GLIF Architecture Task Force
DRAFT—DRAFT--DRAFT

GREEN PAPER

January 6- 2013

Prepared by:

Bill St. Arnaud bill.st.arnaud@gmail.com
Erik-Jan Bos bos@nordu.net
Inder Monga imonga@es.net

During the 11the Global LambdaGrid Workshop in Rio de Janeiro, Brazil in October 2011, the Architecture Task Force was created under the GLIF Tech Working Group. Subsequent to that meeting a Charter for the task force was drafted on September 12, 2012. At the October 2012 GLIF meeting in Chicago it was a Use Case Analysis document was presented and it was agreed that the next step would be to draft a Green Paper to be submitted to the GLIF Tech a the January 2012 meeting in Hawaii.

The *Green Paper* is a *Consultation* instrument to incorporate researcher's experience, vision and expectations for end to end lightpath connectivity across GLIF infrastructure including campus networks. The intent of the document is to focus on future technical directions and does not address policy issues, accepted use, governance, and/or cost sharing. The document is also not intended to propose detailed technical solutions, but instead identify gaps, opportunities and where possible identify best practices.

Concurrent with this document there are several programs around the world to address the challenges of researcher connectivity across campus to high speed research networks. These include the NSF CC-NIE program and the Terena led GN3 "Campus Best Practices" as well as Internet 2 and GENI are investigating technical solutions to provide end-to-end lightpath support.

In what follows, high-level technical issues are briefly introduced and classified into generic subject areas:

1.0 **Vision of GLIF**
2.0 **Use Cases and Applications**
3.0 **Infrastructure Capabilities**
4.0 **Topics of Discussion and Next Steps**

The GLIF Tech Committee will subsequently process the results of this consultation which will result in a *Final Report* as a *White Paper* to the GLIF community.

1.0 Vision

GLIF, the Global Lambda Integrated Facility, is an international virtual organization that promotes the paradigm of lambda networking. GLIF provides lambdas internationally as an integrated facility to support data-intensive scientific research, and supports middleware development for lambda networking.

GLIF's accomplishment has been its ability to provide lambda networking to a vast, global and diverse user base. Via its constituent National Research and Education Networks (NRENs) on almost all continents and GLIF Optical Lightpath Exchanges (GOLEs) and campus networks it connects researchers and end-users worldwide, thus creating the world's largest and most advanced R&E networking ecosystem for supporting high end data intensive science and collaboration.

The success of GLIF is due to the foresight of the NREN community in recognizing that dedicated lightpaths and lambdas will be critical to the huge data flows that are now part and parcel of data intensive science. It is therefore imperative that the GLIF ecosystem further develop robust multi-domain lightpath services to advanced users as the demands of data intensive science continue to evolve, particularly with the introduction of clouds, mobile applications and commercial science services.

On top of advanced multi-domain dynamic lightpath networking, we are also witnessing the emergence of a multitude of software driven virtual networks generally referred to as service or software oriented architectures. GILF might evolve into a federated infrastructure with the network itself as an entity capable of creating/hosting complex objects made of virtual circuits, routers, switches and nodes, enabling community oriented virtual services. Architecturally, the general adoption of service oriented architectures will allow the integration of services from many different providers creating a vibrant and competitive set of offerings built on top of the network. The GLIF community, working together with NREN initiatives around the world, will be vital in realizing the opportunities that the service oriented evolution will bring.

The world of science and research based networking and computing is rapidly changing and GLIF must adopt to these challenges. While lambda networking will remain the primary objective of GLIF, the demands of any where, any time, and any device will place new requirements on how we will use lambdas to support advanced science in the future. Cloud computing, science as service, commercial data providers, wireless access and large distributed sensor networks, campus out sourcing and off loading are some the factors that must be taken into account in the future design of lambda networks.

In addition to the demands of data intensive science researchers and institutions are also trying to mitigate the energy and environmental impact of high performance computing and networks. Numerous research groups are exploring how to use lightpath and software based networks to build highly distributed and distributed infrastructure in order to locate large energy hungry sensors and computation facilities at low cost, preferably green locations.

GLIF and the participating NRENs will need to continue to expand existing network services and innovate at all layers, particularly as follows:

- Building scalable and discoverable applications that can automatically request end-to-end network services. Federated identity and middleware platforms offering a suite of composable network services to enable such automatic end to end provisioning will be a critical part of the solution mix. The ability to build heterogeneous solutions from campus based SDN, lightpaths or IP networks to GOLEs and across myriad of GLIF enabled infrastructure will be essential.

- The global Internet and its suite of services will still remain the predominant form of networking for most researchers. Dedicated high speed lightpaths, server to server, for the foreseeable future will be used by only the high-end of researchers particularly in the fields of physics, astronomy and climate modelling. High quality access to the internet, both wired and wireless is the most important network service for the majority of researchers and other users. Unfortunately the commercial Internet suffers from serious degradation at major Internet Exchange points (IXs). Traffic engineering by reducing round trip time and multi-hop AS through direct peering is the surest way of improving Internet performance for researchers. Using GLIF lambdas for the exchange of peering routes to create a federated Tier 1 global ISP network will allow GLIF to represent community of global NRENs in peering arrangements with commercial ISPs and distribute scientific and educational content via lambda content networking. As well off loading Wifi, 3G and LTE traffic will provide significantly better performance for users who are off campus.

- Although layer three packet switching (IPv4, IPv6) and its multi-domain control via BGP will remain the cornerstone for the many-to-many connectivity needs, it is expected dedicated high bandwidth, private and discipline specific, bandwidth on demand integrated with applications will become a major trend. LHCONE is an early example of this type of architecture. Other examples include the Greenstar network, the NLR Genomics network and plans for a global SKA network.

.

- The convergence of computing (both cluster and supercomputing), clouds, databases and instruments in a multi-domain hierarchical environment (campus, NREN, GLIF) may have direct implications at network layers. Virtualization within high performance computing and networking is expected to provide a seamless workplace for researchers and campuses, leading to “Science as a Service” allowing a researcher access to their computation and instruments from anywhere, anytime and on any device. Providing researchers with simple tools to configure, manage and debug distributed infrastructure and undertake performance testing will be an essential part of the puzzle.

At the end of the day a GLIF end-to-end architecture is likely to be a heterogeneous mix of network solutions such as MPLS VRFs, SDN networks, lightpaths and

switched lambdas, collectively providing a range of services from global IP peering, big data flows, content routing and integrated wireless networks.

2.0 Use Cases and Applications

In terms of use cases and applications it is important to understand the different types of users, and the applications that will serve their specific needs. Although Lightpath networking may not be directly accessible or relevant to a large number of researchers and users, indirectly it can enhance and improve scientific research through a number of lightpath based applications and middleware. This section first looks at the different types of users and the possible lambda network applications and middleware that will support these needs.

2.1 End-User Profiles

Users can be classified according to their requirements, along with their respected number and degree of technical maturity. The types of users can be broken into three profiles largely based on technical maturity and degree of management/control functionalities delegated to an end-user: *Small and Medium Science Users* (the vast majority of users, requiring IP connectivity), *Big Science Users* (few users, mainly e-Science projects, requiring dedicated circuit high capacity linking) and *Guinea Pig Users* who to test out new technologies and architectures. Based on these the following profiles can be identified:

2.1.1 Small and Medium Science Users

The vast majority of users in the R&E communities are small and medium science users, satisfied with IP connectivity. Even though they only require IP connectivity high performance with low latency and high throughput is critical to this community. Campus network bottlenecks and limited bandwidth connectivity to commercial ISPs can be a major impediment for these users. Common usage, apart from IP connectivity require federated services, including ID management and credential delegation via a federate IdM, access to video-conferencing & collaborative tools etc. As well, small and medium science users are increasingly taking advantage of commercial clouds and Science as a Service (SaaS) providers for their research needs. For large file transfers this community largely relies on IP based file transfer services such as Globus Online rather than using dedicated lambdas. Clearly, following the lead of Big Science users by integrating applications and middleware with bandwidth on demand (BoD) middleware in a transparent fashion may expedite file transfer process, and many such initiatives are in the way. While composable service oriented architecture and BoD lightpath networks may not likely be required on an individual basis for many small and medium science users, aggregating their demands at the campus or backbone lambda level to enable direct connectivity to cloud and commercial providers through dedicated or BoD lightpath services will be important.

2.1.2 Big Science Users

They need the greatest degree of bandwidth, extending to 10 Gig and well beyond dedicated circuit provisioning and most usually are bundled within large campuses or facilities, with high capacity aggregate access to their NREN. Power users share interconnected storage and computing resources (e.g. clusters within a Grid, distributed super-computing centres) and often require dynamic lambda facilities. It is this type of users that initially triggered the demand for lambda networking. Big

Science Users often require international connectivity and federated computing and network resources. They are currently the poster child for GLIF types of services.

The biggest challenge now for lambda networking is how to integrate middleware and applications with bandwidth on demand services or re-routing flows to dedicated or BoD services. As well, multi-domain management plane tools for measuring performance and throughput are critical to this community as even small losses can cause significant throttling.

2.1.3 Guinea Pig Users

An emerging community of Guinea Pig users includes researchers on the Network of the Future, GENI, etc. This category, on top of high capacity connectivity at low protocol layers, may need access to the GLIF infrastructure to serve as a testbed and to support virtualization facilities for emulation experiments e.g. *PlanetLab*, *GENI*, *FEDERICA*. These are advanced users, e.g. University laboratories in networking and distributed computing that are willing to try novel architectures and services. They are not expected to request the support level of production environment, but rather require high level experts support. They might provide useful feedback during service development. Candidate use cases may include beta-testing of shared virtual data repositories and elastic computing services, and evaluation of logical routers (i.e. physically located in NREN PoPs, but providing a virtual slices to end-user campuses). The *Guinea Pig* user group extends worldwide, with intercontinental cooperation for global proofs of concept.

2.2 Lightpath Network Applications

Lightpath Network applications can be broken into two broad categories – *direct lightpath connectivity* to end users and *underlying lightpath connectivity* to support traffic engineering, middleware and IP connectivity where the lambda and lightpath networking is largely hidden from end users. This differentiation is important as the end-to-end architecture requirements, performance management tools, etc may vary quite a different depending on the application.

To date, most lightpath network demonstrations have focused on switched short lived lightpaths (Circuit on Demand) for large data flows between a computer database (or instrument) and an end user's server or border router. Lightpath networking is seen as a more workable substitute to using QoS on IP networks. Most IP routed networks don't have the bandwidth headroom to accommodate large data flows. As well data packet loss from campus mis-configuration and congestion on the network can result in significant throttling of throughput on IP networks especially those with long Round Trip Times (RTTs). Direct lightpath networking is intended to obviate these attendant problems with routed IP networks.

While direct lightpath connectivity will likely remain a predominant application for lambdas within the GLIF environment for some time, it is important that we don't ignore other network paradigms that will also involve lightpath networking, but where direct manipulation of the lightpaths may be hidden from the end user. In this environment lightpaths are often used as part of a traffic engineering toolset by network engineers and eventually applications to segregate different types of traffic

flows. Application performance is enhanced by providing direct, uncongested IP routes. These types of applications may have greater impact for small and medium size science communities who make up the overwhelming portion of the scientific community.

Traffic engineering lightpaths, in theory, can use BoD lightpaths services but route advertisement, convergence and the threat of flapping means that most traffic lightpaths will need to long term and persistent.

Clarification on Bandwidth on Demand (BoD) and Circuit on Demand (CoD)
The Bandwidth on Demand (BoD) approach is to build a long-lived circuit and set its bandwidth to near-zero. This will allow the keep-alive connectivity to live between the applications while it does not use a lot of reserved bandwidth. When the application initiates a bandwidth-intensive workflow, it can ask the network to modify the available bandwidth for the existing circuit to its desired value. This works fine and avoids the route flapping/convergence problem, though it is quite feasible, like any BoD system, the bandwidth required between two points may not be available.

Unlike the above approach, one approach is to only request a circuit when the application needs it. This can be typically longer time to provision, and there is no resource availability guarantee. Additionally, the end-to-end IP issues need to be resolved manually ahead of time in order to make the circuit useful.

The network provider and/or the user can determine what approach works for them the best and leverage either approach to take benefit of a B/CoD system.

The big advantage of a traffic engineering approach to lambda networking, as opposed to direct lambda connectivity to users or applications is that traffic engineering interfaces, aka lightpaths, are much better with existing applications and Operating Systems. Most science applications are built on the Unix stack where interconnection is done through software ports and the routing of traffic to a specific interface, in most cases, is based on default routes mapped through an ARP request of a local LAN. With BoD a host application must not only be able to setup a specialized channel across the LAN as well as across the GLIF infrastructure it must then arrange to insert static routes at both ends of the connection, or arrange for a new ARP request at each end (usually on the same subnet), so that the application can transfer data. Considerable back and forth communication between host and destination, as well as all points in between, is necessary to complete a single transfer. A traffic engineering approach, on the other hand, with static and/or dynamic routes is more consistent with global network routing and yet can achieve much of the same result of a dedicated path between host and destination. But in order to insure routes are properly advertised network route with no flapping, paths and interfaces must have persistence considerably longer than the time required for the actual data transfer. As well most server and router interfaces use “keep alive” messages in the forwarding plane as a way to signal link failure to the control plane. Dynamic optical paths therefore need to insure that they don’t inadvertently create frequent link failures. With OpenFlow,

where the control plane is separated from the data plane, it is conceivable that persistence of the forwarding path may not be required.

This section attempts to explore the entire range of possible lambda network applications for both direct and underlying lambdas.

2.2.1 Global Tier 1 Peering Applications

As mentioned previously the overwhelming need for most researchers is high quality access to the global Internet. While lambda networks traditionally have not been seen as a vehicle to address IP performance problems, it has been increasingly recognized that lambdas can significantly improve Internet throughput and latency by reducing round trip times (RTT) and the number of Autonomous System (AS) hops.

Given the low and dropping cost of Internet transit prices some larger NRENs see little value in establishing no cost peering at various IXs around the world. They believe that the cost of circuit to a peering point can outweigh the advantages of purchasing IP transit locally. However, NORDUnet, SURFnet, AARNet and other networks who have replaced transit connectivity with direct IP peering have noted an immediate jump in traffic of approximately 25%. They have attributed this change to fact that the direct peering via lambda reduces RTT and the number of AS hops which results in many applications being able to push or pull more traffic through the direct peering connection, rather than sitting idle because of longer RTTs via the commercial Internet. This is especially important for sites that are a long distance away as the RTT can severely slow down throughput, especially if there is any packet loss anywhere along the route. The TCP congestion control mechanism throttles back data volumes in the event of any packet loss. Using directly connected lambdas minimizes the risk of any packet loss and can significantly reduces AS hop count.

As well several NRENs offer a “content routing service” to enable smaller regional networks and institutions to connect to major content providers, Content Distribution Networks (CDNs) and Tier 2 ISPs. This is especially beneficial to smaller and more remote networks or institutions who don’t have access to low cost Tier 1 transit providers. Expanding the extent and reach of the “content routing service” globally will further help reduce the high Internet costs these organizations face.

Most of the challenges of building a federated global Tier 1 ISP network will undoubtedly be in the business relationships between participating NRENs, particularly in how the costs of transit traffic will be attributed to each NREN. As well, some major Tier 1 ISPs require contractual agreements in order to peer with them and it will be necessary to decide whether GLIF itself, or some other identity, can speak on behalf of the participating NRENs in terms of entering into peering agreements with these major Tier 1 providers.

Nevertheless, despite the business and political challenges a number of network technical issues need to addressed as well. The most complex architectural decision will be whether participating NRENs will be able to peer directly at major IXPs or have their routes advertised by the local NREN or perhaps GLIF itself.

Most organizations that peer at an Internet eXchange Point (IXP) prefer to keep the number of peers to a minimum, as there is a cost associated with the number of interfaces and size and power of the route server or router. Usually connecting networks make a trade off between traffic volumes and the number of connected peers. Most IXPs offer a mix of connectivity via a shared network and direct point to point connections, as such a participating network needs to maintain as a minimum an Ethernet interface on a device and some sort of routing engine. In theory an Ethernet interface with a direct circuit connection for remote peering is also possible. Virtual routers are also a possibility.

At the end of the day it is likely there will be a mixture of solutions depending on traffic volumes, peering relationships, cost of equipment for each participating NREN. These solutions can be summarized as follows:

(a) Direct peering with NREN owned router and circuit;

(b) Shared peering with virtual router/interface and virtual circuit, with local NREN or GLIF owning and operating underlying circuit and router at the IXP; and

(c) NREN peering behind local NREN or GLIF AS at the IXP.

In the examples (a) and (b) above most likely lightpaths such as MPLS VRF VPNs, VLANs, etc will need to be setup from the NREN virtual or physical router back to the participating NREN core router. As the number of peers and interconnected bandwidth at IXPs constantly changes NREN engineers will need the capability to increase (or decrease) the number of virtual router instances and/or the number of dedicated or child lightpaths from their presence at the IXP. Clearly this is an ideal application for Software Defined Network (SDN) technology built on lambdas.

It should be also noted that many larger campuses also do their own direct peering at IXPs and are expected to be similarly interested a SDN peering service.

2.2.2 R&E Content Distribution Network

It has been long recognized that the vast majority of data on the Internet is relatively static and the same data is often retrieved multiple times. As such Content Distribution Networks (CDNs) have been a major integral part of the global Internet infrastructure.

Although CDNs are often associated with the distribution of multimedia content such as movies, music and so on, they are also used to distribute courseware and research material. The LHCONE network is a specialized example of a CDN network for a specific dedicated application – the distribution of data from the CERN accelerator.

Popular databases from the Human Genome project, virtual astronomy, virtual anatomy, etc are often distributed via these networks. But since many CDN networks charge for distribution of content the vast majority of educational and research material is not distributed on these networks. To date CDN facilities have not been critical for R&E networks because of the ample bandwidth, but as more and more

users are accessing the R&E networks through wireless connection, or through the commercial Internet (i.e. for Citizen Science or courseware applications), performance and throughput can be significantly enhanced with a CDN network. It is not only receiving content and data that CDN networks are important, but also for delivering content from universities and research institutes to the global Internet community.

CDNs significantly improve user perceived performance and throughput because the content is stored locally, rather than on a distant server. The transmission rate of data over the Internet is directly related to the Round Trip Time (RTT) and any packet loss on route. The shorter the RTT and reduced packet loss can dramatically increase throughput and performance response for users.

It is interesting to note that commercial CDNs such as Google, Akamai, Limelight, etc have the largest deployment of wide area optical networks in the world far exceeding that of commercial carriers. They are also the first networks to deploy SDN to manage their network and content distribution infrastructure.

The IETF has started up a working group called CDNi which is looking at developing standards for interconnection and distribution of CDN networks globally. It is generally desirable that a given content item can be delivered to an end user regardless of that end user's location or attachment network. However, a given CDN in charge of delivering a given content may not have a footprint that expands close enough to the end user's current location or attachment network, or may not have the necessary resources, to realize the user experience and cost benefit that a more distributed CDN infrastructure would allow. This is the motivation for the IETF initiative for interconnecting standalone CDNs so that their collective CDN footprint and resources can be leveraged for the end-to-end delivery of content from Content Service Providers (CSPs) to end users. As an example, a CSP could contract with an "authoritative" CDN Provider for the delivery of content and that authoritative CDN provider could contract with one or more downstream CDN provider(s) to distribute and deliver some or all of the content on behalf of the authoritative CDN Provider.

Building a federated CDN network, for the delivery of research and educational content, compliant with the IETF CDNi standards for the R&E community is an excellent application for a lightpath SDN architecture similar to that recently deployed by Google on their optical network.

A R&E CDN network might possibly involve use of both dynamically switched lightpaths and long term persistent flows. For example a researcher may need to update a database distributed by CDN on an infrequent basis and as such a short lived dynamic lightpath may be all that is required. On the other hand an institution may want to be part of the CDNi distribution network and therefore would require a complex mesh of VPNs or equivalent to interconnect to various commercial, as well R&E CDN services.

2.2.3 Cloud Applications

Access to commercial clouds and science service providers is becoming increasingly important for many researchers, especially those outside of the physical sciences.

"Science as a Service" is a new and increasing popular resource from many university researchers who use commercial providers for a variety for specialized analytical and processing tasks. Many companies in Europe and the US, especially in the fields of genomics and humanities are being established to provide a variety of research services.

Establishing lightpaths and IP services to these companies is becoming increasingly important issues for many NRENs and is challenging many preconceived notions of Acceptable Use Policies (AUP) and membership in the NREN.

It is expected that NRENs will establish large optical interconnections to commercial cloud or SaaS providers. In most cases traffic will be aggregated to end users with appropriate billing arrangements made between the NREN and institutions. In some situations where large traffic volumes are expected between a cloud or SaaS provider and a specific user or institution, a smaller child or ancillary lightpath might prove useful.

In the case of private clouds deployed at universities and research institutions which interconnect physical servers the movement of virtual machines (VMs) between sites for restoral or failover processes will be required. Mega pipe networks will also be required for interconnecting storage tiers. These configurations will be closely related to CDN and Global peering applications as discussed previously.

There are a number of possible large data flow scenarios for cloud applications:

(a) Raw data from sensors and instruments flows directly to a commercial cloud where all computation, processing and storage is done in the cloud ;

(b) Raw data from sensors and instruments flows to campus computational resources and a subset of processed is stored in the cloud; and

(c) Cloud is used as a storage medium while processing is done on campus computational facilities.

The configuration and management of lightpaths will vary with these different scenarios.

Even though we state this as a Cloud Application, the real differentiator here is the ability for the researcher to request and acquire computing cycles on demand, the ability to move their data to the compute and get the results back. Many of these services and models are being investigated by the traditional Supercomputer facilities as well, who so far have been offering a batch-processing model with scheduling for large-scale computation. With the on-demand supercomputing or cluster model, the requirements of lightpaths are similar to that of the cloud.

A real world example of the requirement to map lightpaths to cloud services is Amazon Web Services (AWS) new VPN service to interface to external networks. Carriers can now interconnect to AWS with large 10 Gbps pipes and partition those links into multiple VPNs dedicated to different customers in order for them to connect

to the Amazon Cloud. Clearly this will likely be an important service for GLIF and the NRENs to offer to their research clients.

2.2.4 Big Data Applications

The quintessential application for *direct lambda connectivity* is to support big data flows between instruments, computer databases, computational facilities and so forth. The poster child for big data is the distribution of data from CERN to the various Tier 1 data centers around the world. Another growing example is the transmission of large climate modelling data sets.

Within certain management domains such as ESnet, switched lighpaths for Big Data Applications has been hugely successful with over 30%?? of traffic volume on the ESnet network being carried on switched lightpaths. Outside of ESnet adoption of switched lightpaths has been slower and most production based big data applications instead use software tools such as Globus On Line.

Generally with large big data applications lightpath capability is only needed for short durations while the data is being transmitted. As a result a lot of work in GLIF has been focused on developing protocols to support short term setup, path finding an tear down of these lightpaths. Software packages that implement reservation and path computation capability such as OSCARS, DRAC, AutoBahn, etc have been under development for some time to support this establishment of end to end lightpaths.

Existing software that implements multi-domian BoD services are based around two primary protocols: NSI and IDCP. IDCP is the predecessor to NSI, but is widely deployed in R&E production environments in some countries. The NSI protocol is the more recent standards based protocol and comes in two flavours: NSI v1 and NSI v2.0
The protocols have implemented in the following software stacks:

(a) OSCARS v0.5 and v0.6 which supports IDCP and being extended to support NSI 2.0 interconnect

(b) AutoBahn Supports IDCP connect and NSI 1.0 and being extended to support NSI 2.0

(c) OESSSupports OESS and OSCARS (i.e. IDCP) and through OSCARS (i.e. IDCP and eventually NSI 2.0)

(d) Open NSA Implements NSI 1.0 being extended to support NSI 2.0

(e) Open DRAC Implements NSI 1.0 being extended to support NSI2.0

To date use of BoD or dedicated lambdas is a manual process between the application and the BoD service. The most common solution is to establish static routes between source and destination once an end-to-end circuit is established. This allows existing applications to re-route traffic over the dedicated circuit.

Currently there is considerable discussion on the need to integrate point-to-point transport into the applications. But this will requires considerable development effort where in the scientific workflows so that point-to-point circuits can be requested and terminated. Network engineers need to understand the full science data production cycle, with point-to-point BoD being part of the picture.

For a true GLIF vision BoD services need to scalable and discoverable. While considerable success has been achieved with layer 1 and 2 topology discovery and circuit setup, GLIF engineers are still stymied by the fact that most applications are based on the UNIX stack which assumes ports and packet buffers for forwarding and storing data. The forwarding of packets to a particular IP address is carried out by an entirely separate process – usually a TCP cron job. As a result most applications are separated by many layers from the actual transport of data. This also makes scalability and wide scale discoverability of NSI services a challenge. Currently for example all layer 2 networks protocols including NSI assume a single IP domain or subnet for the end-to-end circuit. After a circuit is established a process is required to agree upon a common subnet in order to transfer data. Generally this is done manually. Clearly this will not scale. We need a process where local IP address can be used with an end-to-end circuit. Ideally the discovery and advertisement of these IP addresses that can be reached with an end-to-end circuit will be an important part of interfacing with applications.

Composable services might be one possible approach where users can construct a number of data management workflows of which one element of the workflow is a BoD service such as NSI. To move in this direction BoD protocols or their implementations need to expressed as a composable service as for example used in OpenNaaS.

Another architectural approach is having specialized “Science as a Service” organizations that manipulate, transfer and process data on behalf of the science user. This is an application/network model most popular amongst the genomics and bio-informatics communities who are frequently using commercial companies to transfer and process their data. Integrating BoD or user defined traffic engineering with these specialized applications might be of value to these specialized organizations.

Recently, some promising developments have occurred with the integration of Sciencce DMZ with Globus On Line to transparently setup an end to end lightpath as requested for Globus file transfer.

The demand for end to end switched lightpaths for BoD will largely depend on how much future data will be processed, stored and managed within commercial clouds. While there will be a need for some specialized high performance computation facilities, in the next few years the overwhelming volume of computational science may be done within the cloud. Lightpaths in that case will only then be needed to transfer data to the cloud and provide uncongested bandwidth for visualization of data etc.

For example, in the high energy physics world, it is conceivable that the bulk of Tier 2 and Tier 3 computation and storage could be done entirely within commercial clouds.

What impact would this have on the need and design of switched lambda networks between NRENs?

2.2.5 Large Sensor Applications

Probably one of the most enduring *direct lambda connectivity* requirements will be for delivering data from large instruments and sensors to computational and storage devices around the world. It should be noted that this application is a specialized subset of "Big Data" as the assumption is that this involves the transmission of raw data directly from the sensor instrument.

Although the LHC data from CERN is often included in this category it is not necessarily the case as the raw data from the CERN instruments is initially processed and stored on computation facilities locally on CERN. It is only the processed data that is transferred from the CERN Tier 1 to the global set of Tier 1 sites around the world. A more accurate example of large sensor applications is the transmission of raw data from radio telescopes to distant processing sites such as the proposed Square Kilometre Array (SKA) project where each satellite dish may be transmitting continuously up a Terabit per second of data. Other examples include the various ocean based observatories, remote optical telescopes, etc

Even if all scientific data computation and storage eventually moves to commercial clouds there will still be a requirement for global lambda networks to link large sensor arrays to the clouds. Genomic sequencers, material testing, remote sensor arrays, etc will still need to get their data to the cloud.

However, it is expected that most of these large sensor applications will have predefined sources and sinks for most data flows and therefore are less likely to require globally discoverable end points. Preconfigured or pre-computed topologies may also be acceptable in this environment.

2.2.6 Aggregating High Speed Wireless Network Applications

A related application to large sensor will be aggregating traffic from large sensor arrays. It is expected future demand for multi-domain lightpaths will be aggregating traffic from Wi-Fi-based hotspots and 3G/4G off-load that are provided by third parties. Various research applications in wearable health sensors, vehicle sensor and environmental sensor arrays will require wide scale wireless connectivity via 4G and WiFi networks. Some of these highly mobile sensor arrays will span multiple NRENs and continents such as tracking shipments of highly sensitive research specimens, disease monitoring, migratory patterns, etc.

As well many cell phone companies are interested in deploying 4G/Wifi towers on or near university campuses. Their biggest data users are students streaming videos and downloading music. The faster and sooner this traffic can be offloaded to a Wifi or optical network the better. The biggest challenge in these environments will be how to define and segregate commercial traffic from R&E traffic. Cell phone coverage does not stop at the campus perimeter. Many cell phone companies are also deploying direct lambda connections to individual antenna on radio towers (RF over optical). It

is likely that these lambdas will need to traverse NREN networks to reach cell phone facilities serving university campuses.

Many NRENs are also entering into agreements to extend eduroam across entire nations serving coffee shops, municipal wireless networks and other hot spots. As well many NRENS would like to enable global wireless roaming through Eduroam for both their WiFi and 4G connected clients. Integrated 4G/WiFi networking will also enable anytime, anywhere, any device access to research and education content and services.

Such capabilities will have significant implications for future lambda network architecture and interfacing to commercial networks. It is expected that the lambda architecture for most wireless sensor applications will be to interconnect aggregation points at various points around the world. For eduroam and campus applications the requirement will be likely for parallel independent managed networks to separate commercial and R&E traffic.

2.2.7 Low carbon emission applications

A growing demand for both direct lambda and underlying lambda capability is to reduce the energy and carbon impact of computing and storage. As researchers and funding councils become aware of the high energy cost of computing and storage on there is a growing incentive to move these facilities off campus or to a commercial cloud. Up to 50% of a research university's electrical consumption can be due to the electrical consumption of the computing and networking equipment. Even for non-research intensive universities, or those without a data centers, computing and networking can be 20-40% of the total electrical energy consumption.

SURFnet for example has established lambda connectivity to GreenQloud in Iceland to enable Dutch researcher to use a low carbon computation cloud. NORDUnet is locating is also locating some computational facilities in Iceland in order to reduce the carbon impact of research computing in the Nordic countries. Universities in the Boston area have built a data center 80 miles west of Boston at a municipal owner power dam to enable low carbon computing for their respective institutions.

The demand for low carbon computing and storage has been slow to take off because as yet there has been no international agreement to place a price on carbon. If countries individually or collectively do decide to put a price on carbon, either through cap and trade or a carbon tax the cost of campus based computing and storage could jump dramatically. For example in a paper published in Educause on Green computing it was estimated that for a university like University of Michigan whose power is completely coal based, a cap and trade system would increase the cost of campus computing alone by $7 million per year, if carbon is priced at $20 per metric tonne.

Hurricane Sandy and other large powerful storms have also awoken researchers that in addition to make computing and networking more green, we also need to build networks and research facilities that can survive climate change. Low lying countries with major research facilities like the Netherlands, parts of the UK and USA are particularly vulnerable.

While this still remains a controversial research topic, a number of research teams are exploring how to deploy lambda and cloud networks that are only powered by renewable resources such as solar and wind power facilities. Since these networks are not connected to the electrical grid it also means they are more likely to survive and continue to operate through major storms like Hurricane Sandy. It is expected that the frequency and intensity of such storms will increase in the coming decades.

There are many "green" network initiatives around the world largely looking at measuring energy efficiency. Within GLIF "green" Perfsonar is a new project that intends to measure energy consumption of GLIF networks. Currently there are only two research projects looking a building survivable networks – the GreenStar project in Canada and Mantychore in Europe. Because optical networks require very little power, compared to IP routed networks they can be easily powered by renewable sources. But because renewable energy is unreliable, as it is dependent on the wind and sun, the need to quickly and frequently switch and setup optical paths is critical for such an architecture.

2.2.8 Experimental Testbeds

There are a number of next generation Internet and optical testbeds that require lambda connectivity to support international collaborative research.

Although these testbeds will be critical for the future direction of the Internet, it is unlikely that they will have much impact on any end-to-end architecture design requirements for GLIF. Most testbed, as part of their essence of being a testbed, establish their own interconnection and peering policies with like minded researchers around the world. For the most part, all they require is persistent, manually configured lightpaths.

New testbed initiatives including the GENI and GEANT3+ are looking at creating 'sliced' testbeds on-demand by leveraging unused production network capacity. They have elaborate mechanisms for user admission control and resource management, and are looking at leveraging on-demand lightpaths to build a testbed. These initiatives can leverage the GLIF lightpath capabilities though the GLIF and GOLE operators will need to learn and deploy a policy management infrastructure that will allow them to seamlessly allocate resources and participate in multi-domain testbeds.

2.2.9 Private Lightpath or SDN networks across Multi-domain optical networks

Private Lightpath Networks (often referred to as Private Optical Networks) are to date the largest production use of GLIF and NREN optical infrastructure. Many NRENs have deployed Private Lightpath Networks for a variety of research and multi-institution applications. There are also several Private Lightpath Networks that span multiple NRENs. Examples include LHCOPN, CAVEwave, GLORIAD, HPCLnet, etc. As well there are at least two SDN testbed networks that span multiple NRENs.

To date, these private lightpath networks are for the most part stitched together manually. It is expected that in the coming years there will be a demand for end-to-end multi-domain private SDN networks similar in nature to what we are seeing for in

terms of private lightpath networks. These private SDN networks initially may not be trans-continental, but address the challenge to link a campus SDN network to a backbone SDN network and ultimately to a destination SDN campus network.

Over the past number of years there have a plethora of proposals to develop routing protocols for inter- for multi-domain VPNs at all 3 layers of the network stack. For a variety reasons, none of these protocols have yet achieved production in the commercial world. The complexity of business relationships to deliver inter-domain or multi-domain VPN services has significantly hindered their deployment. In the R&E world the NSI perhaps has come closest to enabling multi-domain private lightpath networks – but as mentioned previously the issues of routing and address space naming have not been fully resolved.

One of the challenges of building multi-domain VPNs is whether the VPN should extend across both the forwarding plane, control plane or even the management plane. In the commercial world most of the focus has been on forwarding plane VPNs. But in the R&E world, particularly with SDN networks, it would also be useful to have VPNs that also extended all, or portions, of the both the control and management plane of the VPN across the multiple underlying optical networks.

Proposed protocols that address the control and management plane as well the forwarding plane have included overlay routing where a single domain network is deployed over multiple underlying single management domain networks. The argument made for this approach is that optical or SDN inter-domain or multi-domain signalling and path finding is too difficult and too intractable as it not a technology challenge but a business issue that needs to be resolved. Instead by allowing virtual switches or routers to be made available by the underlying networks to the overlay network as composable elements, an overlay network can have a separate end-to-end management and control plane. This was the essence of UCLP, now called OpenNaaS – “Network as a Service”.

Other approaches to propagating VPNs across multiple domains included adding to BGP attributes to give preference to an optical path or MPLS VPN, such as back to back VRFs and BGP PE-CE Routing Protocols. Another related development in this area is mapping OpenFlow Flow table splitting to MPLS VRFs. Today, virtual networks are defined by MPLS VRFs, or VLANs or by overlay tunnelling. OpenFlow capabilities of flow mapping allow network engineers to define virtual networks using any criteria you like. It could be source MAC and destination MAC (roughly equivalent to VLANs), or source physical port to destination physical port. Network engineers could also define a virtual network by source and destination IP addresses.

Addressing the inter-domain or multi-domain end-to-end architecture of either optical or SDN networks will probably be the biggest challenge for GLIF in the coming years. For direct lambda connectivity with short duration connections inter-domain or multi-domain optical signalling will be essential. But where underlying lambda networks are deployed as part of a persistent network service such as CDN, global peering, etc other approaches may be possible.

3.0 Infrastructure Capabilities

Although great success has been achieved in establishing lightpaths across multiple independent managed networks and GOLEs achieving connectivity across campus networks to the researcher's desktop remains an elusive goal. Recently technologies such as Science DeMilitarized Zones (Science-DMZs) and campus Software Defined Networks (SDN) promise to alleviate some of the campus lightpath challenges. But the interconnection and interoperability of DMZs, SDNs and other campus network architectures with global interconnected lightpaths as a seamless architectural vision remains an unrealized objective. Not only is a seamless physical interconnection required, but the specification of all the user interface, management, measurement, operational and control aspects of the architecture must be detailed as well.

Compounding the problem of defining an end-to-end lightpath architecture is the increasing need for researchers to interconnect lightpaths or SDN flows to commercial databases, clouds and computational resources. In some cases the end-to-end solution may not even touch the campus network. Instead a researcher may wish to connect to the output of a remote instrument directly to a commercial cloud. Building end to end solutions in the academic/research world with its commitment to openness and collaboration is one thing, but this can be quite a bit more challenging in the commercial world with its concerns about competition, privacy, security etc.

The ultimate vision of the GLIF architecture task force is that a researcher, or an application can compose or create an end-to-end lighpath or SDN solution across a campus, multiple GOLEs and networks using a simple interface such as SURFconext, Globus OnLine or Comanage/NET+. All the necessary management, measurement and control tools would also be incorporated in such an interface.

3.1 Infrastructure Combinations

To simplify the complexity of interconnecting many independent lightpaths across multiple networks, many services may be consolidated into a much smaller number of abstracted services which can also be an advertised service as part of an end-to-end solution.

With any type of end-to-end switched inter-domain or multi-domain architecture the role and process of initiating and terminating parties must be carefully addressed. How does a researcher at one campus initiate an end-to-end lightpath to a research or database another campus if they have no authority or credentials to setup a lightpath at the destination campus? At the network to network level authenticating and accepting lightpath requests across a GOLE or intervening network, although not trivial, is relatively easy in comparison. Can a researcher delegate authority to allow external parties to setup a lightpath across the campus network? Or should a researcher only be authorized to "meet in the middle" at a GOLE or campus border router i.e. all lightpaths or SDN flows terminate at GOLEs , one set from the designated originator and another set from the designated recipient?

To help clarify the requirements for the GLIF end-to-end architecture it would be useful to document the technology challenges of the possible applications described in the previous section. The following list is a summary of the various possible

architectures mapped to the applications described previously with a more detailed analysis of each case given separately:

(a) True lightpath connectivity across campus with direct interconnect to global GLIF BoD services;

(b) SDN network on campus (most likely OpenFlow) to interconnect GLIF lightpaths or MPLS VRF at campus interface for traffic engineering applications;

(c) Campus DMZ outside of campus network interconnected to GLIF facilities on the outward facing connection and IP connection facing inward;

(d) Campus IP network with VPNs (MPLS) or VLANs to interconnect to GLIF facilities at campus egress;

(e) Terminating end-to-end lightpath on a commercial interface: e.g. cloud; and

(e) Establishing lightpath connection on a remote instrument network to a commercial cloud or database

Each of these use cases will be explored more fully in the following sections:

3.1.1 True lightpath connectivity across campus with direct interconnect to global GLIF services

A small number of university and research campuses have local area optical networks with dynamic switching of optical lightpaths across the campus as well as direct connections to GLIF network facilities. Most of these optical networks are operated completely independent of the campus network and are responsible for their own security and global connectivity. In some cases servers connected to the optical campus network are firewalled from the campus IP network.

In many cases the end-to-end lightpath is not from campus to campus but from GOLE to GOLE. In this environment the GOLE acts much like a DMZ where researchers locate their test and computation gear.

Some of these networks support NSI (or its predecessor protocols). As such setting up end-to-end lighpaths is rather trivial compared to the other use cases. Many of these networks are used for specialized applications such as experimental testbeds.

3.1.2 SDN network on campus to interconnect GLIF lightpaths at campus interface

A growing number of universities, research campuses and large data centers are deploying various SDN networks, mostly variations of OpenFlow. SDN or OpenFlow allows the network manager to easily and quickly configure dedicated flows to various researchers and users on campus. With OpenFlow these devices can be

centrally managed and configured – which is often very appealing to a campus network manager.

For the most part ingress and egress to the campus is at the IP layer through a campus border router. Considerable research is going on to map OpenFlow VLANs to MPLS and GMPLS VPNs using VRFs. A few examples include proof of concept with OSCARS demonstrated at SC11 (http://sc11.supercomputing.org/schedule/event_detail.php?evid=rsand110) and work at Internet2 with NDDI.

SDN networks, particularly OpenFlow, because they separate control plane from forwarding plane, may allow advertisement of special IP routes to complete BoD applications, that are persistent regardless of the status of the flow path itself. With many campuses implementing multi Gigabit interfaces, establishing specialized routes on a given interface for either BoD or traffic engineering applications will be useful.

3.1.3 Campus DMZ outside of campus network interconnected to GLIF facilities on the outward facing connection and IP connection facing inward

To get around many of the bandwidth and connectivity limitations of campus networks, ESnet in particular has been promoting the concept of DMZs. With a DMZ a campus researcher can upload or download large data files to a server outside of the campus firewall. The DMZ is directly connected to GLIF optical infrastructure. For the most part the DMZ is considering the terminating device and the rest of the campus network including researcher's services remain hidden from external users. DMZ also come configured with PerFSonar and other network management devices which makes measurement easier.

As researcher's progressively move to using commercial clouds for storage and computation the DMZ may in fact become an intermediate stop point for a data flow between an instrument and a commercial cloud facility. The interconnection to the campus network becomes less relevant and would let researchers do large data analysis from their local coffee shop. In that case the ability to set up lightpaths from the DMZ or the originating instrument itself to a commercial cloud becomes important.

3.1.4 Campus IP network with VPNs (MPLS) or VLANs to interconnect to GLIF facilities at campus egress

This is the most common interconnection, other than using general IP for interconnecting researchers with GLIF infrastructure. In many cases, campus configuration problems bedevil the setup of end-to-end lightpaths which has resulted in the deployment of Science DMZs.

Considerable work has been done in NSI, IDCP and other lighpath switched protocols to map optical lightpaths to MPLS tunnels.

3.1.5 Terminating end-to-end lightpath on a commercial interface: e.g. cloud

As mentioned previously there is growing demand by researchers to use commercial clouds and databases for the uploading downloading of large datasets, as well as directly forward data from instruments.

In most environments the connection to a commercial cloud provide such as Google, Amazon, Azure, GreenQloud, etc is owned and controlled by a NREN. Connectivity is provided at the IP layer through standard IP addressing and naming. However, the need for a researcher to have a direct connection independent of the IP service layer is growing. This will introduce a host of problems of how to terminate individual lightpaths through perhaps a single 10G pipe to a cloud service provider. Most commercial cloud providers have not yet scaled up to handle this type of large IO data flows (although they do handle teabits of IP flows).

As most commercial cloud providers are not yet ready to accept lightpaths, it is likely that the NREN will have to offer a proxy service and do the traffic engineering to terminate and manage lightpath requests to a commercial cloud – in effect operating a "reverse" DMZ on behalf of the commercial cloud operator. The ability, therefore to terminate originating lightpaths from third parties will be an essential feature.

Considerable more research has to be done for this use case. SURFnet in partnership with GreenQloud in Iceland is probably the most advanced in this field. Their experience will be a useful in helping other NRENs and researchers use lightpaths to transmit and receive data from cloud providers.

It is this environment, as well as that of global peering, where inter-domain or multi-domain control plane and management plane extensions may be required.

3.1.6 Establishing lightpath connection on a remote instrument network to a commercial cloud or database

This example is very similar to that of interfacing users to clouds, but where both ends of a network connectivity are outside the management domain of the user and the institution's network. This use case is the ultimate example of third party delegation of lighpaths – where an independent researcher may, for example, want to setup a lightpath from CERN to a commercial cloud provider such as Amazon. None of the lightpaths may terminate or come even close to touching the researcher's own campus network. Delegation of control plane and management plane resources to a third part is the primary challenge in this scenario.

3.2 GOLEs

GOLEs are the linchpin of the global community of R&E networks, much like IXPs are the major interconnection points for the global Internet. Besides support the interconnection of lambdas, many GOLEs provide additional functionality such as hosting for performance measurement equipment and test equipment for various network research experiments.

In many situations GOLEs also act as DMZs for the termination of lightpaths and hosting computation and data storage facilities outside of university campuses. It is likely that GOLEs role will continue to expand as they are the most logical place to host CDN nodes and do traffic handoffs between commercial and R&E networks for wireless and SaaS applications.

Currently the various GOLEs around the world support a variety of network services. However, gradually there appears to be a convergence on Ethernet as the common layer 2 transport protocol with NSI as the common BoD protocol. As yet there is no common standard for sub channel partitioning, with some GOLEs supporting MPLS-TE and others using SONET or SDH channel partitioning. However, with the recent decision of one major optical network manufacturer to abandon PBT, it would appear that MPLS-TE (or MPLS-TP) will become the standard for creating sub-channels on lambdas.

4.0 Topics of Discussion and Next Steps

A measure of success for GLIF will be the ability of the NRENs and GLIF to collectively work together to develop and offer scalable and discoverable networking services across the multi-domain management and technological environment in which they co-exist.

The three main areas identified where further work needs to be done with the auspices of GLIF can be summarized with these inter-related objectives as follows:

> (a) Developing Bandwidth on Demand and Traffic Engineering toolsets that use both NSI and SDN but interoperate with (G)MPLS-TE;
>
> (b) Integrating lambda and SDN networking within applications that are routeable, discoverable and scalable at Internet layer 3 and layer 2; and
>
> (c) Developing inter-domain and multi-domain SDN for both the forwarding, control and management planes.

These objectives are described in more detail in the following sections.

4.1 Developing Bandwidth on Demand and Traffic Engineering toolsets that use both NSI and SDN but interoperate with (G)MPLS-TE

As noted earlier user initiated BoD applications for big data transfers are only a small subset of use cases for lambda networking. In fact the argument for big data transfers, other than raw data from remote instruments, may largely disappear as a lot of data and computation is done and remains with the cloud. To date, traffic engineering using lambda networking has largely been ignored by the GLIF community.

Historically protocols like MPLS and others were originally intended to enable circuit like quality for QoS applications much like lambda networking, but found their use largely for traffic engineering. It is not unreasonable that the bulk of applications for lambda network will be applied to traffic engineering as well.

Traffic engineering will require dynamic and static configurations requiring protocols like NSI, but more likely with entail considerable wide are SDN development.

Most traffic engineering applications to date, have been within a single management. But big science and the NRENs will require traffic management tools that span multiple domains. Needless to say this is a challenge fraught with both political and technical issues. But it only under the auspices of GLIF where a community of trusted federated NRENs might have the ability to implement inter-domain and multi-domain traffic management without commercial constraints.

4.2 Integrating lambda and SDN networking within applications that are routeable, discoverable and scalable at Internet layer 3 and layer 2

The holy grail of lambda networking is to allow both end-to-end and traffic engineering applications automatically set up lightpaths or SDN links across multi-domain networks.

The challenge is not only to make applications aware, but also to insure that lambda and SDN networking are fully compliant and inter-operable with existing IP layer 3 and layer 2 routing. Constraining inter-domain lambda or SDN links to a single IP subnet is not scalable and unlikely discoverable.

The ultimate success and survivability of GLIF will not to be an specialized "island" of unique networking protocols that exist outside of the global Internet, rather the tools developed by GLIF should complement and enhance the existing suite of Internet networks, services and applications. It is therefore important to that NSI and SDN inter-domain and multi-domain tools be discoverable and scalable within the global Internet. As yet network engineers have only begun to scratch the surface on this topic.

4.3 Developing inter-domain and multi-domain SDN for both the forwarding, control and management planes.

Control plane interoperability is still not achieved in a inter-domain, multi-domain, multi-vendor environment, albeit standardization promises such as the OIF UNI/NNI or the IETF GMPLS and control plane interworking coordination in GLIF. This may have an adverse effect on automated end-to-end provisioning in a complex environment that consists of interconnecting several backbones relying on various vendor architectures.

Ultimately control plane or management plane operation in a inter-domain or multi-domain, multi vendor environment may not be realistic. Instead composable services to create an overlay single domain solution may be more practical.

4.4 Content, Storage and its intersection with Networking

Named-data networking is a new method being proposed on building the content lookup, caching and distribution function within the network. Like any new paradigm, this has its doubters, but the multi-domain GLIF and its participants can effectively experiment with such a new technology and vets its applicability to science, universities and other constituents globally. This fits in especially with the CDNi initiative in the IETF discussed earlier in the document