation costs there. SURFnet contributes by providing capacity on the Geneva-Amsterdam 100Gbps link and its associated share of hardware resources, and the manpower to operate it.

The annual cost breakdown is shown in [Table 2](#). The total cost to NSF, summed up over the grant duration of 5 years, amount to **\$ 4,522 k**, equivalent to an average of **\$ 905 k per year**. The optical and switch-router Equipment costs include port costs on switching and routing equipment, and also \$ 50k for monitoring and DCN servers' installation in the first year. The costs for 2010 and 2011 are based on current pricing obtained from USLHCNet's RFP evaluation (Spring 2009), its vendor pricing tables, and its effective operational costs. Costs in 2012-2014 include a projected evolution of -15% per year on the cost per 10 Gbps of bandwidth of transatlantic links, based on recent and long-term US LHCNet experience, and recent market studies [78].

On the road to 100Gbps technology in production use by or before 2014, USLHCNet's roadmap foresees an upgrade of the optical switching equipment (CIENA CoreDirectors) to the next generation platform in 2010/2011. As a fair share contribution from DYNESTAR, we include the upgrade of the Force10 switch-routing (Layer 2 and 3) equipment, from the current "Terascale" E600s to the corresponding latest "Exascale" model. The total estimated upgrade costs of \$ 400 k are included in the Optical and Switch-Router Costs line in the above table, split between 2010 (\$ 200k) and 2011 (\$ 200k).

Year	2010	2011	2012	2013	2014
Bandwidth (Gbps)	20	40	60	80	100
Transatlantic Circuit Lease Cost (k\$)	230	390	500	570	600
Optical and Switch-Router Equipment (k\$)	370	335	160	150	160
Maintenance and colocation (k\$)	28	42	59	76	32
Manpower (1.25 FTE, k\$, incl. travel)	152	158	164	170	176
DYNESTAR Cost to NSF (k\$)	780	925	883	966	968

Table 2: DYNESTAR annual cost breakdown.

H. Management plan

H.1. Management and Organization

The management team consists of the DYNESTAR PI (H. Newman) and USLHCNet co-manager (D. Foster, CERN). The management team and senior personnel has vast experience in managing large scale international networks as well as collaborative research and education projects on a global scale. The management team, in consultation with operations management (see [Figure 7](#)) is responsible for the roadmap, strategic plan, external collaborations, and relationships with vendors including an annual RFP for circuits to optimize costs; monitors the project milestones and deliverables; and provides guidance to DYNESTAR team members related to both the strategic plan, and daily operations where needed to ensure that the project milestones are met.

H.2. Interaction with the DICE Group and Other External Groups

The management team collaborates with the DICE policy group on setting the directions regarding all policy issues. H. Newman, D. Foster and A. Barczyk are members of the DICE policy group since 2008. The policy group meets regularly twice a year in person, with regular conference call meetings in between.

The DICE operations group coordinates all aspects regarding inter-operation between its member networks. Two DYNESTAR/USLHCNet members (Artur Barczyk and Azher Mughal) are active members of the DICE operations group, and participate in the meetings. Operations meetings are organized as face-to-face meetings twice a year, and every 2-3 months via conference calls.

The entire project team will participate in the DICE and GLIF consortia and work with the OGF to ensure that the services developed and deployed conform to emerging standards for dynamic circuits and network monitoring. Together with GLIF, US LHCNet, Internet2, NLR, SURFnet, NORDUnet, and other European NRENs and GEANT3, the DYNESTAR team will work on evolving the architecture, operational methods and services of the global fabric of Open Lightpath Exchanges.

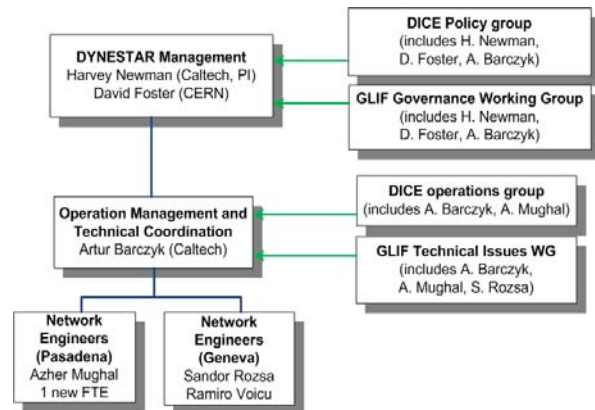


Figure 7: DYNESTAR management structure.

H.3. Project Schedule and Milestones

The general project schedule foresees activation of IPv4 and IPv6 services already within the first month following the award. Dynamic circuit services are planned for deployment and operation within the first year of award. DYNESTAR initially plans to provide 50% of the total capacity for IP transit peerings supporting general purpose networking (including supporting flows to and from physics groups at LHC Tier3 sites), and 50% for dynamic circuit services (including transatlantic flows among the LHC Tier1s and Tier2s as well as other large data flows). Based on round the clock in-depth monitoring (as currently in operation, plus new PerfSONAR services to be developed in partner projects) the balance between IP transit and circuit-oriented bandwidth will be adjusted as needed to best meet the needs of the community.

H.3.1. Phase 1: Design, Planning, Tender Procedure and First Operations (4 months)

Phase 1, prior to the delivery of the first new DYNESTAR circuits, will consist of design and planning, together with our partner networks, of the services to be delivered. A Request For Proposals (RFP) will be prepared, based on the successful USLHCNet RFPs [83][84], and issued within one month following the award, with a 3 month deadline. While awaiting the RFP replies, DYNESTAR will prepare the services to be delivered. Peering agreements, and any mutual backup arrangements between DYNESTAR and its partner networks will be prepared and activated. Hardware for PerfSONAR monitoring will be installed, the services deployed, and integrated with the partner networks' monitoring infrastructure.

In this phase, building on the cost and operations sharing with US LHCNet, we will use the bandwidth currently deployed for general purpose networking, to immediately activate IPv4 and IPv6 transit peerings with our partner networks in the US and Europe, providing routed connectivity services nearly from day one. Any Layer 3 peerings not already present will be enacted, tested and tuned, and the IP connections will be integrated into the common monitoring infrastructure based on PerfSONAR.

H.3.2. Phase 2: Advanced Services Deployment (6 months)

Phase 2 includes hardware deployment to support the additional circuits, deployment and test of new circuits (expected to be available 3 months after the announcement of the RFP results), and an increase in the bandwidth for routed as soon as the new circuits come into operation. Layer 1 and Layer 2 connections to exchange points and partner networks will be installed, and monitoring will be immediately available through preparation during the prior phase. The acquired links will be integrated into the dynamic circuit provisioning system, and tested in collaboration with US and European partners. The DANTE AutoBAHN project, which is expected to be available as a Service Activity by summer 2010, will be integrated with the DYNESTAR DCN in this phase.

H.3.3. Phase 3: First Year of Operation (October 2010 – September 2011)

By October 2010 the DYNESTAR network will become fully operational, and provide services between partner networks across the Atlantic, with an initial capacity of 20 Gbps. Development and refinement of

operational procedures and methods and integration of services among the partner networks and organizations mentioned above will continue throughout this phase.

H.3.4. Phases 4-6: Annual Upgrades

The upgrade plan, which will be made part of the initial RFP, includes annual transatlantic capacity increases by 20Gbps, reaching 100Gbps by 2014. The corresponding ports required in the CIENA multiplexers and Force10 routers are included in the budget plan presented in Section G (and in detail in the Budget Justification section). Given the resilient circuit-mesh already in place, the deployment of new links will be fully transparent to the partner networks and end-users, as has been the case in USLHCNet.

The first 40Gbps wavelength service is offered by Hibernia Atlantic since August 13, 2009. By 2013/2014 we therefore foresee first transatlantic 100Gbps circuits to be commercially available. The upgrade of the existing switching and routing hardware to next-generation platforms is planned to occur in late 2010 for the CIENA multiplexers, and in 2010/2011 for the Force10 switch-routers. While the chassis upgrade cannot be made completely transparent, we will minimize the impact on services by staging the installation of each PoP in time, as demonstrated in similar upgrades in US LHCNet. Due to the resiliency in the fabric, services are expected to be only marginally reduced during the installation.

Refinement of operational procedures and methods and integration of services among the partner organizations mentioned above will continue throughout these phases. The development of monitoring services and high throughput applications for efficient use of 40 and 100 Gbps links will take place in partner projects, and will be periodically integrated into DYNESTAR following pre-production testing.

H.3.5. CIENA Multiplexer and Force10 Switch-Router Upgrades

The CIENA CoreDirector multiplexers currently deployed in US LHCNet will be upgraded to the next generation platforms in 2010. In order not to disturb the USLHCNet operation during the LHC run that year, the installation and test of the new devices will start in mid-2010, with the CoreDirectors remaining in-service through the end of the LHC run (November 2010).

The new devices are operationally compatible with the existing CoreDirectors, which will allow us to switch services in a nearly-transparent manner. Starting from November 2010, after the end of the LHC run, US LHCNet will initially migrate one link between one pair of PoPs to the new devices. A test phase of 2 months is foreseen before more circuits will be migrated, one link at a time. Since DYNESTAR/USLHCNet operates two PoPs on each continent, services will be guaranteed through the other, unaffected devices.

The Force10 routers currently used in USLHCNet are limited to 10Gbps per port, and will be upgraded in order to provide higher-bandwidth services in the future. Here again, we will upgrade one link at a time. In contrast to the CIENA upgrade, we foresee a fork-lift upgrade, i.e. the new chassis will be put in place of the old one. This will result in a short (1 day) interruption of IP services through the PoP in question, however overall the services will be maintained by way of the remaining three locations. We plan to upgrade two devices (Geneva and Chicago) in 2010, and two in 2011 (New York and Amsterdam).

I. Results of Prior NSF Support

Harvey Newman: (PHY-0427110: UltraLight – An Ultrascale Information System for Data Intensive Research, \$ 1.97M; 9/15/04 - 8/31/09) PI, developed a four-continent network testbed and facility, including dynamic circuits and state of the art high throughput transfer methods and tools. (PHY-0622423: Physics Lambda Network System, \$598,285; 8/1/07 - 7/31/10) PI, developed high throughput storage, monitoring and control systems with dynamic circuits for globally distributed data using the UltraLight testbed. (ANI-0230967: Multi-Gbps TCP Project, \$468,240; 10/1/2002 - 9/30/05) Co-PI, developed and deployed high throughput TCP-based methods for large-scale long-distance data transfers. (ANI-0230937: Next Generation Collaborative System Across Advanced Networks, \$750,000; 10/1/02 - 9/30/06) PI, developed a global autonomous collaborative system for the LHC and other major programs. (ACI-0086044: Grid Physics Network Subaward \$762,000; 9/1/00-8/31/04) Co-PI, application of virtual data over networks in the CMS experiment. (EAI-0303620: Wide Area Network in a Lab Subaward, \$110,000), integrated state of the art TCP protocols across WANInLab and the UltraLight network testbed.

References

- [1] ATLAS: <http://atlas.web.cern.ch/Atlas/index.html>
- [2] CMS: <http://cms.web.cern.ch/cms/>
- [3] Open Science Grid: <http://www.opensciencegrid.org/>
- [4] IRNC:Exp Federated Experimental Network Resources for International Research, FENRIR, proposal submitted by University of Houston, George Mason University and NORDUnet in response to solicitation NSF 09-564
- [5] Originally, DICE stood for DANTE, Internet2, Canarie, ESnet. Caltech/USLHCNet joined in 2008
- [6] Virtual Concatenation protocol. See http://en.wikipedia.org/wiki/Virtual_concatenation and http://www.lightreading.com/document.asp?doc_id=30194&page_number=5
- [7] Link Capacity Adjustment Scheme specified in [ITU-T G.7042](#). See for example <http://en.wikipedia.org/wiki/LCAS>
- [8] Fast Data Transfer: <http://monalisa.cern.ch/FDT/>
- [9] FDT and dCache Integration: http://www.ultraviolet.org/FDT_Hadoop/FDTdCache_Status122008.doc
- [10] FDT and Hadoop Integration: http://www.ultraviolet.org/FDT_Hadoop/hdfs-fdt-writeup.doc
- [11] GLIF Network Interface (GNI): http://wiki.glif.is/index.php/GNI_API_Working_Group
- [12] DANTE manages the pan-European Research and Education network GEANT3 <http://www.dante.net/>
- [13] Dutch National Research and Education Network, <http://www.surfnet.nl>
- [14] Joint collaboration between the Nordic NRENs (in Denmark, Sweden, Norway, Iceland and Finland), <http://www.nordu.net>
- [15] French National Research and Education Network, <http://www.renater.fr>
- [16] NSF Award # 0943314, “The Taj: A New Model for Global Federated Network Infrastructure for Science and Education”, <http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0943314>
- [17] We note that the TransLight/StarLight team has committed to work with our DYNESTAR team on the planned first tests of transatlantic 100G links (between New York and Amsterdam for example)
- [18] Large Hadron Collider: <http://lhcb.web.cern.ch/lhcb/>
- [19] Bringing the benefits of the DYNESTAR system, services, and tools to the communities in the many fields of data intensive science mentioned in this proposal will be facilitated by our planned work with the StarLight/TransLight team, through their wide-ranging work with NSF-supported science groups in many disciplines.
- [20] NP Science Network Requirements. ESnet Report of the Nuclear Physics Network Requirements Workshop. May 2008 <http://www.es.net/pub/esnet-doc/NP-Net-Req-Workshop-2008-Final-Report.pdf>
- [21] Earth System Grid: <http://www.earthsystemgrid.org/>
- [22] *FES Science Network Requirements*. ESnet Report of the Fusion Energy Sciences Network Requirements Workshop. March 2008 <http://www.es.net/pub/esnet-doc/FES-Net-Req-Workshop-2008-Final-Report.pdf>
- [23] ITER: <http://www.iter.org/>
- [24] *BES Science Network Requirements*. ESnet Report of the Basic Energy Sciences. June 2007 <http://www.es.net/pub/esnet-doc/BES-Net-Req-Workshop-2007-Final-Report.pdf>

- [25] Ravot, S and H. Newman, J. Bunn, I. Legrand, F. van Lingen, " Meeting the Challenges of High-Energy Physics", CENIC Interact Winter 2005, Partnership award extract (see also: <http://www.ultralight.org>)
- [26] Newman, H and J. Bunn, I. Legrand, S. Low, D. Nae, S. Ravot, C. Steenberg, X. Su, M. Thomas, F. van Lingen, Y. Xia, R. Cavanaugh, S. McKee. "[The UltraLight project: The Network as an Integrated and Managed Resource for Data Intensive Science](#)", in [Computing In Science and Engineering](#), Issue on grid computing, 2005 (see also: <http://www.ultralight.org>)
- [27] Data Intensive Sciences University Network: <http://www.disun.org/>
- [28] Newman, Harvey. *ICFA SCIC Report: Networking For High Energy Physics*. International Committee on Future Accelerators. February 2009
http://monalisa.caltech.edu:8080/Slides/Public/SCICReports2009Final/ICFASCIReport2009_020909.pdf
- [29] Fermi Telescope: <http://fermi.gsfc.nasa.gov/>
- [30] Chandra Observatory: <http://chandra.harvard.edu/>
- [31] NuStar: <http://www.nustar.caltech.edu/>
- [32] GALEX Satellite: <http://www.galex.caltech.edu/>
- [33] Joint Dark Energy Mission: <http://jdem.gsfc.nasa.gov/>
- [34] Spitzer Space Telescope: <http://www.spitzer.caltech.edu/>
- [35] Herschel Satellite: <http://sci.esa.int/herschel/>
- [36] WISE Satellite: <http://wise.ssl.berkeley.edu/>
- [37] UKIDSS Survey: <http://www.ukidss.org/>
- [38] VISTA Telescope: <http://www.vista.ac.uk/>
- [39] WMAP Satellite: <http://map.gsfc.nasa.gov/>
- [40] PLANCK: <http://www.esa.int/SPECIALS/Planck/>
- [41] Expanded VLA Project: <http://www.aoc.nrao.edu/evla/>
- [42] LOFAR: <http://www.lofar.org/>
- [43] Automated Bandwidth Allocation across Heterogeneous Networks: <http://www.geant2.net/server/show/ConWebDoc.2544>
- [44] Square Kilometer Array: <http://www.skatelescope.org/>
- [45] Catalina Real-Time Transient Survey: <http://crts.caltech.edu/>
- [46] Palomar Transient Factory: <http://www.astro.caltech.edu/ptf/>
- [47] Pan-STARRS: <http://pan-starrs.ifa.hawaii.edu/public/>
- [48] SkyMapper: <http://www.mso.anu.edu.au/skymapper/>
- [49] Large Synoptic Survey Telescope: <http://www.lsst.org/lsst>
- [50] AutoBAHN Project: <http://www.geant2.net/autobahn>
- [51] Ciena: <http://ciena.com>
- [52] Also part of the next generation ITU OTN standard (ITU-T Recommendation G.7042).
- [53] On-demand Secure Circuits and Advance Reservation System: <http://www.es.net/OSCARS/>
- [54] DRAGON – Dynamic Resource Allocation via GMPLS Optical Networks: <https://dragon.maxgigapop.net>
- [55] Dynamic Circuit Network Software Suite: <https://wiki.internet2.edu/confluence/display/DCNSS/>

- [56] Inter-domain Controller (IDC) Protocol: <http://www.controlplane.net>
- [57] DICE: <http://www.geant2.net/server/show/conWebDoc.1308>
- [58] Open Grid Forum: <http://www.ogf.org/>
- [59] Global Lambda Integrated Facility (GLIF): <http://www.glif.is/>
- [60] Open Grid Forum (OGF) Network Service Interface Working Group: http://www.ogf.org/gf/group_info/view.php?group=nsi-wg
- [61] Open Grid Forum (OGF) Network Measurement Working Group (NM-WG): http://www.ogf.org/gf/group_info/view.php?group=nm-wg
- [62] Vollbrecht, John, et al. *Using DCN: A Brief Tutorial [presentation]*. 29 April 2009 <http://www.internet2.edu/presentations/spring09/20090429-DCNTutorial-lake.pdf>
- [63] Internet2 DCN Usage Statistics: <http://ion.net.internet2.edu/dcn-usage/>
- [64] TeraPaths End-to-End QoS Networking Project: <https://www.racf.bnl.gov/terapaths/>
- [65] LambdaStation: <http://www.lambdastation.org/>
- [66] Newman, Harvey. International Committee on Future Accelerators (ICFA) Standing Committee on Interregional Connectivity (SCIC) [presentation]. Presented at February 2009 ICFA meeting http://monalisa.cern.ch:8080/Slides/Public/SlidesReportsforJohn/ICFASCIIPresentation2009_hbnShorterExtract072809.ppt
- [67] DRAGON Supported Switches. Internet2 (Accessed 24 July 2009): <http://wiki.internet2.edu/confluence/display/DCNSS/DRAGON+Supported+Switches>
- [68] Mannie, L ed. *Generalized Multi-Protocol Label Switching (GMPLS) Architecture*. Internet Engineering Task Force (IETF). RFC 3945. October 2004 <http://www.ietf.org/rfc/rfc3945.txt?number=3945>
- [69] perfSONAR. perfSONAR: <http://www.perfsonar.net/>
- [70] perfSONAR Information Service Tree(Accessed 24 July 2009): <http://dc211.internet2.edu/cgi-bin/perfSONAR/tree.cgi>
- [71] Open Grid Forum (OGF) Network Measurement and Control Working Group: http://www.ogf.org/gf/group_info/view.php?group=nmc-wg
- [72] Summer 2009 ESCC/Internet2 JointTechs: <http://events.internet2.edu/2009/jt-indy/index.cfm>
- [73] Robb, Chris. *Evolving Internet2 DCN to Production [presentation]*. JointTechs. 20 July 2009 <http://www.internet2.edu/presentations/jt2009jul/20090720-robb01.pdf>
- [74] DANTE/GEANT: AutoBAHN, SURFnet: DRAC. Also interoperability with the European Phosphorus project has been demonstrated.
- [75] Dynamic Network System, DYNES: proposal submitted by Internet2, Caltech and University of Michigan and Vanderbilt University in response to the NSF MRI-R² solicitation NSF 09-561, proposal number 0958998
- [76] dCache: <http://www-dcache.desy.de/>
- [77] US LHCNet monitoring repository is available at <http://repository.uslhcn.net.org>
US LHCNet Real-time link status: <http://repository.uslhcn.net.org/Panel?page=panel>
US LHCNet link availability statistics: <http://repository.uslhcn.net.org/status/>
US LHCNet Flow based statistics: http://repository.uslhcn.net.org/display?page=MLEFLOWS/ciena_hist_mleflows&dont_cache=true&interval.min=86400000&interval.max=0
- [78] See e.g. Telegeography Global Bandwidth Research study (2009) Executive Summary, available online at <http://www.telegeography.com/products/gb/download/executive-summary.pdf>.

- [79] International Committee on Future Accelerators: <http://www.fnal.gov/directorate/icfa/>
- [80] Hadoop: <http://hadoop.apache.org/>
- [81] SuperComputing 2008: <http://sc08.supercomputing.org/>
- [82] *CalTech SuperComputing 2008 Report*. CalTech. November 2008
<http://supercomputing.caltech.edu/results.html>
- [83] Laser Interferometer Gravitational Wave Observatory: <http://www.ligo.caltech.edu/>
- [84] USLHCNet has been issuing yearly RFPs for its transatlantic circuit infrastructure in order to optimize the costs since 2006.
- [85] *BER Science Network Requirements*. ESnet Report of the Biological and Environmental Research Network Requirements Workshop. July 2007 <http://www.es.net/pub/esnet-doc/BER-Net-Req-Workshop-2007-Final-Report.pdf>
- [86] Bunn, J. and H. Newman. "Data Intensive Grids for High Energy Physics", in *Grid Computing: Making the Global Infrastructure a Reality*. edited by Fran Berman, Geoffrey Fox and Tony Hey, March 2003 by Wiley
- [87] Cotter, Steve. *ESnet4 Update [presentation]*. JointTechs Conference. July 2009
<http://www.internet2.edu/presentations/jt2009jul/20090721-cotter.pdf>
- [88] Johnston, William E, et al. *The Evolution of Research and Education Networks and their Essential Role in Modern Science*. To be published in: *Trends in High Performance & Large Scale Computing*. Gandinetti, Lucio and Gerhard Joubert editors, 2009 <http://www.es.net/pub/esnet-doc/The-Evolution-of-Research-and-Education-Networks-and-their-Essential-Role-in-Modern-Science.v5.pdf>
- [89] *Virtual Bridged Local Area Networks*. IEEE Standard 802.1Q. 2005
<http://standards.ieee.org/getieee802/download/802.1Q-2005.pdf>
- [90] Metzger, Joe and Ndousse, Thomas. *Performance Measurement in a Multi-Domain Infrastructure [presentation]*. JointTechs. 21 July 2009
<http://www.internet2.edu/presentations/jt2009jul/20090721-metzger-ndousse.pdf>