Optical Network Infrastructure for Grid

Grid High Performance Networking Research Group	Dimitra Simeonidou (Editor) *University of Essex*
GRID WORKING DRAFT	Bill St. Arnaud *CANARIE*
draft-ggf-ghpn-opticalnets-0	Micah Beck *University of Tennessee*
Category: Informational Track	Bela Berde *Alcatel CIT Research*
http://forge.gridforum.org/projects/ghpn-rg/	Freek Dijkstra *Universiteit Van Amsterdam*
	Doan B. Hoang *University of Technology, Sydney*
	Gigi Karmous-Edwards *MCNC Institute*
	Tal Lavian *Nortel Networks Labs*
	Jason Leigh *University of Illinois at Chicago*
	Joe Mambretti *Northwestern University*
	Reza Nejabati *University of Essex*
	John Strand *AT&T*
	Franco Travostino *Nortel Networks Labs*

Status of this Memo
This memo provides information to the Grid community in the area of high performance networking. It does not define any standards or technical recommendations. Distribution is unlimited.

Comments: Comments should be sent to the GHPN mailing list (ghpn-wg@gridforum.org).

Contents

1. Introduction

1.1 Background

During the past years it has become evident to the technical community that computational resources cannot keep up with the demands generated by some applications. As an example, particle physics experiments [1,2] produce more data than can be realistically processed and stored in one location (i.e. several Petabytes/year). In such situations where intensive computation analysis of shared large scale data is needed, one can try to use accessible computing resources distributed in different locations (combined data and computing Grid).

Distributed computing & the concept of a computational Grid is not a new paradigm but until a few years ago networks were too slow to allow efficient use of remote resources. As the bandwidth and the speed of networks have increased significantly, the interest in distributed computing has taken to a new level. Recent advances in optical networking have created a radical mismatch between the optical transmission world and the electrical forwarding/routing world. Currently, a single strand of optical fiber can transmit more bandwidth than the entire Internet core. What's more, only 10% of potential wavelengths on 10% of available fiber pairs are actually lit [3]. This represents 1-2% of potential bandwidth that is actually available in the fiber system. The result of this imbalance between supply and demand has led to severe price erosion of bandwidth product. Annual STM-1 (155 Mbit/sec) prices on major European routes have fallen by 85-90% from 1990-2002 [4]. Therefore it now becomes technically and economically viable to think of a set of computing, storage or combined computing storage nodes coupled through a high speed network as one large computational and storage device.

The use of the available fiber and DWDM infrastructure for the global Grid network is an attractive proposition ensuring global reach and huge amounts of cheap bandwidth. Fiber and DWDM networks have been great enablers of the World Wide Web fulfilling the capacity demand generated by Internet traffic and providing global connectivity. In a similar way optical technologies are expected to play an important role in creating an efficient infrastructure for supporting Grid applications [5], [6].

The need for high throughput networks is evident in e-Science applications. The USA National Science Foundation (NSF) [7] and European Commission [8] have acknowledged this. These applications need very high bandwidth between a limited number of destinations. With the drop of prices for raw bandwidth, a substantial cost is going to be in the router infrastructure in which the circuits are terminated. "The current L3-based architectures can't effectively transmit Petabytes or even hundreds of Terabytes, and they impede service provided to high-end data-intensive applications. Current HEP projects at CERN and SLAC already generate Petabytes of data. This will reach Exabytes (10^{18}) by 2012, while the Internet-2 cannot effectively meet today's transfer needs."

The present document aims to discuss solutions towards an efficient and intelligent network infrastructure for the Grid taking advantage of recent developments in optical networking technologies.

1.2 Why optical networking for the Grid

1.2.1 Grid applications and their requirements for high speed, high bandwidth infrastructure

It is important to understand the potential applications and the community that would use lambda or optical Grids.

In today's Internet we have a very rich set of application types. These applications can possibly be categorized as follows:

- Large file transfer between users or sites who are known to each other e.g. high energy physics, SANs
- Anonymous large file transfers e.g. music and film files
- Small bandwidth streams - e.g. audio and video
- Large bandwidth streams - e.g. Data flows from instrumentation like radio telescopes
- Low bandwidth real time interactive - e.g. web, gaming, VoIP, etc
- High bandwidth real time interactive e.g. large distributed computing applications
- Low bandwidth widely dispersed anonymous users - e.g. web pages

It is still unknown what will be the major applications for lambda or optical Grids. How many of these application types will require dedicated high speed optical links in the near future? It would seem unlikely that all the application types we see on the Internet today will require optical grids. One early obvious application is large data file transfers between known users or destinations. Some researchers have hypothesized the need for a couple high bandwidth interactive applications - such as interactive HDTV.

Currently those who require lambda Grids for large data file transfers are well defined communities where the members or destination sites are known to each other. Such communities include the high energy physics facilities around the world (which are broken into smaller specific application communities - ATLAS (CERN), CMS (CERN), D0 (Fermilab), KEK (Japan). Other examples are the virtual observatories, SANs and very long base line interferometer projects. These communities are relatively small and maintain long lived persistent networked relationships.

The need for "anonymous" large file transfer to unknown users outside of their respective communities is currently a limited requirement.

This is not to say there will be no need for optical networks for traffic engineering, aggregation and similar "network" requirements.

Emerging network concepts such as Logistical Networking (described below) impose a new requirement for high bandwidth infrastructure and promise a wide range of applications.

Logistical Networking

Difficult QoS requirements (for instance, latency lower than the speed of light allowing access to remote data) can in some cases be achieved by using large bandwidth, aggressively prefetching data across the network and storing it in proximity to the endpoint. If the data required by the application can be predicted "accurately enough" and "far enough in advance", and storage availability close to the endpoint and wide area bandwidth are high enough, then the latency seen by application may be reduced to the latency of local access, except for an initial delay in start-up.

But, what if the data being prefetched is produced on demand by a cluster capable of filling the large pipe? Then the high bandwidth pipe is in fact tying together two halves of a distributed system, one the server and one the client, and the data being transferred may never exist in its entirety at the server, and it may never exist in its entirety at the client (if storage is limited, and prefetched data cannot be cached indefinitely). This is called a "terapipe," and it may have very broad applicability as an application paradigm for using high bandwidth networking and storage.

This approach is an example of Logistical Networking that may have practical applications as shown in a data visualization application (Remote Visualization by Browsing Image Based Databases with Logistical Networking Jin Ding, Jian Huang, Micah Beck, Shaotao Liu, Terry Moore, and Stephen Soltesz Department of Computer Science, University of Tennessee, Knoxville, TN, to be presented at SC03)

In this case the application had to be rewritten somewhat to produce data access predictions and supply them to a layer of Logistical Networking middleware that was responsible for the prefetching. In the experiments reported, the bandwidth of the pipe is not that high (20-40 Mbps) so the resolution of the images being browsed had to be limited (latency seen by the application was equivalent to local at 300x300, but not at 500x500). The size of the entire dataset was just 10GB. Increasing the resolution increases the storage and bandwidth requirements proportionately; full screen at 1400x1050 would require 100s of Mbps; serving a Power Wall at that resolution would easily require multiple Gbps of bandwidth and TBs of storage.

This "logistical" approach to using bandwidth can generate speculative transfers of data that are never used by the application. And if predictors are not good enough to mask circuit setup time, it may be necessary to keep a pipe open in order to respond to unexpected demands. On the other hand, it can allow an application to achieve latencies that are better than the lower bound imposed by the speed of light. It has the charm of not requiring a lot of detailed network programming - just a "good enough" predictor of data accesses and "high enough" bandwidth. If prestaging became a popular approach to achieving QoS, the demand for large pipes might increase greatly, particularly if good predictors were hard for application developers to supply.

1.2.2 Optical networking for high bandwidth applications

Grid applications can differ with respect to granularity of traffic flows and traffic characteristics such as required data transaction bandwidth, acceptable delay and packet loss. Here we specifically consider applications with high bandwidth requirements. Some of these applications (e.g. particle physics, CERN [9]) are sensitive to packet loss and require reliable data transmission. In contrast, there are high bandwidth Grid applications (e.g. radio astronomy [10]) that are sensitive to the packet loss pattern rather than the packet loss. There are also specific applications [11] that they may require bulk data transfers for database replication or load balancing and therefore packet loss minimisation is necessary to increase performance. Finally some emerging Grid applications (e.g. video-games for Grid [12]) require real time (short delay), long lived, relatively small bandwidth but potentially large number of users. Foster [13] proposes that Grid computing can support a heterogeneous set of "Virtual Organizations" (VO), each composed of a number of participants with varying degrees of prior relationship who want to share resources to perform some task.

Despite the above mentioned differences, there are two main common requirements generated by a large number of Grid applications:

- Large amounts cheap bandwidth provisioned and scheduled on-demand
- User or application management and control of the network resources (i.e. set-up self-organized distributed computing resources and facilitate bulk data transfers)

A number of other requirements concerning throughput, priority, latency, QoS and storage capacity will also influence the Grid network design but they are more specific to the type of application. Grid applications are also likely to differ in the number and type of participants, and also in the degree of trust between the participants [13].

A new network concept is now emerging to satisfy Grid application requirements. This is a network where resources such as ports, whole equipment, even bandwidth are controlled and maybe owned by the user. Furthermore, in contrast to traditional (telecommunications) networks where applications are allocated resources and routed over fixed network topologies, in Grid networks, resources under user/application control are organized in an automated way to provide connectivity without getting the permission from a carrier or a central authority. In other words, the user will drive its own virtual network topology.

Optical Technologies are best suited to fulfill some of these requirements, i.e. to offer huge capacity (theoretically up to 50 Tb/s/fiber) and relatively low latency. What’s more, WDM & tunable technologies in combination with optical switching can provide dynamic control and allocation of bandwidth at the fiber, wavelength band, wavelength or sub-wavelength granularity in optical circuit, burst, or optical packet systems. Today’s optical technologies support fast and dynamic response of bandwidth offering the capability to provide bandwidth services dynamically controlled by individual users/applications. This has been made possible by the development of a distributed control plane based on established IP/MPLS protocols

Based on this capability, future data-intensive applications will request the optical network to provide a point-to-point connection on a private network and not on the public Internet. The network infrastructure will have the intelligence to connect over IP network (packet) or to provide λ (circuit) to the applications. A λ service provided through OGSI will allow Virtual Organizations to access abundant optical bandwidth through the use of optical bandwidth on demand to data-intensive applications and compute-intensive applications. This will provide essential networking fundamentals that are presently missing from Grid Computing research and will overcome the bandwidth limitations, making VO a reality.

Despite these features, optical networks have been developed with telecommunications applications in mind and the implementation of a Grid optical network imposes a lot of new challenges.

General requirements in this type of optical network can be summarized as follows:

- Scalable, flexible, and reconfigurable network infrastructure
    - It can be argued that initially optical grids are going to serve a small set of specialized applications and thus scaling becomes a minor and unimportant issue. However, we have already identified new applications requiring optical infrastructure and there seems to be a strong possibility that other applications will emerge. It is therefore significant addressing issues of scale. Scalability is an inherent attribute of the Grid vision, and enables the creation of ad hoc virtual organizations. Scalability considerations would be a big factor on the design and engineering decisions one would make in deploying an optical grid
- Ability to support very high capacity - Bulk data transfer
- Low cost bandwidth
- Bandwidth on demand capabilities for short or long periods of time between different discrete points across the network. Various schemes will be supported, for the management and exchange of information between Grid services (i.e. point and click provisioning, APIs and/or OGSI/OGSA services) that an application can use to exploit agile optical networks
- Variable bandwidth services in time
- Wavelength and sub-wavelength services (STS-n, optical packet/flow/burst)
- Broadcasting/multicasting capabilities
- Hardware flexibility to be able to support wide range of different distributed resources in the network
- High resilience across layers. In particular, a resilient physical layer will entail an number of features including resilient wavelengths, fast and dependable restoration mechanisms, as well as routing diversity stipulations being available to the user
- Enhanced network security and client-network relationship both at user-network level (UNI security) and network-network level (NNI and data path security)

- Ability to provide management and control of the distributed network resources to the user or application (i.e. set-up self-organized distributed computing resources and facilitate bulk data transfers)
-

1.2.3 Other factors supporting the need for optical infrastructure

1.2.3.1 Limitations of packet switching for data-intensive applications

In order to understand why optical networking for Grid, we need also to understand the current limitations of packet switching for Grid and data-intensive applications. The current Internet architecture is limited in its ability to support Grid computing applications and specifically to move very large data sets. Packet switching is a proven efficient technology for transporting burst transmission of short data packets, e.g., for remote login, consumer oriented email and web applications. It has not been sufficiently adaptable to meet the challenge of large-scale data as Grid applications require. Making forwarding decisions every 1500 bytes is sufficient for emails or 10k -100k web pages. This is not the optimal mechanism if we are to cope with data size of six to nine orders larger in magnitude. For example, copying 1.5 Terabytes of data using packet switching requires making the same forwarding decision about 1 billion times, over many routers along the path. Setting circuit or burst switching over optical links is a more effective multiplexing technique.

1.2.3.2 End-to-end Transport protocol Limitations

Responsiveness

TCP works well in small Round Trip Time (RTT) and small pipes. It was designed and optimized for LAN or narrow WAN. TCP limitations in big pipes and large RTT are well documented. The responsiveness is the time to recover form single loss. It measures how quickly it goes back to using a network link at full capacity after experiencing a loss. For example, 15 years ago, in a LAN environment with RTT=2ms and 10Mbs the responsiveness was about 1.7ms. In today's 1Gbs LAN with RTT, if the maximum RTT is 2ms, the responsiveness is about 96ms. In a WAN environment where the RTT is very large the RTT from CERN to Chicago is 120ms, to Sunnyvale it is 180ms, and to Tokyo 300ms. In these cases the **responsiveness is over an hour** [9]. In other words, a single loss between CERN and Chicago on a 1Gbs link would take the network about an hour to recover. Between CERN and Tokyo on a 10GE link, it would take the network about **three hours to recover** [9].

Fairness

In packet switching, the loss is an imperative mechanism for fairness. Dropping packets is in integral control mechanism to signal end-system to slow down. This mechanism was designed in multi streams sharing the same networking infrastructure. However, there is no sharing in dedicated optical link; thus, fairness is not an issue. There is no competition for network resources. Fairness need to be addressed in the level of reservation, scheduling and allocating the networking resources.

1.2.3.3 New transport protocols

In order to address some of the above packet switching limitations, new transport protocols have started to evolve. Examples are GridFTP FAST, XCP, Parallel TCP, and Tsunami. The enhancements in these protocols are done via three mechanisms: 1) tweaking the TCP and UDP settings; 2) transmitting over many streams; and 3) sending the data over UDP while the control is done in TCP.

Transmitting over TCP without the enhancements results in about 20Mbs over the Atlantic. Recent tests have seen GridFTP to achieve 512Mbs , Tsunami at 700Mbs , and in April 2003, FAST achieved 930Mbs from CERN to SLAC.

None of the above protocol can fully utilize OC-192 links. Statistical multiplexing of multiple streams of the above protocols can do current utilization of OC-192.

2. Photonic Grid network Characteristics

2.1 Network topology

The Grid enabled optical network will require the network topology to migrate from the traditional edge-core telecom model to a distributed model where the user is in the very heart of the network. In this type of network the user would have the ability to establish true peer-to-peer networking (i.e. control routing in an end-to-end way and the set up and teardown of light-paths between routing domains).

To facilitate this level of user control, users or applications will be offered management/control or even ownership of the network resources of network resources from processing and storage capacity to bandwidth allocation (i.e. wavelength and sub-wavelength). These resources could be leased and exchanged between Grid users. The network infrastructure, including network elements and user interface, must enable and support OGSA. Through OGSA the Grid user can only have a unified network view of its owned resources on top of different autonomous systems. The resources can either be solely owned or shared with other users.

Another topological alternative that could be used in conjunction with user-owned capacity is an OVPN. This means leasing wavelengths on commercial DWDM systems on a link-by-link basis. The status of these would be advertised to the Grid participants and they could dynamically connect capacity on a series of links together along a route they define by signaling messages.

These new topological solutions will have a direct impact on the design of optical network elements (optical cross-connects, add-drop multiplexers etc) and will impose new demands to the interface between the Grid user and network (GUNI[1]): i.e. The user through GUNI (see 3.3 for further for further details) will be able to access and manipulate the network elements. This requires propagation of significant network

[1] GUNI is the GRID User Network Interface with functionality not fully covered by the OIF UNI

element information to the application interface, information that today resides almost exclusively in the provider's domain. It also implies new types of network processes for discovery, naming, and addressing.

As an example:

- The optical network elements:
    - must be able to dynamically allocate and provision bandwidth on availability
    - have knowledge of adjacent network elements, overall network resources, and predefined user and network constrains
    - depending on application requirements, perform optical multicasting for high performance dynamic collaboration

- The GUNI will be able to schedule huge bandwidth (i.e. OC768) over predefined time windows and establish optical connection by using control domain signaling (e.g. GMPLS)

2.2 Optical switching technology and transport format considerations

An important consideration that would influence optical Grid network architecture is the choice of switching technology and transport format. Optical switching offers bandwidth manipulation at the wavelength (circuit switching) and sub-wavelength level through technologies such as optical packet and burst switching offering not only high switching granularity but also the capability to accommodate a wide variety of traffic characteristics and distributions.

A number of optical switching technologies and transport formats can be considered:

- Wavelength switching
    - Wavelength switching (sometimes called photonic switching, or λ-switching) is the technology used to switch individual wavelengths of light onto separate paths for specific routing of information. In conjunction with technologies such as DWDM, λ-switching enables a light path to behave like a virtual circuit. λ-switching requires switching/reconfiguration times at the msec scale
- Hybrid router-wavelength switching
    - This architecture extends the wavelength switching architecture by adding a layer of IP routers with OC-48/192/768 interfaces between the Grid nodes and the optical network
- Optical burst switching
    - An optical transport technology with the capability of transmitting data in the form of bursts in an all-optical, buffer-less network, using either circuit switching (light paths), flow switching (persistent connection), or per-hop switching (single burst) services, depending on connection set-up message. The network is transparent to the content of a burst (analogue or any digital format) as well as to the data rate. Switching timescales will depend on the

length/duration of bursts in a particular network scenario. Typical values vary from few µsec to several msec.

- Optical flow switching
  - The switched entity is a set of consecutive packets in an active connection (ie packets form one source going to the same destination). Flow can be shorter than bursts (may be just 1 packet). A header is attached to the flow and it is routed and switched like a single packet. Buffering needed, which must be large enough to encompass the flow. Hop-by-hop path set-up. Advantages include integrity of transmitted sequence. The minimum flow duration will define the requirements for switching timescales. For optical networking at 10-40 Gb/sec, switching times at the nsec scale may be required
- Optical packet switching
  - The header is attached to the payload. At the switch the header is examined to determine whether payload is switched or buffered. Hop-by-hop path set up. Generally thought of as synchronous, but not necessarily so. Buffering may be a problem, due to lack of optical memory. Typical optical packet lengths vary from 50 bytes-15,000 or 30,000 bytes which clearly imposes a requirement for nsec switching technology

Most of the work to date assumes wavelength routing [14], because equipment such optical cross-connects (OXCs) is currently available. There is good evidence that optical burst or packet switching may eventually provide even better bandwidth and finer granularity [15]. In addition, application friendly switching such as optical flow switching can result in an improved end-to-end network performance [16].

The choice of format will be mainly driven by an understanding of the traffic characteristics generated by Grid applications. The expectation is that ongoing work on Grid will generate this information. It is likely that the right solution is going to vary between types of Grid applications. For example, wavelength switching may be the preferred solutions for moving terabytes of data from A to B, but appears to be inappropriate for video games applications, and the terabit router/OXC option may provide a competitive ready to deploy solution.

Decisions on switching and transport formats will also influence the design of optical network equipment as well as the management and the control of the network.

2.2.1 Wavelength Switching

Recent advances in Grid technology have promised the deployment of data-intensive applications. These may require moving terabytes or even Petabytes of data between data banks. However, the current technology used in the underlying network imposes a constraint on the transfer of massive amounts of data. Besides the lack of bandwidth, the inability to provide dedicated links makes the current network technology not well suited for Grid computing. A solution is needed to provide data-intensive applications with a more efficient network environment. This solution should provide higher bandwidth and dedicated links, which are dynamically allocated on-demand or by scheduled reservation.

Wavelength switching (WS) is a promising solution, and the required infrastructure to realize this promise is now within reach.

Future data-intensive applications will ask the optical network for a point-to-point connection on a private network or an OVPN. Intelligent edge devices will decide to connect via a packet-based IP network or via circuit-based lambda allocations.

2.2.1.1 Wavelength Switching – Hardware Infrastructure

In this architecture the long haul networking backbone would be provided by agile all-optical networking equipment such as ultra long-haul DWDM with integrated optical cross-connects (IOXC's) providing OADM-like functionality with extensions to support degree n (n>2) nodes. Fiber could be user-owned, obtained via an IRU (Irrevocable Right to Use) agreement, or carrier owned; in the latter case the Grid network would contract for the number of wavelengths on each link which they need. Bandwidth would be available in increments of OC-48, OC-192, and eventually OC-768. Optical maintenance and optical fault isolation/recovery would primarily by the responsibility of the EMS and control plane software provided by the optical vendors.

The backbone network would be controlled by a distributed control plane using GMPLS or similar technology, with sub-second connection set-up time. To allow control by the Grid infrastructure, internal network state information needed for routing and capacity management would be advertised by the network to the infrastructure. Connection changes would be controlled by signaling messages (RSVP or CR-LDP in the case of GMPLS) initiated by the Grid infrastructure. When capacity is shared between applications where there is not trust the OVPN mechanism could be used to provide firewalls and prevent unwanted contention for resources.

In the event that all nodes involved in a single Grid application could not be connected to the same optical network, inter-domain connectivity would be provided using an ONNI. The ONNI would also be used to provide interworking between dissimilar technologies or different vendors where necessary.

The strengths of this architecture include:

- The hardware and control technologies exist or are low-risk extensions of current work. Many vendors are at work in this space, as are the standards bodies.
- Little doubt about scalability.
- Compatible commercial networks providing the necessary functionality already have a large footprint in the U.S. and elsewhere.
- Likely to be the lowest cost, fastest, most secure, and most reliable way of transporting vary large (multi terabyte) data sets between two points (or from 1 to N points) on demand.
- Transmission times should have less variance than any of the options using packet or flow switching. This might allow improved scheduling.
- Compatible with both users owned and carrier provided networks, and also hybrids.
- Short-lived Grid relationships can establish and then tear down their optical infrastructure by use of carrier OVPN's.

The issues for this architecture include:

- Not competitive for small (< ?? GB) data transfers.
- Not appropriate for highly interactive applications involving a large number of nodes or for N-to-N multipoint applications (large N).
- Vendors need to be persuaded to make the necessary control plane extensions, and (for use of carrier facilities) carriers need to be persuaded to offer OVPN's at a reasonable price.

2.2.1.2 Wavelength Switching–Software Infrastructure for Network Scheduling

In many circumstances, Grid applications will need to make similar requests for bandwidth at specific times in the future ("future scheduling"). For these applications, there should be a facility for scheduling future allocations of wavelengths without knowledge of the underlying network topology or management protocols. In addition, other applications will need traditional "on-demand" allocations, and both models must be supported.

Grid applications typically need to schedule allocation of computing and data resources from multiple sources. With the advent of wavelength switching, network bandwidth is another such resource that requires scheduling. Services such as the Globus Resource Allocation Manager (GRAM) job scheduler have been developed to coordinate and schedule the computing and data resources needed by Grid applications. Some Grid network allocation proposals are based on DiffServ configuration and do not take into account the optical layers. These services will need to be extended to handle network resources as well. To do so, they will require facilities for scheduled allocation of wavelengths. Simple coordinating and scheduling services may need only high-level facilities. However, services that attempt to optimize network resources will need a richer interface. For example, optimization of schedules with multiple possible paths and replicas will require the ability to schedule individual segments of wavelength paths.

A facility for scheduled allocation of wavelengths on switched optical networks should present a standardized, high-level, network-accessible interface. A natural choice for Grid applications is an Open Grid Service Interface (OGSI). Such interfaces are compliant with the GGF's OGSA specification and conform to widely used Web Services standards (WSDL, SOAP, XML).

In addition to presenting an OGSI-compliant interface, the wavelength service should have a standard way of representing wavelength resources for communicating with clients. Unfortunately no such standard currently exists. For the Grid community, a promising approach would be to extend the XML form of the Resource Specification Language (RSL). This RSL schema is currently used by GRAM to schedule other resources. Adding network extensions to RSL would make it possible to enhance GRAM to handle network resources as well.

GARA is the GGF proposal for General-purpose Architecture for Resource Allocation. GARA architecture provides task scheduling and queuing. Without changing this model, computation or storage might be available when the network is not. The current model is not fully distributed and does not allow remote access; the data must be copied to a local store. For example, GridFTP is a mechanism to copy the data from remote storage to the local storage near the computation. This process is called "data pre-staging." The GARA design schedules the start of computation after the data is available locally. Each task must be completed before the next step is decided.

Most storage and computation exist within a single administrative domain or "points"; a network connection may cross administration boundaries and can be thought of as a "line". A network path has a start point and an end point. This makes network resources different from CPU, and storage resources. CPU and storage resources are isolated and local, while network resources are combined and global. For example, a network path between a CPU and storage may involve a number of small networks.
To solve this problem, a service layer is needed to modify GARA for time synchronization between available resources, including network resources. This network service layer must interact with the optical network discovery facility, find the availability of network resources, and optimize the schedule and availability of the optical network resources. This service layer interfaces with the optical control plan and make the decision to use traditional IP networks or optical networks.

2.2.1.3 Wavelength Switching – Economics

Recent cost structure changes have generated new economic considerations that drive fundamentally different architecture principles for high bandwidth networking.

Inexpensive optical bandwidth - DWDM provides multiple Lambdas, and each one of them accommodates high bandwidth over long distances. Thus, now the transmission cost per data unit is extremely low. This is a departure from the assumptions prevalent for the past 20 years. When the bandwidth is almost free, old assumptions must be reconsidered.

Optical HW costs **-** Depending on the specific Grid application, simplifications and cost reductions may be possible. These include use of dumb CWDM optics rather than agile IOXC or OBS optical networks. For example, a star network with a small number of simple MEMS OXC in the center (and OBGP as protocol), might be adequate in many situations. When all the GRID nodes are close together, there are no trust issues, and the relationships are expected to be long-lasting.

Optical costs - L3 routers can look into packets and make routing decisions, while optical transmissions do not require this functionality. Therefore, the L3 architecture in traditional routing requires substantially more silicon budget. The routing architecture in OC-192 costs about 10x more than the optical transmission equivalent. Specifically, an OC-192 router port costs about 5x as much as the Optical Cross Connect (OXC) equivalent. Furthermore, at intermediate nodes the router ports are in addition to the optical costs.

Connectivity costs - Until recently, an OC-192 connection coast-to-coast has cost about one million dollars. The design of the new optical ultra-long-haul connection reduces the economic fundamentals of big-pipe, long-haul connections.

Last mile costs - Previously, the last-mile connections were expensive and very narrow. Due to recent technology advances and economic restructuring, Optical Metro service has changed the principles of the access. Therefore, we believe that eventually last mile big optical pipes will be affordable for many Grid Computing and data-intensive applications.

Inexpensive LAN bandwidth - 1GE NICs become extremely inexpensive with a new price point of $50 for copper and $100 for optical. 1 GE becomes a commodity for servers and the desktop, while the cost per port of 1Gbs switching port has fallen substantially. With the aggregation of 1 Gbs ports, we believe that this will drive a domino effect into 10GE. With this price point per bit, bandwidth is almost free in the LAN.

Storage costs - Presently, disk prices are very inexpensive. One terabyte currently costs less than $1,000. This affordability has encouraged Grid applications to use larger amounts of data. In particular, 1 Petabyte storage systems cost approximately $2-3 million, which is within the budget of large organizations. With this new economic cost structure and affordability, it is reasonable that many Grid projects will build large data storage.

Computation costs - Many Grid applications require massive amounts of computational power, which is nonetheless inexpensive. The computational power that we have on our desks is larger than a super computer of 10 years ago, and at a price point which is orders of magnitude lower. This phenomenon drives massive amounts of computation at low prices and in many cases require massive amounts of data transfer.

Based on these fundamental cost structure changes in many dimensions, we can expect substantial growth. It looks like Grid applications will be the first to use these new inexpensive infrastructures. The design of optical networking infrastructure for Grid applications must address these challenges in order to allow for predicted growth.

2.2.2 Hybrid Router/Wavelength Switching

This architecture extends the wavelength switching architecture just discussed by adding a layer of IP routers with OC-48/192/768 interfaces between the Grid nodes and the optical network. The GRID node would connect optically to these interfaces, as would the optical network. In addition there might also be connectivity directly from the Grid nodes to the optical network so that the previous architecture could be used where appropriate.

The routers would be capable of providing full line-rate packet switching.

Connectivity between the routers would be dynamically established by use of the UNI or extensions. This could be done under control from the Grid connectivity API, presumably. Packet routing/forwarding from the Grid node, through the router and the optical network, and to the remote Grid node could be controlled by the Grid node by use of GMPLS.

The strengths of this architecture are:

- Full IP packet networking at optical speeds.
- Delay, packet loss, and costs associated with intermediate routers can be minimized by dynamically establishing direct router-router pipes for periods when they are needed.
- Can be used in conjunction with the wavelength switching architecture.
- The necessary networking capabilities are mostly commercially available.

The weaknesses include:

- Uses more resources than wavelength switching if the routers are used for giant file transfers.
- The Grid/router control interface needs definition.
- The addition of another layer will complicate OAM.

2.2.3 Optical Burst Switching

Many in the networking research community believe that optical burst switching (OBS) can meet the needs of the scientific community in the near term (2-3 years). For clarification, the 2-3 years timescale is relevant to early adopters such as Universities and government institutions (usually the same organizations pushing the technology envelope to meet their un-met applications' requirements), pre-standardization. The Grid community seems to fit this definition. Large carrier deployment for the public arena will come later, in practice, since network management and standards need to be in place prior to widespread deployment.

OBS brings together the complementary strengths of optics and electronics [17,18, 19, 20, 21, 22, 23, 24 ,25]. The fundamental premise of OBS is the separation of the control and data planes, and the segregation of functionality within the appropriate domain (electronic or optical). This is accomplished by an end-user, an application, or an OBS edge node initiating a set-up message (control message) to an OBS ingress switch. The ingress switch is typically a commercial off-the-shelf (COTS) optical cross-connect (OXC). The control processor forwards the message along the data transmission path toward the destination. Control messages are processed at each node (requiring OEO conversions); they inform each node of the impending data burst, and initiate switch configurations to accommodate the data burst. The data burst is launched after a small offset delay. Bursts remain in the optical plane end-to-end, and are typically not buffered as they transit the network core. A burst can be defined as a contiguous set of data bytes or packets. This allows for fine-grain multiplexing of data over a single lambda. Bursts incur negligible additional latency. The bursts' content, protocol, bit rate, modulation format, encoding (digital or analog) are completely transparent to the intermediate switches. OBS has the potential of meeting several important objectives: *(i)* high bandwidth, low latency, deterministic transport required for high demand Grid applications; *(ii)* all-optical data transmission with ultra-fast user/application-initiated light path setup; *(iii)* implementable with cost effective COTS optical devices.

2.2.3.1 OBS architectures

There are several major OBS variants. They differ in a number of ways: ***(i)*** how they reserve resources (*e.g.,* 'tell-and-wait', 'tell-and-go'), ***(ii)*** how they schedule and release resources (*e.g.*, 'just-in-time' 'just-enough-time'), ***(iii)*** hardware requirements (*e.g.*, novel switch architectures optimized for OBS, commercial optical switches augmented with OBS network controllers), ***(iv)*** whether bursts are buffered (using optical delay lines or other technologies), ***(v)*** signaling architecture (in-band, out-of-band), ***(vi)*** performance, ***(vii)*** complexity, and ***(viii)*** cost (capital, operational, $/Gbit, *etc*.).

Most OBS research has focused on edge-core, overlay architectures [26, 27, 28]. However, some research is focusing on OBS network interface cards (NICs) for peer-to-peer, distributed networking.

TCP and UDP variants will almost certainly be the predominant transport protocols for data communications. However, some high demand applications might require novel transport protocols which can better take advantage of OBS. OBS allows for bursts of unlimited length, ranging from a few bytes to tens or hundreds of gigabytes. This has led some in the OBS research community to rethink some of the IP protocols to better take advantage of OBS technology – no buffering, ultra-high throughput, ultra-low error rates, etc. Others are investigating simplified constraint-based routing and forwarding algorithms for OBS (e.g., that consider dynamic physical impairments in optical plane when making forwarding decisions [29, 30, 31, 32]) and on methods based on GMPLS.

OBS is deployed in several laboratory test-beds and in at least one metropolitan area dark fiber network test-bed (with a circumference of about 150 Km). Proof-of-concept experiments are underway, and will continue to provide further insights into OBS technology.

Also, there is an effort underway to extend GridFTP to utilize Just In Time (JIT) TAG protocol for possible improvements in performance.

2.2.3.2 OBS and Grid

Many in the scientific research community are of the opinion that today's production, experimental and research networks do not have the capabilities to meet the needs of some of the existing e-science and Grid applications. Many of these applications have requirements of one or more of these constraints: determinism (guaranteed QoS), shared data spaces, real-time multicasting, large transfer of data, and latency requirements that are only achievable through dedicated lambdas, as well as the need to have user/application control of these lambdas. Key for OBS technology is to determine early on, how the technology, protocols, and architecture must be designed to provide solutions to these requirements. This is an opportunistic time within the development stage (pre-standardization) of OBS to incorporate these solutions. Key concepts of interest to the OBS community are as follows:

- Network feedback mechanisms to user
- Status
- Alarms
- Availability and reach
- Creation of hooks to provide policy based control of network behavior
- Policy based routing algorithms – user or carriers decide on how forwarding tables are created.
- Integrating security concerns at both the protocol level as well as control and management plane.
- Incorporating necessary inter-domain information exchange in protocol definitions.
- Providing necessary flexibility in architectures to meet both carrier-owned and user-owned networks.
- Understanding the requirements for both physical layer QoS and application layer QoS and incorporating them into protocol definitions.
- Determine how users will get billed for the Grid network service
- Determine what is meant by Grid SLAs and how the network can provide them.

3. Optical network elements for the Grid

3.1 Optical switching nodes

The network nodes combine edge and core switch functionalities. The edge nodes provide the interface between the electrical domain and optical domain in different layers (i.e. from control layer to physical layer). The core switches, based on the control information configure the switch matrix to route the incoming data to the appropriate output port, and resolve any contention issues that may arise.

A generic structure of an optical switch consists of an input interface, a switching matrix and an output interface. The input interface performs delineation and retrieves control information, encoded in the control packets. The switching block is responsible for the internal routing the wavebands/wavelengths or bursts/packets - depending on technology used - to the appropriate output ports and resolving any collision/contention issues, while the output interface is responsible for control update and any signal conditioning that may be required such as power equalization, wavelength conversion or regeneration.

The optical switch architecture will offer features such as:

- dynamic reconfiguration with high switching speed (<ms, although a more relaxed requirement will be acceptable for very large data transfers and long duration of optical connectivity)
- strictly non-blocking connectivity between input and output ports
- broadcasting and multicasting capabilities in dedicated devices (i.e. near the source or destination)
- capability to address contention issues
- scalability

- protection and restoration capabilities
- minimum performance degradation for all paths and good concatenation performance

In terms of optical switch architectures there are a number of options already proposed in the literature, but the different proposals need to be adjusted to the set of requirements imposed by this new application framework. Especially, waveband and transparent switching are challenging issues. Features such as broadcasting/multicasting are central and need to be addressed by the proposed solution. The broadcast and select architecture may be the obvious choice, but architectures utilizing tunable wavelength converters and wavelength routing devices offer an alternative solution as optical wavelength converters may offer capabilities such as creation of multiple replicas of a single optical signal.

In terms of switching technology, different options are available. Among the main selection criteria would be the switching speed. Depending on the transport format, options may include certain switching technologies such as opto-mechanical or micro-electromechanical system (MEMS) supporting slower switching speeds (typically μsec-msec). For faster switching speeds, more appropriate switch choices are based on electro-optic or SOA technologies supporting ns switching times. These technologies commonly suffer by reduced switch matrix dimensions that can be overcome using multistage architectures. The alternative solution based on the broadcast and select architecture utilizes passive splitters/couplers and tunable filters instead of a switch fabric and in this case the challenging technology choice is associated with the tunable filtering function. A third option in terms of switching functionality is provided through the use of tunable wavelength converters and wavelength routing devices.

3.2 Multicasting in Photonic Network Elements

3.2.1 Motivation for Photonic Multicasting

Multicasting has traditionally found greatest use in multi-site video conferencing, such as on the AccessGrid where each site participating in the conference multicasts or broadcasts several 320x200 video streams to each other. However in the context of Grid computing new uses for extremely high speed multicast are emerging. These are usually data-intensive applications for which there is a real time data producer that needs to be accessed simultaneously by multiple data consumers. For example, in collaborative and interactive Grid visualization applications, extremely high resolution computer graphics (on the order of 6000x3000 pixels and beyond,) that are generated by large visualization clusters (such as the TeraGrid visualization server at Argonne,) need to be simultaneously streamed to multiple collaborating sites (we call this egress multicasting). In another example, data from a remote data source may need to be “cloned” as it arrives at a receiving site and fed into distinct compute clusters to process the data in different ways. Again using large scale data visualization as an example, a single data stream could be used to generate two or more different visual representations of the data using distinct compute clusters running different visualization algorithms (we call this ingress multicasting).

3.2.2 Photonic Multicasting

Strictly speaking photonic multicasting is 1:N broadcasting rather than N:N as in the classical router-based multicast. Hence this 1:N broadcast is often called a Light Tree. A Multicast-capable photonic switch (also called a multicast-capable optical cross connect switch) is a photonic switch that uses optical splitters, also referred to as power splitters, to split a lightpath into N>1 copies of itself. For an N-way split, the signal strength in each split is reduced by at least 1/N. In practice there is always a few dB loss as the light beam passes through the splitter. Hence depending on the size of N and the distance to the termination point, optical amplifiers may need to be incorporated to boost the signal. However optical amplifiers may also amplify any noise in the signal. Rouskas, Ali and others [33, 34, 35] have proposed several possible designs for power-efficient multicast-capable photonic switches and Leigh [36] in collaboration with Glimmerglass Networks, is building a low-cost multicast-capable photonic switch to support collaborative Grid visualization applications.

To support multiple wavelengths, wavelength demultiplexers can be used to split the light into W individual wavelengths which can then be fed into W multicast-capable photonic switch units. The outputs would then reconverge onto a set of W wavelength multiplexers. This solution would support any permutation of photonic multicast and unicast in a non-blocking manner, however its use of W photonic switches with W inputs makes this solution prohibitively expensive to build [33]. Hence simpler and more modularly approaches, such as the one proposed in [36], are needed in the interim until we gain a clearer understanding of practical use-patterns for data-intensive Grid multicast applications.

3.2.3 Controlling Light Trees

It is well known that the problem of Routing and Wavelength Assignment (RWA) in photonic networks is far more difficult than electronic routing. When establishing a lightpath between two endpoints one needs to select a suitable path AND allocate an available wavelength. Dutta [37] shows that optimal solutions for point-to-point RWA cannot be practically found. The Multicast RWA (MC-RWA) problem is even more challenging because, if wavelength conversion is not employed, wavelength assignment must also ensure that same wavelength is used along the entire photonic multicast tree [38].

This will require the development of new control plane algorithms and software in three areas: Firstly the topology and resource discovery algorithms must be extended to include consideration for the availability and location of the multicast switches and their relevant attributes such as maximum splitter fan-out. Secondly multicast extensions to classical RWA algorithms must be made to support both lightpath and lighttree route and wavelength determination. Some excellent initial simulation-based research has already been done by [39, 40, 41, 42, 43, 44]. Thirdly, control plane software needs to be extended to handle setup and teardown of lighttrees. Consequently GMPLS protocols such as CR-LDP and RSVP-TE must be augmented to handle lighttrees.

3.2.4 Application of Photonic Switches as Cluster-interconnects and Ingress Multicasting for Data Replication

The use of photonic switches as interconnects for compute clusters [36] is sparked by the growing trend to move optics closer to the CPU. Savage [45] believes that in 2-5 years optical connections will move between circuit boards inside computers, and in 5-10 years chip-to-chip optical connections will emerge. Today, using multiple optical gigabit network interface cards in each node of a Grid compute cluster, it is possible and potentially advantageous to create dedicated connections between compute nodes using a photonic switching [36]. Since the paths do not go through any electronics, higher speed optical gigabit NICs (at 10G and perhaps 40G) can be used as they become affordable. Furthermore the application-level programmability of the photonic switch allows for the creation of a variety of computing configurations- for example one could connect a collection of compute nodes in several parallel chains or as a tree. This allows applications to reconfigure computing resources to form architectures that are best suited for the particular computing task at hand.

In the photonic cluster-interconnect paradigm, photonic multicasting can be an effective way to take incoming data from a remote source, duplicate it and pass it on to a number of parallel computing units that may be performing different tasks on the same data (for example, generating different types of visualizations at the same time). What this suggests is that the photonic control plane software that is currently focused on assigning wavelengths between remote domains will in the future also need to provide control for a hierarchy of subdomains at a finer granularity level than previously anticipated. That is, RWA for lightpaths and lighttrees will need to be extended to support lambda allocation in the photonic cluster-interconnect paradigm.

3.3. GUNI

3.3.1 Definitions

To facilitate used control and management of the optical network resources, interoperable procedures for signalling and data transport need to be developed between Grid users and the optical transport network. These procedures constitute the Grid User Network Interface (GUNI), the Grid service interface between the Grid user and the optical transport network.

The GUNI functionalities are grouped in the following categories:

- Signalling
    - Bandwidth allocation
    - Automatic light-path setup
        - Automatic neighbour hood discovery
        - Automatic service discovery
    - Fault detection, protection and restoration
    - Security at signalling level
- Transport
    - Traffic classification, grooming, shaping and transmission entity construction

- Data plan security

The signalling mechanism will be responsible for requesting, establishing and maintaining connectivity between Grid users and Grid resources while the data transport mechanism will provide a traffic/bandwidth mapping between the Grid service and the optical transport network.

3.3.2 Functionalities

Bandwidth allocation: will provide a mechanism for allocation of the required bandwidth (i.e. Wavelength or sub-wavelength) for the Grid user/service. Also it would be required to support a lambda time-sharing mechanism to facilitate scheduling of bandwidth over predefined time windows for the Grid users/service. (i.e. lambda time-sharing for efficient/low cost bandwidth utilization). The GUNI signalling also would be required to support ownership policy of bandwidth.

Automatic light-path setup: users can automatically schedule, provision, and set up light-paths across the network. To setup a light-path for a particular Grid service, user must be able to discover and invoke the Grid service (automatic service discovery).

Fault detection, protection and restoration: as Grid services have wide variety of requirements and different level of sensitivity to transport network faults (see section 1.2) the GUNI must be able to support/invoke different protection and restoration signalling schemes.

Traffic classification, grooming, shaping and transmission entity construction: The GUNI performs traffic classification and aggregation under supervision of service control and management plan. At transport layer (physical layer) the GUNI must be able to map the data traffic to a transmission entity (e.g. optical burst). In case of in band signaling the GUNI will provide a mapping mechanism for transmission of control messages (e.g. control wavelength allocation).

Security: the GUNI would be necessary to support a security mechanism for both control plan (signalling) and data plan (transport). (See section 6)

3.3.3 Implementation (technology consideration)

The GUNI implementation will be influenced mainly by the transport network switching paradigm described in section 2.2. For example OBS technology will require a fast tuneable and reconfigurable GUNI to facilitate dynamic bandwidth allocation and lambda sharing between users.

In terms of GUNI technology, fast tuneable laser and high-speed reconfigurable hardware (e.g. fast field programmable gate arrays) are promising technology for realising required functionality at the user interface of the optical enabled Grid network.

4. Optical network control and signaling

It is well known that a separation into a control plane and a data transport plane is necessary for an agile optical network. The control plane includes protocols and mechanisms for discovering, updating available optical resources in the data plane; the mechanisms for disseminate this information; and algorithms for engineering an optimal path between end points. In particular, it requires protocols for routing, protocols for establishing paths between end points, and protocols for configuring and controlling the OXCs.

An architecture that separates control plane functions from transport plane functions been the focus of development activity among standards bodies for a number for years. For example, the IETF has long been involved in developing IP switching methods such as Multi-Protocol Label Switching (MPLS), which provides for signaling protocol that separates forwarding information from IP header information [46, 47, 48, 49, 50]. Forwarding, therefore, can be based on label swapping and various routing options. The IETF is now developing mechanisms, derived on these concepts, for IP-based control planes for optical networks as well as for other IP-optical networking processes [51]. The majority of current standardization activity (IETF, ITU, OIF) is focused on the development of the Generalized Multiprotocol Label Switching protocol (GMPLS), which, conceptually a generalized extension of MPLS, expanding its basic concepts to switching domains. [52, 53, 54, 55, 56, 57, 58].

The Generalized Multiprotocol Label Switching (GMPLS) protocol is an important emerging standard. GMPLS provides for a distinct separation between control and data planes. It also provides for simplified management of both functions, for enhanced signaling capabilities, and for integration with protection and survivability mechanisms. GMPLS can be used for resource discovery, link provisioning, label switched path creation, deletion, and property definition, traffic engineering, routing, channel signaling, and path protection and recovery.

GMPLS has extensions that allow it to interface with traditional devices, including L2 switch devices (e.g., ATM, FR, Ethernet), devices based on time-division multiplexing (e.g., SONET/SDH) and newer devices, based on wavelength switching and fiber (spatial) switches [59] [DAV98]. Therefore, GMPLS allows forwarding decisions to be based on time slots, wavelengths, or ports. Path determination and optimization are based on Labeled Switched Path (LSP) creation. This process gathers the information required to establish a lightpath and determines its characteristics, including descriptive information [60] [LAN00a]. This type of IP control plane provides for extremely high-performance capabilities for a variety of functions, such as optical node identification, service level descriptions (e.g., request characterizations), managing link state data, especially for rapid revisions, allocating and re-allocating resources, establishing and revising optimal lightpath routes, and determining responses to fault conditions.

An optical network needs to provide a user-network-interface UNI [61] to allow client network devices to dynamically request connection through it. For this to work across vendor boundaries, and across administrative boundaries, network nodes such as optical cross-connects must also be signaling one another to carry the dynamic provisioning forward hop-by-hop using a network-to-network interface (NNI).

The Generalized MPLS extension provides a way to use MPLS for provisioning in optical networks. This involves OSPF routing protocol (with optical extensions), either RSVP-TE or CR-LPD for establishing the required path, and some control protocol that allows the control plane to configure the OXCs on demand. Also, addressed are path protection, [62], detecting and locating faults at the IP and optical layers, rapid responses, and restoration [63].

The utility of new signaling methods for Grid applications based on these new methods for dynamic lambda provisioning, within metro areas, and even globally, is being demonstrated in prototype on metro and international test-beds. [64]
If the path is wholly contained within an administrative domain, it is possible to engineer an optimal path with GMPLS. However, if the path traverses multiple administrative domains, more complicated negotiation is necessary. OBGP [65] is needed to bridge the path between end points that are in different domain, each domain may deploy different strategy to allocate its resources.

Optical Border Gateway Protocol (OBGP) is building on the Boarder Gateway Protocol (BGP), the well established inter-autonomous routing system protocol [66]. OBGP is very much oriented toward the Grid concept of enabling applications to discover and utilize all required resources, including light-paths. OBGP was designed in part to motivate the migration from today's centralized networking environments with their complex hierarchies of protocols and control methods to an environment where optical network resources are shared and managed by individual organizations and communities [67]. OBGP is an interdomain lightpath management tool with capabilities for discovery, provisioning, messaging, and adjustment.

In many cases, some higher authority may be involved to sort out various problems concerning policies within a domain.

OBGP can be used in conjunction with GMPLS to interconnect networks and maintaining the lightpath between end-to-end connections. OBGP can also perform some optimisation in term of dynamically selecting autonomous domains and therefore improving the performance of Grid.

The combination of GMPLS, OBGP and/or other multi-domain protocols under evaluation will enable control of optical nodes, peer-to-peer connections, secure data exchange and QoS required by the Grid.

4.1 Optical Grid networks serving well defined scientific communities and/or high volume users

In a dedicated optical Grid network where high volume data transfers between well known users and or sites are the major application there are 2 approaches how an optical network could be deployed:

(a) A shared optical "cloud" with rapid switching of lambdas between users (OBS, GMPLS, ASON)
(b) A fixed optical point to point (partial) mesh between users with slow "automatic fiber patch panel" switching (OBGP)

The advantage of the first approach is the efficient use of a potential costly optical infrastructure. The disadvantage is a complex management system and signaling systems are required e.g. OBS, ASON, and GMPLS. This could be further complicated by the need to provide bandwidth on demand across multiple domains.

The second approach allows for end users to acquire customer owned facilities to common interconnection points and have bandwidth available all the time without signaling other than that is required for an automatic fiber patch panel. The assumption is that topological changes are rare and infrequent. The disadvantage is that this architecture will not scale to a large anonymous community of unknown users at unknown locations. It also makes inefficient use of the bandwidth as the optical links are nailed up for long periods of time. At a given site users may have to schedule or reserve access to the nailed up optical link.

As always a large determining factor is cost. In some cases nailed up optical links with a high number of big files transfer, as opposed to deploying a complex optical network with on demand bandwidth may in fact be cheaper for a small community of users who need a lambda Grid. If there is a large community of users, where only a small unknown subset needs high bandwidth at any given time then an optical cloud using GMPLS, ASON is probably a better solution. As well if a common carrier deploying an optical network to meet its traffic engineering need as well as providing optical VPNs, can in theory, have a cost equation that tilts towards to the optical managed cloud. On the other hand, the cost of dark fiber and customer owned optical networks and/or wavelengths is continuing to decline.

4.2 Access issues

There are 2 possible approaches for connecting to an optical network:
(a) Campus aggregation at a border optical switch or border optical router
(b) Server to server using a dedicated lambda or sub-lambdas with a layer 1 or layer 2 VLAN

With the border aggregation architecture there are two possible approaches of how data is routed:
(a) Automatic detection of large flows and the setup of an end to end VPN
(b) Application signalling for the rapid setup of an end to end VPN

The rapid setup of VPNs across multiple domains still remains a significant challenge.

With server to server applications layer 1 or layer 2 VLANs/VPNs lambdas or sub lambdas would be required. The advantage of server to server based connections is the possibility of bypassing campus firewalls and congested campus networks.

There are 5 ways of setting up VLAN/VPNs:
(a) Protocol transparent lambdas (G.709)
(b) SONET/SDH STS channels (with virtual concatenation)
(c) Generic MPLS tunnel
(d) Ethernet 802.1 p/q
(e) IP VPN tunnel

VLAN/VPN architecture that use (a) and (b) can support both campus aggregation and/or server to server connection. Generic MPLS tunnels in theory could support both type of access - but few campuses have plans for extending MPLS across the campus. Ethernet and IP VPNs/VLANs imply a campus aggregation switch.

4.3 Framing protocols

The choice of framing protocols will be dependent on the choice of VLAN/VPN technology that is chosen.

Transparent lambdas allow for the greatest flexibility in terms of choice framing protocol, MTU, etc. They will also be suited for up coming future protocols such as RDMA

SONET/SDH allows for STS VLANs supporting Ethernet framing or POS (also maybe FiberChannel??)

Ethernet 802.1 p/q requires an aggregation switch and limits framing to Ethernet

IP VPN tunnels requires IP framed traffic

5. Optical Networks as Grid service environment

Optical networks can be viewed as essential building blocks for a connectivity infrastructure for service architectures including the Open Grid Service Architecture (OGSA) [68], or as "network resources" to be offered as services to the Grid like any other resources such as processing and storage devices.

This section offers some definitions of a Grid service, explores how optical network resources can be created and encapsulated as a Grid service.

5.1 Grid Services

Grid services are self-contained, self-describing applications that can be published, located, and invoked over an internet. Grid services can perform a range of functions, from simple resource requests to complicated business or scientific procedures. Once a Grid service component is deployed, other Grid services can discover and invoke the published service via its interface. A Grid service must also possess three additional properties. First, it must be an instance of a service implementation of some service type. Second, it must have a Grid Services Handle (GSH), which might be the Web Service Description Language (WSDL) document (or some other representations) for the service instance. Third, each Grid Service instance must implement a port called "GridService" which has three operations:

FindServiceData. This operation allows a client to discover more information about the service's state, execution environment and other details that are not available in the GSR.

Destroy. This operation allows an authorized client to terminate the service instance.

S*etTerminationTime*. This operation allows the lifetime of a service to be set

OGSA defines the semantics of a Grid service instance including service instance creation, naming, lifetime management and communication protocols. The creation of a new Grid service instance involves the creation of a new process in the hosting environment, which has the primary responsibility for ensuring that the services it supports adhere to defined Grid service semantics.

5.2 Optical Network Resources

If optical networks are considered as network resources to be shared among virtual organizations one needs to specify exactly what are meant by optical network resources, how to encapsulate these resources into services, how to manage these services.

So what would be a meaningful optical network resource that could be offered at a level most useful to an application? In optical networks, possible resources may include an optical cross connects (OXCs) or generally photonic switching devices (i.e. OBS, OPS), a fiber, a wavelength, a waveband, a generalized label, an optical timeslot, an interface, etc. [69]. These and other choices are normally coupled tightly with the intended application. For the purpose of this document, let’s assume some typical network resources: 1) an optical path with a specific bandwidth requirement across two end points and 2) an optical tree with adequate bandwidth across multiple end points in a multicast situation. To be more specific, one may specify QoS constraints on these paths in terms reliability, delay, jitter, protection, alternative path, or even the exact time and duration for which the resource is needed.

Whatever the choices, it can be seen that an optical resources (as defined) will involve two or more network entities, not wholly contained within a network element. This makes the situation a bit more complicated since any reservation and allocation will involve cooperation of more than one network elements. Other Grid services such as processors, storage devices can be simply controlled and allocated (booked, reserved) by one network element without external constraints.

The situation is further complicated when a desired path traverses multiple heterogeneous administrative domain. Local management of the resource at the originating end of the path may not able to negotiate a path without involvement of some higher authority. Issues involved security and cooperation among different administrative domains have to be considered.

5.3 Optical network as a Grid service

OGSA framework demands that a service has to be represented as self contained, modular entity that can be discovered, registered, monitored, instantiated, created and destroyed with some form of life cycle management.

For this to be conformed to the OGSA, an optical network resource has to be wrapped up into an object that has name, characteristics, and facilities for invocation, monitoring. It is thus necessary for a Local Grid Resource Allocation and Management (LGRAM) [70], situated above the Optical Control Plane, to manage its resources. The LGRAM is responsible to create as well as manage the required optical resources using GMPLS or other form of signaling.

To assist the messaging, discovery, instance creation and lifetime management functions required by a Grid service, the OGSA standard Grid Service ports include
NotificationSource and NotificationSink ports. This service provides a simple publish-subscribe system.
HandleMap. This service provides the mapping between the Grid Service Handle and the current Grid Service Reference.
Registry. This service allows service instance to be bound to a registry. The Registry port also allows services to be unregistered.
Factory. A Factory service is a service that can be used to create instances of other services. In Grid applications the factory service can create instances of transient application services.
A Grid service hence always requires a hosting environment to provide supplementary functions including Global Information Services, Grid Security Infrastructure, and to ensuring that the services it supports adhere to defined Grid service semantics.

5.4 Grid Resource Management issues

Few people in the Grid community thought of network as a resource in the same way as processing or storage. They are inclined either to view the network as a bottleneck or, if bandwidth resources are plentiful, to take the network for granted without the need for reserving options for their applications. This view was reflected in the early architecture of the Globus Resource Allocation Management (GRAM) architecture. Advances have been made, however, the issues of managing “network resources” is far from being solved. This section takes a look at various issues concerning the encapsulation and allocation of optical network resources.

In the network community, network resources are often statically allocated, or allocated on demand. In the Grid community, resources are often reserved, allocated, and even scheduled. If only allocated on demand, existing reservation techniques may be adequate.

In other cases, co-reservation and co-allocation may be necessary to cope with staging in a heterogeneous environment [71]. In case of scheduling resources, additional protocols involving cooperation are required to make sure a scheduled plan is acceptable among all participants.

A Resource Management Architecture for Metacomputing Systems [72] was proposed to deal with the co-allocation problem where applications have resource requirements that can be satisfied only by using resources simultaneously at several sites. In this architecture, an extensible resource specification language (RSL) is used to communicate requests for resources between components: from applications to resource brokers, resource co-allocators and resource managers. A Monitoring and Discovery Service (MDS) is a service that houses information pertaining to the potential computing resources, their specifications, and their current availability. Resource brokers are responsible for taking high-level RSL specifications and transforming them into more concrete specifications (ground requests) that can be passed to a co-allocator which is responsible for coordinating the allocation and management of resources at multiple sites. Resource co-allocators break a multirequest that involves resources at multiple sites, into its constituent elements and pass each component to the appropriate resource manager. Each resource manager (GRAM, Globus Resource Allocation Manager) in the system is responsible for taking a RSL request and translating it into operations in the local, site-specific resource management system.

The realization of end-to-end quality of service (QoS) guaranteed in emerging network-based applications requires mechanisms that support first dynamic discovery and then advance or immediate reservation of resources that will often be heterogeneous in type and implementation and independently controlled and administered.

The GRAM architecture does not address the issue of advance reservations and heterogeneous resource types. The absence of advance reservations means that we cannot ensure that a resource can provide a requested QoS when require. The lack of support for network, disk, ands other resource types makes it impossible to provide end-to-end QoS guarantees when an application involves more than just computation.

To address this problem, the Globus Architecture for Reservation and Allocation (GARA) was proposed [71]. By splitting reservation from allocation, GARA enables advance reservation of resources, which can be critical to application success if a required resource is in high demand. Also, if reservation is cheaper than allocation, lighter-weight resource reservation strategies can be employed rather than expensive and immediate allocation of actual resources.

The most challenging issue in the management of resources in Grid environments is the scheduling of dynamic Grid services where negotiation may be required to adapt application requirements to resource availability, particularly when requirements and resource characteristics change during execution. The deployment of such environments requires the ability to create Grid services and adjust their policies and behavior based on organizational goals and application requirement.

OGSI-Agreement negotiation model was proposed [73] allowing management in these environments where centralized control is impossible. The OGSI-Agreement model uses agreement negotiation to capture the notion of dynamically adjusting policies that affect the service environment without necessarily exposing the details necessary to enact or enforce the policies.

Negotiation is a stateful dialogue. It may be as simple as a single request message being allowed (or not) by policy, or it may involve a complicated scenarios where the policies and intermediate commitments of the two parties are revealed piece by piece over a long sequence of message exchanges, resulting in an agreement capturing an intersection in their policies.

As a result of the negotiation process, an Agreement service may be created. An Agreement service should always relate to a "delivered service "behavior which may involve a Grid service. It may relate to an "existing service" known by the agreement provider. In this case the Agreement represents an aspect of policy affecting the behavior of that service. Alternatively, the Agreement service may relate to a "new service" which will be created due to the agreement. In this case, the Agreement service may represents, on the part of the agreement provider, both a commitment to create the new service and policy affecting the behavior of the new service.

It is believed that OGSI-Agreement model presents a very useful framework for effective scheduling of Grid resources. Adopting this model of cooperating agreement is essential in providing interoperability in the Grid heterogeneous environments. However, it is equally important to ensure that an OGSI-Agreement model remains simple and realistic. It has the potential of evolving into an over-complicated model which cannot be deployed effectively.

6. Security

6.1 Threats

Active/passive attacks are grouped in the following three categories (A, B, and C) according to their target.

A. Attacks on out-of-band user-network and network-network signaling:
A.1) Acquire confidential data and identities by snooping traffic
A.2) Modify packets (e.g., a downgrade attack to lessen security agreements)
A.3) Inject new packets
A.4) Man-in-the-middle attack at setup time, with user or network impersonation, and hijacking of traffic
A.5) Mount DoS attack against legitimate signaling traffic
A.6) Disrupt the security negotiation process
A.7) Traffic analysis
A.8) Covert channels

A.7 and A.8 are the most speculative ones (no evidence of grid communities with sensitivity to these types of attack).

B. Attacks on in-band user-network signaling (as seen in flavors of OBS):
B.1) A malicious user can wreak havoc by abusing semantics (e.g., get authorization to proceed with "tell and wait" and use "tell and go" instead). A stratum of strong up-front authentication/authorization is required, and out-of-band solutions make the most sense (e.g. due to heavy-duty crypto processing and database handling). This is vulnerable to the threats identified in out-of-band user-network signaling (see [A]).
B.2) Past this barrier, a user must be trusted to use the lightweight in-band signaling in a sensible way. Therefore, "door-rattling" attacks on the control processor (e.g., by announcing silly burst sizes) are ruled out.

C. Attacks on the data plane (assuming that L3 and above data are already end-to-end authenticated, with integrity, confidentiality, and replay prevention):
C.1) Forging of logical capabilities granting access to lightpaths (hence circumventing signaling)
C.2) Violation of non-TDM sharing rules (e.g., OBS) within a lightpath

6.2 Strengths

When compared to packet switching, the circuit-oriented technologies described in this document show noteworthy points of strength in security. Chiefly, a circuit is a practical way to limit trust relationships to a small, tractable set of users (e.g., the two peers in a dedicated lightpath, or a small set of peers in an OBS setup).

Conversely, in a packet-switched network a user must trust any and all of its users to "play nice" and execute their end-to-end protocols in the IETF sanctioned terms only. For instance, experimental, faulty, or outright malicious TCP implementations[74] can dramatically alter fairness, often reaching the extreme case of (D)DoS attack. Access capacity, QoS, and policy boundaries are known to lessen this exposure, though in practice these boundaries are soft-boundaries when compared to a circuit's “hard” boundaries. As a case in point, research testbeds can be easy exploitation targets due to the mix of experimentation, high access capacity, and non-commercial-grade QoS/policy stipulations.

Circuit-oriented technologies are seen having the following strengths:
a) isolation and non-interference among users
b) compartmentalization in the face of failure or compromise
c) friendly end-to-end protocol experimentation with a limited trust base
d) traceable and accountable access (no need for firewalls)
e) hitless circuit setup/teardown

In scenarios with out-of-band signaling, the separation of signaling vs. data concerns has merits as well as inherent risks. The key strength is that security measures can now be designed to custom fit signaling channel and data channels. That is, the a-priori knowledge of their two different traffic patterns can lead to a security schema with tighter protection. A key risk is that the signaling plane represents a manifest and highly rewarding target to attackers. It is easy to imagine that an intrusion into signaling and control planes can generate catastrophic failures. While optical networks typically use physically isolated networks for the signaling/control functions, it is also the case that researchers are advocating greater and more direct control of the network (with potential vulnerabilities at the testbed level at least). The theft of optical resources via circumvention of the whole signaling phase is another noteworthy risk area.

The scoreboard of strengths vs. risks suggests that Grid experimentation can proceed on optical networks with a remarkably good security potential, starting with the early research testbeds.

6.3 Design options

Out-of-band user-network and network-network signaling are typically IP-based. Network-level security (e.g., IPsec [75]) can thwart the attacks in A.1 to A.4. [76] describes a possible implementation. The use of key exchange protocols (e.g., IKE [77]) is highly recommended (see [76]) . Access-control limits exposures to A.5 and A.6. Dummy traffic and/or link-level security are canonical defenses against traffic analysis (A.6).

With regard to in-band signaling and C.2 attacks, rate-control fixtures can force traffic to fit into agreed-upon envelopes. This aptly complements the trust given to the (small) set of users sharing a lightpath via, say, OBS techniques (e.g., a user can still be faulty).

The C.1 attack requires that the capability to a lightpath (i.e., the outcome of successful signaling to the network) be closely guarded. In optically-attached systems, the point of ingress to a lightpath is integral part of the TCB, and standard OS security considerations apply. In setups where traffic is groomed on lightpaths one or more hops away, an attacker can infer that, for instance, VLAN IDs correspond to lightpaths, and sweep the VLAN ID space with spurious traffic until a lightpath is found. These setups can be secured by protecting the access ramps to lightpaths from traffic injection, or using on-the-wire IDs stronger than VLAN IDs.

OVPNs [78] are an emerging solution to increase the granularity of a circuit's capacity (e.g., to scale a circuit in STS-1 increments). Additionally, they can restrict connectivity and isolate domains of addressing/routing. As such, they are a powerful step towards securing these circuit-oriented optical technologies.

When optical resources are exposed as an OGSI-based service, the above-mentioned security techniques can be thought of as operating in the back-end of the service. The front-end of the service should conform to the GGF’s Grid Security Infrastructure, enabling a seamless integration of the optical resource with other resources.

7. Authors Information

1. Dimitra Simeonidou (Editor), Photonic Networks Laboratory, University of Essex,UK, dsimeo@essex.ac.uk
2. Bill St. Arnaud, CANARIE, Canada, bill.st.arnaud@canarie.ca
3. Micah Beck, Logistical Computing and Internetworking, Computer Science University of Tennessee, USA, mbeck@cs.utk.edu
4. Bela Berde, Ph.D, Alcatel CIT, Research & Innovation, Bela.Berde@alcatel.fr
5. Freek Dijkstra, Universiteit Van Amsterdam, freek@macfreek.nl
6. Doan B. Hoang, Department of Computer Systems, University of Technology, Sydney, dhoang@it.uts.edu.au
7. Gigi Karmous-Edwards, Advanced Network Research, MCNC Research and Development Institute, USA, gigi@anr.mcnc.org
8. Tal Lavian, Nortel Networks, tlavian@nortelnetworks.com
9. Jason Leigh, Electronic Visualization Lab, University of Illinois at Chicago, spiff@evl.uic.edu
10. Joe Mambretti, International Center for Advanced Internet Research, Northwestern University, Illinois,USA, j-mambretti@northwestern.edu
11. Reza Nejabati, Photonic Networks Laboratory, University of Essex, ,UK, rnejab@essex.ac.uk
12. John Strand, AT&T Transport Network Evolution Dept, jls@research.att.com
13. Franco Travostino, Nortel Networks, travos@nortelnetworks.com

8. Intellectual Property Statement

The GGF takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights. Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the GGF Secretariat. The GGF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights which may cover technology that may be required to practice this recommendation. Please address the information to the GGF Executive Director (see contacts information at GGF website).

9. Full Copyright Notice

10. References

[1] Information about the Large Hadron Collider at CERN: lhc-new-homepage.web.cern.ch

[2] Information about the BarBar experiment: www.slac.stanford.edu/BFROOT/

[3] World Economic Forum, New York 2001, Digital Divide Report

[4] Telegeography Inc, Terrestrial bandwidth 2002

[5] “The GRID, Blueprint for a new computing infrastructure”, Ian Foster and Carl

[6] Training presentation: www.globus.org

[7] "Revolutionizing Science And Engineering Through Cyberinfrastructure", Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, January 2003, http://www.cise.nsf.gov/evnt/reports/atkins_annc_020303.htm

[8] "Research Networking in Europe - Striving for global leadership", European Commission, 15 sep 2002, http://www.cordis.lu/ist/rn/rn-brochure.htm

[9] Information about CERN, The CERN Grid Deployment group: http://it-div-gd.web.cern.ch/it-div-gd/

[10] Ralph Spencer, Steve Parsley and Richard Hughes-Jones “The resilience of e-VLBI data to packet loss”, 2nd eVLBI workshop, 15-16 May 2003, Netherlands

[11] Allcock, W., A. Chervenak, I. Foster, C. Kesselman, C. Salisbury, and S. Tuecke, “The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets”, Network and Computer Applications, 2002.

[12] David Levine, “Grid Computing for the Online Video Game Industry “, GlobusWorld January 14, 2003

[13] Foster, I., Kesselman, C. and Tuecke, S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. International Journal of High Performance Computing Applications, 15 (3). 200-222. 2001. www.globus.org/research/papers/anatomy.pdf.

[14] www.canarie.ca

[15] M. J. O'Mahony, D. Simeonidou, D. K. Hunter, A. Tzanakaki, " The application of optical packet switching in future communication networks", IEEE Communications Magazine, pp. 128-135, March'01

[16] J. He, D. Simeonidou, "Flow routing and its performance analysis in optical IP networks", Photonic Network Communications, Vol 3, pp 49-62 (Special issue for IP over WDM), 2001

[17] J. S. Turner. Terabit burst switching. Journal of High Speed Networks, 8(1): 3{16, January 1999

[18] C. Qiao and M. Yoo. Choices, features, and issues in optical burst switching. Optical Networks, 1(2): 36{44, April 2000.

[19] J. Y. Wei and R. I. McFarland. Just-in-time signaling for WDM optical burst switching networks. Journal of Lightwave Technology, 18(12):2019{2037, December 2000.

[20] Y. Xiong, M. Vandenhoute, and H.C. Cankaya. Control architecture in optical burst-switched WDM networks. IEEE Journal on Selected Areas in Communications, 18(10):1838{1851, October 2000.

[21] Ikegami, Tetsuhiko. "WDM Devices, State of the Art," In Photonic Networks, Giancarlo Prati (editor), Springer Verlag 1997.

[22] C. Qiao and M. Yoo. Optical burst switching (OBS)-A new paradigm for an optical Internet. Journal of High Speed Networks, 8(1):69{84, January 1999.

[23] Baldine, G. N. Rouskas, H. G. Perros, and D. Stevenson. JumpStart: A just-in-time signaling architecture for WDM burst-switched networks. IEEE Communications, 40(2):82{89, February 2002.

[24] Kozlovski E., M. Duser, I. De Miguel, P. Bayvel, "Analysis of burst scheduling for dynamic wavelength assignment in optical burst-switched networks", IEEE, Proc. LEOS'01, vol. 1, 2001.

[25] Dolzer K. "Assured horizon - an efficient framework for service differentiation in optical burst switched networks." Proc. SPIE OptiComm. 2002. 1 page. (Proc. SPIE Vol 4874.)

[26] http://www.ind.uni-stuttgart.de/~gauger/BurstSwitching.html#Publications

[27] http://www.utdallas.edu/~vinod/obs.html

[28] http://www.cs.buffalo.edu/~yangchen/OBS_Pub_year.html

[29] Dimitri Papadimitriou, Denis Penninckx, " Physical Routing Impairments in Wavelength-switched Optical networks", Business Briefing: Global Optical Communications, 2002.

[30] John Strand, Angela Chiu and Robert Tkach, "Issues for Routing in the Optical Layer," IEEE Communications Magazine, February 2001.

[31] Daniel Blumenthal, "Performance Monitoring in Photonic Transport Networks", Bussiness Breifing: Global Photonics Applications and Technology 2000.

[32] Byrav Ramamurthy, Debasish Datta, Helena Feng, Jonathan P. Heritage, Biswanath Mukherjee, "Impact of Transmission Impairments on the Teletraffic Performance of Wavelength-Routed Optical networks", IEEE/OSA Journal of Lightwave Technology Oct '99.

[33] G. N. Rouskas, "Optical Layer Multicast: Rationale, Building Blocks, and Challenges," IEEE Network, Jan/Feb 2003, pp. 60-65.

[34] M. Ali, J. Deogun, "Power-efficient Design of Multicast Wavelength-Routed Networks", IEEE JSAC, vol. 18, no. 10, 2000, pp. 1852-1862.

[35] M. Ali, J. Deogun, "Allocation of Splitting Nodes in Wavelength-routed Networks," Photonic Net. Comm., vol. 2, no. 3, Aug. 2000, pp. 245-263.

[36] Leigh et al. An Experimental OptIPuter Architecture for Data-Intensive Collaborative Visualization, the 3rd Workshop on Advanced Collaborative Environments (in conjunction with the High Performance Distributed Computing Conference), Seattle, Washington, June 22, 2003. [http://www-unix.mcs.anl.gov/fl/events/wace2003/index.html]

[37] R. Dutta, G. N. Rouskas, "A Survey of Virtual Topology Design Algorithms for Wavelength Routed Optical Networks," Opt. Net., vol. 1, no. 1, Jan 2000, pp.73-89.

[38] G. N. Rouskas, "Light-Tree Routing Under Optical Layer Power Budget Constraints," Proc. 17th IEEE Comp. Comm. Wksp., Oct. 14-16, 2002.

[39] X.H. Jia et al., "Optimization of Wavelength Assignment for QoS Milticast in WDM Networks," IEEE Trans. Comm., vol. 49, no. 2, Feb. 2001, pp. 341-350.

[40] G. Sahin, M. Azizoglu, “Milticast Routing and Wavelength Assignment in Wide-Area Networks.” Proc. SPIE, vol. 3531, Nov. 1998, pp. 196-208.

[41] A. E. Kamal, A. K. Al-Yatama, “Blocking Probabilities in Circuit-switched Wavelength Division Multiplexing Networks Under Multicast Service,” Perf. Eval., vol. 47, no. 1, 2000, pp.43-71.

[42] S. Ramesh, G. N. Rouskas, H. G. Perros, “Computing Call Blocking Probabilities in Multi-class Wavelength Routing Networks with Multicast Traffic,” IEEE JSAC, vol. 20, no.1, Jan. 2002, pp. 89-96.

[43] K.D., Wu, J. C., Wu, C.S. Yang, “Multicast Routing with Power Consideration in Sparce Splitting WDM Networks,” Proc. IEEE ICC, 2001, pp. 513-517.

[44] X. Zhang, J. Y. Wei, C. Qiao, “Constrained Multicast Routing in WDM Networks with Sparce Light Splitting,” J. Lightwave Tech., vol. 18, no. 12, Dec. 2000, pp. 1917-1927.

[45] Savage, N. “Linking with Light”, IEEE Spectrum, pp. 32-36, Aug, 2002.

[46] R. Callon, et al. 1999. A Framework for Multiprotocol Label Switching. ID: draft-ietf-mpls-framework-03.txt

[47] E. Rosen, et al. 1999. Multiprotocol Label Switching Architecture. ID: draft-ietf-mpls-arch-05.txt

[48] E. Rosen, A. Viswannthan, R. Callon, “Multiprotocol Label Switching Architecture,” IETF RFC - 3031, January 2001.

[49] T. Nadeau, C. Srinivasan, A. Farrel "Multiprotocol Label Switching (MPLS) Management Overview", July 23, 2003

[50] D. Awduche and Y. Rekhter, “Multi-Protocol Lambda Switching: Combining MPLS Traffic Engineering Control with Optical Crossconnects,” IEEE Communications Magazine, March 2001, pp. 111-116.

[51] B. Rajagopalan, J. Luciani, D. Awduche, “IP over Optical Networks: A Framework,” draft-ietf-ipo-framework-03.txt

[52] E. Mannie, et al, GMPLS Extensions for SONET and SDH Control draft-ietf-ccamp-gmpls-sonet-sdh-01.txt

[53] E. Mannie GMPLS Signaling Extension to Control the Conversion between Contiguous and Virtual Concatenation for SONET and SDH draft-mannie-ccamp-gmpls-concatenation-conversion-00.txt]

[54] E. Mannie, et al, Generalized Multi-Protocol Label Switching (GMPLS) Architecture draft-ietf-ccamp-gmpls-architecture-00.txt

[55] A. Bellato G.709 Optical Transport Networks GMPLS Control Framework draft-bellato-ccamp-g709-framework-00.txt

[56] A. Bellato GMPLS Signaling Extensions for G.709 Optical Transport Networks Control draft-fontana-ccamp-gmpls-g709-00.txt

[57] O. Aboul-Magd A Framework for Generalized Multi-Protocol Label Switching (GMPLS) draft-many-ccamp-gmpls-framework-00.txt

[58] G.8080/Y.1304, Architecture for the Automatically Switched Optical Network (ASON), ITU-T

[59] B. Davie, P. Doolan, and Y. Rekhter. 1998. Switching in IP Networks: IP Switching, Tag Switching, and Related Technologies. The Morgan Kaufmann Series in Networking. New York: Academic Press.

[60] J. Lang, et al, Generalized MPLS Recovery Mechanisms draft-lang-ccamp-recovery-01.txt, draft-mannie-ccamp-gmpls-concatenation-conversion-00.txt

[61] User-Network Interface 1.0 Signaling specification, Implementation Agreement OIF-UNI-01.0, October 2001, http://www.oiforum.com/public/documents/OIF-UNI-01.0.pdf

[62] Lang, et al, Link Management Protocol (LMP) draft-ietf-ccamp-lmp-00.txt

[63] R. Doverspike and J. Yates "Challenges for MPLS in Optical Network Restoration," IEEE Communications Mag, Feb 2001, pp. 89-96

[64] J. Mambretti, J. Weinberger, J. Chen, E. Bacon, F. Yeh, D. Lillethun, B. Grossman, Y. Gu, M. Mazzuco, "The Photonic TeraStream: Enabling Next Generation Applications Through Intelligent Optical Networking at iGrid 2002," Journal of Future Computer Systems, Elsevier Press, August 2003, pp. 897-908.

[65]"Optical BGP networks", Canarie OBGP, Internet draft: http://obgp.canet4.net/

[66] Y. Rekhter "A Border Gateway Protocol 4 (BGP-4)", IETF March, 2003 The Border Gateway Protocol, IETF [RFC1518, RFC1519].

[67] M. Blanchet, F. Parent, B. St Arnaud "Optical BGP (OBGP): InterAS lightpath provisioning draft-parent-obgp-01.txt March 2001

[68] Foster, I., Kesselman, C., Nick, J., Tuecke, S., "The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration," Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002

[69] RFC 3036, LMP Specification.

[70] Foster, I., Fidler, M., Roy, A., V, Sander, Winkler, L., "End-to-End Quality of Service for High-end Applications." *Computer Communications, Special Issue on Network Support for Grid Computing*, 2002

[71] Foster, I., Hesselman, C, Lee, C., Lindel, R., Nahrstedt, K, and Roy, A., "A Distributed resource management architecture that supports advance reservation and co-allocation. In Proceedings of the International Workshop on Quality of Service, pp. 27-36, 1999

[72] Czajkowski, K., Foster, I., Karonis, N., Kesselman, C., Martin, S., Smith, W, and Tuecke, S., A "A Resource Management Architecture for metacomputing systems," *In the 4th Workshop on Job Scheduling Strategies for Parallel Processing, pp. 62-82. Springer-Verlag LNCS 1459, 1988.*

[73]Czajkowski, K., Dan, A., Rofrano, J., Tuecke, S., and Xu, M., "Agrement-based Grid Servic Management (OGSI-Agrement)," GWD-R draft-ggf-czajkowski-agrement-00. June 2003.

[74] Stefan Savage, Neal Cardwell, David Wetherall and Tom Anderson, TCP Congestion Control with a Misbehaving Receiver, ACM Computer Communications Review, 29(5):71-78, October 1999.

[75] RFC 2401, The Internet Engineering Task Force

[76] The Optical Internetworking Forum, Security Extensions for UNI and NNI, http://www.oiforum.com/public/documents/Security-IA.pdf

[77] RFC 2409, The Internet Engineering Task Force

[78] Service Requirements for Optical Virtual Private Networks, Internet Draft, http://www.ietf.org/internet-drafts/draft-ouldbrahim-ppvpn-ovpn-requirements-01.txt