

ESnet On-Demand Secure Circuits and Advance Reservation System

(OSCARS)

Chin Guok

Energy Sciences Network, Lawrence Berkeley National Laboratory

chin@es.net

Table of Contents

Project Summary	2
1 The Problem	3
2 Background and Goals	4
2.1 <i>Establishing Service Offerings in the Network</i>	6
3 Technical Approach.....	7
3.1 <i>QoS, MPLS, and RSVP</i>	8
3.2 <i>Reservation Manager (RM)</i>	8
3.3 <i>Monitoring</i>	10
3.4 <i>Traffic Shaping</i>	11
4 Implementation	11
4.1 <i>QoS</i>	11
4.2 <i>MPLS and RSVP</i>	11
4.3 <i>Resource Manager</i>	12
4.4 <i>Monitoring</i>	14
5 Interoperability	14
6 Work Schedule	14
6.1 <i>Year 1</i>	14
6.2 <i>Year 2</i>	14
6.3 <i>Year 3</i>	15
7 References.....	16
8 Budget and Justification.....	17
<i>Budget – Year 1</i>	17
<i>Budget – Year 2</i>	17
<i>Budget – Year 3</i>	17
<i>Budget – Total 3 Years</i>	17
9 Other Support of Investigator(s).....	17
10 Biographies of Key Personnel.....	17
11 Facilities	17

Project Summary

With the advent of service sensitive applications (such as remote controlled experiments, time constrained massive data transfers, video-conferencing, etc.), it has become apparent that there is a need to augment the services present in today's ESnet infrastructure.

Two DOE Office of Science workshops in the past two years have clearly identified both science discipline driven network requirements [1] and a roadmap for meeting these requirements [2]. This project begins to address one element of the roadmap: dynamically provisioned, QoS paths.

The focus of the ESnet On-Demand Secure Circuits and Advance Reservation System (OSCARS) is to develop and deploy a prototype service that enables on-demand provisioning of guaranteed bandwidth secure circuits within ESnet. OSCARS will leverage existing (or in development) products, services, and code (both from the industry and academia) to accomplish its goals. OSCARS will utilize the existing DOEGrids certificate infrastructure and modify Virtual Organization Membership Services (VOMS) software to implement its authentication and authorization schemes. The management and operation of end-to-end circuits (using Label Switched Paths (LSPs)) within the network will be supported using Multi-Protocol Label Switching (MPLS) and Resource Reservation Protocol (RSVP). Quality of Service (QoS) will be used to provide bandwidth guarantees.

The project objective is to develop and deploy an intra-domain service that can be used by ESnet attached sites, but that will eventually be able to be used by a bandwidth broker to set up inter-domain QoS paths.

Interoperability with emerging standards, in particular the OASIS's Web Services Resource Framework (WS-RF) [33] and the Global Grid Forum's Open Grid Services Architecture (OGSA), will be a focal point in the implementation of this project.

The research aspects of this project are the investigation of how all of the various elements of the OSCARS service can properly interact with deployed network tools, and how the overall service can co-exist with the production network.

1 The Problem

The DOE Office of Science (OSC) Roadmap Workshop [2] identified three major areas that need to be addressed in order to meet DOE's science discipline requirements. From the Roadmap Workshop report, the areas may be summarized as:

IP production services with 99.9+% reliability will continue to be critical for achieving DOE science. Collaboratory activities, such as remotely running experiments and steering supercomputer applications, are especially dependent on the reliability of these production services. In addition, the scientific work depends on managed production services that are currently provided in addition to commercial IP services.

High-impact network services refer to services required by a subset of production network sites, whose demands may be extremely high at times but for which a considerably less reliable and potentially schedulable service will nevertheless meet the needs. Provision of such services at greater than 99.9% reliability with the ability to match potential peak demands would otherwise be prohibitively expensive. These high-impact services will be designed to move the massive amounts of data from experiments and simulations with multi-terabyte to multi-petabyte datasets. Other network services with a major impact on science, such as high-end remote visualizations, that cannot be provided cost-effectively by the production network will be candidate high-impact network services.

The research network services will exist in a network test bed environment in which R&D specific to Office of Science requirements for networking can be performed in partnership with other network test beds. Commercial IP providers must deal with the problem of "many small data flows" — moving small/ modest amounts of data around for tens or hundreds of millions of customers. Research network providers such as ESnet face the problem of "a few very large data flows" — scientists need to move terabyte/petabyte scale datasets around to tens or hundreds of scientists collaborating on scientific discovery. The problems are quite different, and significant research is required to provide efficient and effective solutions.

This project will address dynamically provisioned QoS paths which were identified as a key high impact network service.

The two research aspects of this project both relate to providing the service within the context of ESnet as a production network.

One research issue for this project is to design the various subsystems that will manage and monitor the OSCARS service so that they can make use of the deployed production tools in the network. These tools are an integral part of the production environment and they must be used to accomplish the tasks and/or monitoring that they address. It is not possible in a production network to just go in and introduce a lot of new tools that may have security vulnerabilities, may require installation of hardware and software in many locations in the network, etc. One possible outcome of this aspect of the investigation may be the discovery that there are missing tools critical to providing an OSCARS type of service. This in itself would be an important result that would have to be addressed.

Another research aspect of the project is to investigate and determine mitigations for the risks that are engendered by introducing the ability to dynamically allocate capacity in the network — to the exclusion of normal priority traffic — for the exclusive use of the service. Circumstances

under which this might lead to unintended and adverse consequences for the network need to be identified and carefully characterized so that checks may be designed to ensure that this circumstance never occurs.

2 Background and Goals

ESnet (www.es.net) is engineered and operated by the ESnet networking staff located at Lawrence Berkeley National Laboratory in Berkeley, California.

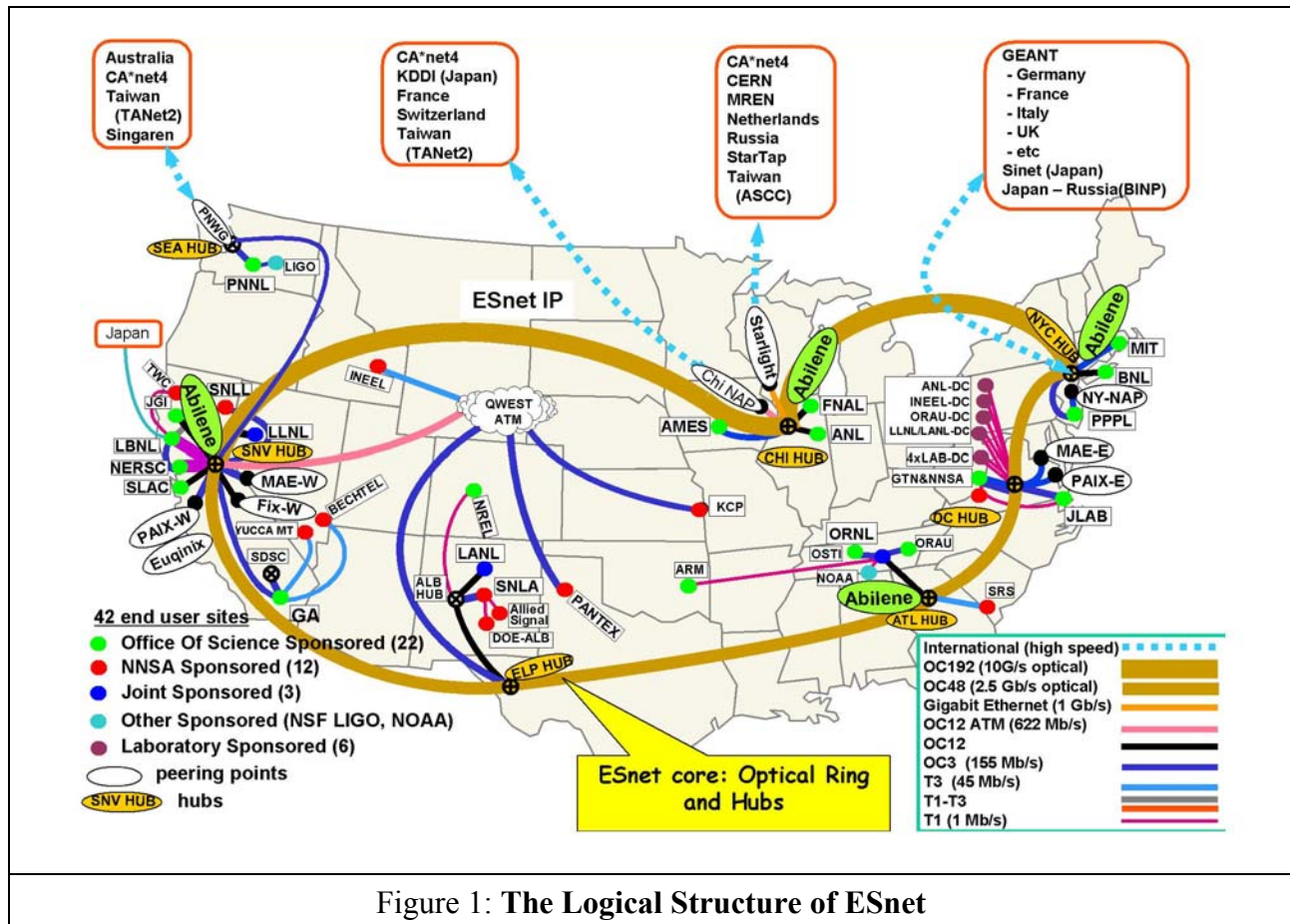


Figure 1: The Logical Structure of ESnet

ESnet is a nation wide network that serves approximately 42 directly connection sites around the country. In addition, there are six core hub sites and two sub-hub sites. There core routers, peering routers, and performance monitoring systems at all hubs. There are ESnet routers at all sites. Overall ESnet manages about 270 routers and systems throughout the network and its operations centers.

The current ESnet architecture (Figure 1) is that of a high-bandwidth (10/2.5 Gb/s) backbone ring around the country, with hubs at strategic locations. The end sites are connected to the hubs in a spoke-like fashion with local loops and last-mile tail circuits that are typically 0.6 or 2.5 Gb/s for the big science Labs.

The network is built from point-to-point circuits are provided by the telecommunications industry and a layer three routing infrastructure designed, built, and operated by ESnet that implements the IP network. Services associated with the network include Overlay Networks, the Domain Name System service that manages the allocation of the IP address space among ESnet

sites, and management of the entire space of routes that are necessary to provide global connectivity to all OSC collaborators and information sources.

The network must also deal with an annual traffic growth of 100% per year while maintaining all of the reliability and services that are needed by the OSC community.

The default characteristic of the Internet today does not provide a user with any service guarantees. There is neither the assurance that a packet will be delivered to its destination, nor any transport predictability (such as latency and jitter) when a packet is in transit. The concept of dynamic provisioning and on-demand bandwidth assignment will go far in assisting high-impact science applications [2] to obtain predictable high bandwidth and/or low latency.

The objective is to deploy a model for an intra-ESnet service supporting on-demand reservations of bandwidth guaranteed paths that is available to ESnet sites via connections to the ESnet border router.

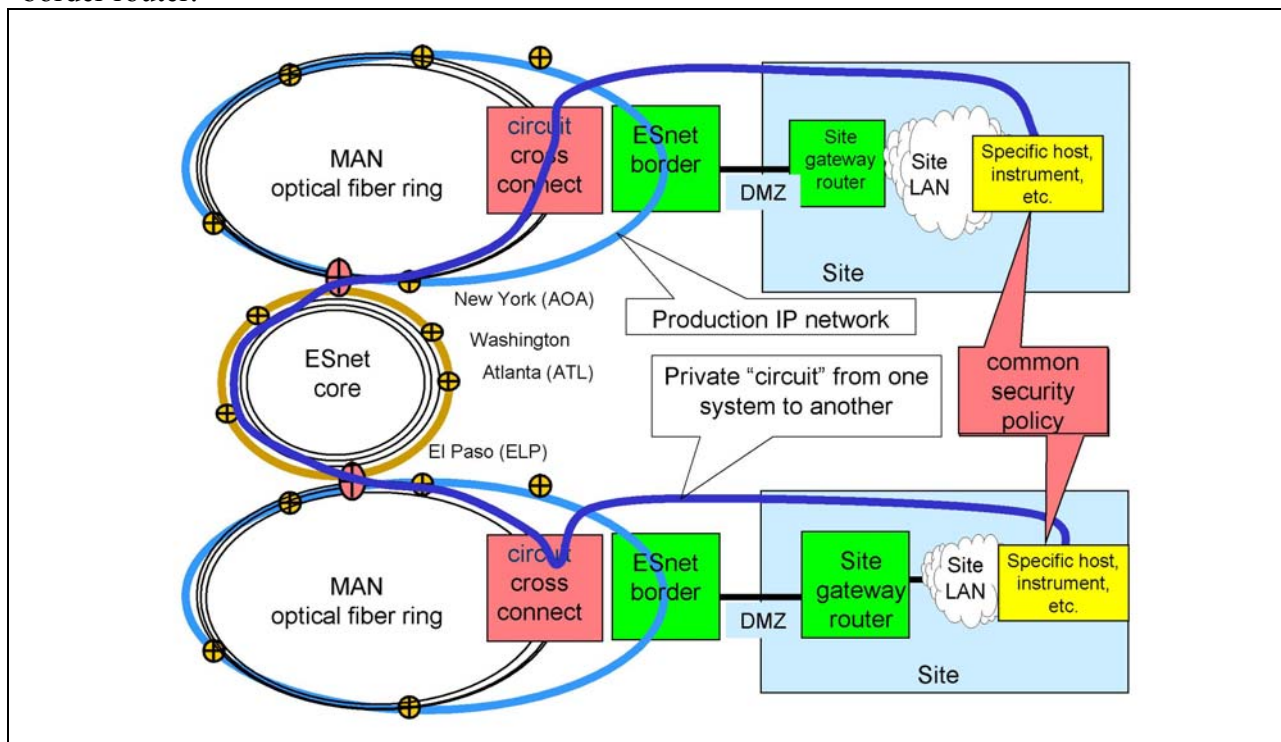


Figure 2: Dynamic provisioning of private, end-to-end “circuits” through the core can provide “high impact science” connections with Quality of Service guarantees.

- A few high and guaranteed bandwidth circuits and many lower bandwidth circuits (e.g. for video, remote instrument operation, etc.)
- The circuits are secure and end-to-end, so if the sites trust each other, and if they have compatible security policies, they should be able to establish direct connections by going around site firewalls to connect specific systems – e.g. HPSS (mass storage) <-> HPSS

To achieve this, the goals of this project are to:

- o Develop and deploy a reservation manager with an appropriate authorization, authentication, and auditing scheme to schedule bandwidth requests.
- o Deploy MPLS and RSVP to enable path establishment, and QoS to guarantee bandwidth availability.

- o Create tools to monitor the operation of these paths to ensure that the quality guarantees are being met.
- o Roll out the service initially to three sites (NERSC, FNAL, and SLAC). UltraScienceNet (USNet) will be a partner in this activity to ensure that USNet lambda paths can be mapped to ESnet MPLS paths in the future.

2.1 Establishing Service Offerings in the Network

QoS, MPLS, and RSVP, although not ubiquitously deployed in the Internet, have been available for some time. ESnet has used these mechanisms to statically deploy a Scavenger Service [8] (using QoS), and an overlay network for SecureNet – an encrypted OC3 ATM network – built using MPLS and RSVP. OSCARS however will require the integration of these three mechanisms, together with a network evaluation and scheduling mechanism, to dynamically create bandwidth assured end-to-end paths.

2.1.1 Network Quality of Service (QoS)

The notion of network Quality of Service (QoS) has also been around for a long time. Current solutions involve the use of selectable service levels [5]. These services distinguish themselves from Best-Effort (BE) by either receiving a higher service response such as Assured-Forwarding [6] or Expedited-Forwarding [7], or a lower one such as Scavenger-Service [8]. Along with selectable service levels, recent advances in router technologies have greatly simplified the problem of setting up a QoS path across a network. These technologies include Multi-Protocol Label Switching (MPLS), Resource Reservation Protocol (RSVP) and RSVP-TE (Traffic Engineering).

Current QoS implementation in the Juniper and Cisco routers used in ESnet allows us to define how packets should be treated if a QoS policy is exceeded. In the event that the egress interface is not congested, we have the option of allowing the packet to be queued for transmission, or simply dropped in compliance with the QoS policy. If the egress interface is congested, the QoS policy will be strictly adhered to, and offending packets in the QoS flow will be dropped. Weighted Random Early Detection (WRED) [15] (instead of tail-drop) can be configured to provide finer control.

2.1.2 Multi-Protocol Label Switching (MPLS) and Resource Reservation Protocol (RSVP)

Multi-protocol label switching (MPLS) is used to speed up packet forwarding and to provide for traffic engineering in Internet protocol (IP) networks. To accomplish this, the connectionless operation of IP networks becomes more like a connection-oriented network, where the path between the source and the destination is pre-calculated based on application requirements. To speed forwarding, an MPLS device uses label matching rather than address matching to determine the next hop for a received packet. To provide traffic engineering, tables are used that represent the levels of QoS that the network can support. The tables and the labels are used together to establish an end-to-end path called a Label Switched Path.

Signaling to establish a traffic-engineered LSP is conducted through use of a label distribution protocol that runs on every MPLS node. Existing protocols have been extended so that label distribution can be piggybacked on them [9]. New protocols have also been defined for the explicit purpose of distributing labels [10–12]. ESnet has already deployed RSVP-TE (Traffic Engineering) as the label distribution protocol to establish static LSP tunnels used for the SecureNet overlay network.

The extended RSVP protocol supports the instantiation of explicitly routed LSPs, with or without resource reservations. It also supports smooth rerouting of LSPs, preemption, and loop detection [14]. An advantage of using RSVP to establish LSP tunnels is that it enables the allocation of resources along the path. For example, bandwidth can be allocated to an LSP tunnel using standard RSVP reservations and service classes [9, 13].

To prevent unauthorized devices from creating an LSP, RSVP uses an HMAC-MD5 message-based digest for authentication. This scheme produces a message digest based on a secret authentication key and the message contents. (The message contents also include a sequence number.) The computed digest is transmitted with RSVP messages. The authentication key is set on a per interface basis. This requires all directly connected neighbors on the interface to be configured with the identical key in order for RSVP to function [20], and this is practical in a single domain such as ESnet.

It is important to note that RSVP and QoS specifications are logically distinct. The RSVP specification does not define the internal format of those RSVP protocol fields, or objects, which are related to invoking QoS control services [14]. When labels are associated with traffic flows, it becomes possible for a router to identify the appropriate reservation state for a packet based on the packet's label value. This means that functions such as traffic shaping, policing, and fire-walling packets are not inherently tied to an RSVP reservation, but can be tied to the LSP.

3 Technical Approach

The intent of OSCARS is to create a service for dynamic QoS path establishment that is simple for users to use. The only task required of a user is to make a bandwidth reservation. The user does not have to configure an alternate routing path, nor mark the packets in any way. All necessary mechanisms needed to provide the user with a guaranteed bandwidth path are coordinated by the reservation manager and managed by the routers in the network.

The procedure of a typical path setup will be as follows (see Figure 3):

- 1) A user submits a request to the ESnet Reservation Manager (RM) (using an optional web front-end) to schedule an end-to-end path (e.g. between an experiment and computing cluster) specifying start and end times, bandwidth requirements, and specific source IP address and port that will be used to provide application access to the path.
- 2) At the requested start time, the RM will configure the ESnet router (at the start end of the path) to create a Label Switched Path (LSP) with the specified bandwidth.
- 3) Each router along the route receives the path setup request (via RSVP) and commits bandwidth (if available) creating an end-to-end LSP. The RM will be notified by RSVP if the end-to-end path cannot be established. The RM will then pass on this information to the user.
- 4) Packets from the source (e.g. experiment) will be routed through the LAN's production path to ESnet's edge router. On entering the edge router, these packets are identified and filtered using flow specification parameters (e.g. source/destination IP address/port numbers) and policed at the specified bandwidth. The packets are then injected into the LSP and switched (using MPLS) through the network to its destination (e.g. computing cluster).

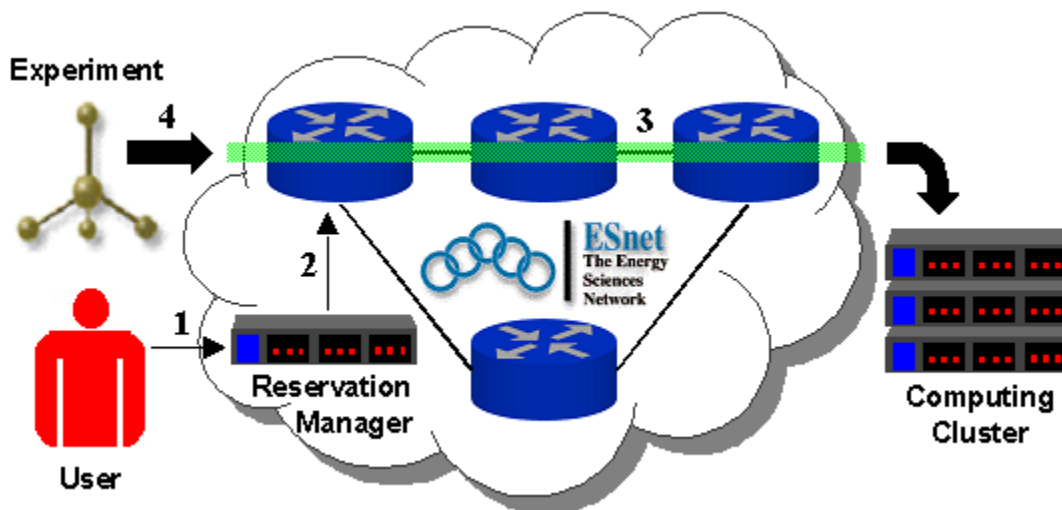


Figure 3: On-demand Bandwidth Path Setup

At the specified end time, the RM will initiate the teardown procedures to remove the path configuration from the router.

3.1 QoS, MPLS, and RSVP

The implementation of an assured bandwidth path consists of two components: 1) The end-to-end path (or circuit), and 2) the guarantee of bandwidth on the path. MPLS and RSVP will allow us to create the circuit using an LSP, and QoS will guarantee the bandwidth.

For this project, ESnet will deploy an Assured-Forwarding service that will be associated with an LSP. The initial implementation will give the LSP traffic better service than BE, up to a predetermined volume (in sync with configured RSVP-TE values). The implementation will also maintain the level of service for overflow traffic if bandwidth is available from the other classes of service. In the event of congestion, overflow traffic will be dropped.

By using the routers' ability to filter packets according to flow specifications (on the ingress interface on the ESnet border router), it is unnecessary to have a separate logical/physical interface to differentiate "regular" packets from "preferred" ones.

The service offering for QoS, MPLS, and RSVP in the network will largely be static (e.g. QoS may be configured to use a maximum of 50% on a specific link). It is therefore important to test different values and configurations that will best suit OSCARS and minimize the impact on the BE production IP traffic.

3.2 Reservation Manager (RM)

The Reservation Manager (RM) will be comprised of three components, the Authentication, Authorization, and Auditing Subsystem (AAAS), Bandwidth Scheduler Subsystem (BSS), and the Path Setup Subsystem (PSS). The RM will have an architecture that is consistent with the WS-RF model [32] of a service, even if it does not initially implement all of the WS-RF conventions. It will define a language independent standard interface using the Web Service Description Language (WSDL) [30]. The system state is managed by the RM and is invoked through a standard service interface using the Simple Object Access Protocol (SOAP) [29]

protocol. This will facilitate the option for other applications, irrespective of programming languages (including middleware) using SOAP to communicate directly with the RM, and will allow for more full integration with WS-RF as implementations mature.

3.2.1 Web-Based User Interface (WBUI)

The initial user interface will be a secured web front-end that authenticates users and sends reservation information to the Reservation Manager (RM). Requests will be sent to the RM in the form of signed SOAP messages.

In principle, this WBUI is just one of many applications that could use the RM service.

The purpose of a web front-end module is to provide the user with a simple interface for making reservations. However, from the point of view of the RM, the WBUI is just an application that speaks the appropriate protocol, and has the appropriate authorization, for the RM service.

3.2.2 Authentication, Authorization, and Auditing Subsystem (AAAS)

The AAAS will be responsible for authenticating the requester and validating the authorization of the requests for bandwidth usage. The AAAS will receive signed SOAP messages (either from the WBUI or an equivalent application) with the reservation information and an authenticated identifier (such as a DOEGrids certificate [17]). This will be done through a standard published interface using WSDL. Authorization and auditing (with further investigation) may potentially leverage off a customized Virtual Organization Membership Service (VOMS) [18-19]. Auditing information generated from usage records will be used in the future for accounting and allocation management.

It is important to note that when an end-to-end path is created, resources will be consumed along the entire path (from host-to-host). It is therefore essential to coordinate with entities along the path (e.g. sender, carrier, and receiver) to gain a confederated agreement. Sender and receiver agreement mechanisms are beyond the scope this project, which is focused on an intra-ESnet service, but end user sites (sender *and* receiver) will have to address this issue.

The aim in designing the AAAS is to integrate existing or emerging standards and services to accomplish the necessary functionality and security, and then to determine if the resulting level of security is compatible with the (fairly stringent) security model of ESnet as a production service critical to DOE's science mission. If the level of security is inadequate, the weaknesses will be analyzed and the AAAS subsystem redesigned.

3.2.3 Bandwidth Scheduler Subsystem (BSS)

The BSS will be responsible for tracking the reservations of the various links within ESnet involved in the end-to-end bandwidth guaranteed paths. The BSS must be able to say whether a reservation request is compatible with the QoS service offering combined with the current state of the network. (I.e., is there sufficient QoS capacity throughout the life of the requested reservation to accept the reservation?) In order to effectively make the reservations, it is essential that the BSS is aware of the state of the network (present and future). To accomplish this, the BSS will need to interact with existing production tools that monitor and coordinate changes in the network.

Issues that the BSS will need to address include:

- o Planned network outage awareness. In the event of a planned network outage, all resources affected by the disruption must be rescheduled. If a backup path is to be used,

conflicts with existing reservations on that path must be resolved (via a set of policies). The BSS will interact with the ESnet Planned Maintenance Calendar [3] system to gain knowledge of all affected routers and links during planned network outages.

- o The Outage Footprint Calculator [3] will be used to determine the impact of unplanned outages and then, if possible, adjust paths as necessary.
- o Policy Enforcement Point (PEP). The BSS must enforce policies to manage resource allocation conflicts. The actions taken in the event of a conflicting reservation might range from a request denial, to preemption. Initially, a first-come-first-served method will be used. This may then potentially evolve into an economy, permitting prioritized reservations with preemption as a possibility. The BSS will use standard Policy Decision Point (PDP) mechanisms to obtain the policies from an external source.

In keeping with a focus on standards, the interface between the AAAS and BSS will evolve out of the work done by the Grid High-Performance Networking Working Group (GHNP-WG) [31].

We will use the background established by the Scalable Trusted Advance Reservation System (STARS) [4] and General-purpose Architecture for Reservation and Allocation (GARA) [21] to develop both the AAAS and BSS.

The effectiveness of the BSS will largely depend on its interaction with network monitoring tools, and its ability to enforce policy. In order to be compatible with a network where the production IP traffic may never be interrupted, the BSS must effectively detect and respond to the impact of unplanned outages. Both of these issues will be examined within this project.

3.2.4 Path Setup Subsystem (PSS)

The PSS is responsible for setting up and tearing down the on-demand bandwidth paths. This is accomplished by making the necessary configuration changes in the routers to create or destroy a Label Switched Path (LSP) at the times dictated by the BSS. The authentication and authorization method for the PSS is internal to the ESnet network and will be specific to the router platform (currently Juniper or Cisco) being configured, and is therefore distinct from the AAA used by the BSS.

Prior to setting up the LSP, the PSS will query routers along the path to ensure that there are no resource conflicts with existing LSPs. (This may occur if there was an unscheduled network outage causing LSPs to utilize alternate paths). A notification of the success or failure of the LSP setup will be passed back to the PSS/RM so that the user can be notified and the event can be logged for auditing purposes.

It is essential for the PSS to validate the path of the LSPs and coordinate changes with the BSS and user when conflicts arise.

The issue of selecting alternate paths is discussed in the implementation section.

3.3 Monitoring

LSPs will be monitored primarily for setup/teardown and usage. This will largely be done via SNMP queries to the routers. Upon teardown of an LSP, all summary data associated with the LSP (e.g. user, source/destination, time of setup/teardown, bandwidth allocated, QoS value, and packets/bytes forwarded) will be sent to the user and logged for future analysis. Detailed data such as traffic graphs may also be generated. The amount and extent of information gathered for

detailed data reports will be a compromise of resource consumption and practicality (e.g. polling every 10 seconds to monitor packets/bytes forwarded on an LSP might be too resource intensive, versus polling every 10 minutes which may not yield useful data).

Several approaches to monitoring the service will be investigated. By comparing the reservation parameters with the router behavior records generated for each LSP, we should be able to monitor the router's view of whether the user's service requirements were met. We will also determine what sort of external active monitoring can validate the router reporting of the guarantees.

3.4 Traffic Shaping

In the dynamic environment envisioned for this service, where the service entry points are interfaces on ESnet border routers and users are outside of the ESnet environment, the input streams may easily be capable of generating more traffic than permitted by the QoS reservation. This implies the need for a traffic-shaping element between the application and the QoS entry point, as most applications cannot shape their own traffic at the packet level. However, this issue is beyond the scope of this proposal, which is limited to the issues of the intra-ESnet service.

4 Implementation

4.1 QoS

Assured-Forwarding (AF) will be the QoS service utilized in this project. Unlike Scavenger-Service (SS) (currently deployed in ESnet) which makes no guarantees, AF must be configured carefully so as to provide the appropriate level of service without greatly impacting traffic on other service classes such as Best-Effort (BE).

We will decide on a suitable maximum bandwidth utilization (per link) that is acceptable for this service and an appropriate policy if this bandwidth is exceeded. Two options are readily available. The first would be to drop offending packets, and the second is to remap them into a different service class such as BE or SS.

The QoS operation is associated with router interface queues. For an LSP to acquire the desired QoS service, each router along the path must insert the LSP packets in the appropriate egress queues. When an LSP is created, a Class-of-Service (CoS) is attached to it. Packets entering the LSP will inherit the CoS value, and routers along the path will use this value to map the traffic to the correct queues for transmission. By using (Cisco) access-lists and (Juniper) firewall filters, "preferred" packets can be identified (by source/destination IP Address/port) from "regular" packets entering the ESnet border router on a single interface. (A separate logical/physical interface for "regular" and "preferred" traffic is unnecessary.)

To ensure correctness of operation, we will run test applications (such as Iperf [23]) and monitor throughput as well as packet drops along the entire path. We will also monitor the impact of the service on the production, BE traffic to see if our estimate of the maximum bandwidth utilization for the QoS traffic is non-interfering. If it is not, then we will investigate the reasons for this discrepancy and adjust accordingly.

4.2 MPLS and RSVP

ESnet employs both MPLS and RSVP to support the SecureNet overlay. OSCARS will use both of these components but with a focus on QoS. When an LSP is created (both ends terminating in

ESnet), an accompanying inbound policy is configured that will map (according to source/destination IP address/port) ingress traffic (from the site) into the appropriate LSP. The inbound policy will also contain policing parameters to dictate the appropriate action to handle overflow traffic on the LSP.

QoS can police traffic per class, but not individual flows within a class. Flow-based shapers could be used to subdivide the classes while still maintaining QoS guarantees, however this is outside the scope of this proposal.

In the event that the LSP is broken due to an unplanned network interruption, MPLS can be preconfigured to automatically establish an alternate path. Issues of over-subscribing resources on the alternate path will need to be resolved. The other approach is to let the RM handle the problem. The PSS (via an SNMP trap) will be notified of the collapsed LSP. It will then chart the alternate route and consult the BSS if a new LSP should be created. These alternatives will be investigated in this project to determine the most effective approach.

4.3 Resource Manager

Based on discussions with several people working on Grid services, we anticipate that a significant amount of code for the RM will be able to be reused from other projects or automatically generated using standard WS tooling, e.g. the SOAP communication, authentication and certificate management for authorization, etc.

4.3.1 Web-Based User Interface

The WBUI will be within the trusted domain of the RM, and therefore security is a primary focus. Hyper Text Transfer Protocol Secure (HTTPS) will be used by the user to connect to the WBUI. The user will use a username/password to authenticate and submit a reservation request. The WBUI will then place the information into a signed SOAP message (including the user's authenticated identifier) and forwards it to the AAAS of the RM. Communication between the WBUI and the RM will be authenticated and secured initially using HTTPS, although as WS-RF matures it is likely that WS-Secure Conversation [34] will be used for this purpose.

4.3.2 AAAS

The AAAS will accept reservation requests in the form of signed SOAP messages. The AAAS will initially only accept requests from the WBUI. This will subsequently change to allow direct request from user applications and reservation managers from other domains (e.g. ESnet and Abilene sites).

Initially, when a signed SOAP message is received, the AAAS will extract the authenticated identification (ID), which will be a distinguished name from a DOEGrids Certificate [17]. The certificate will then be retrieved from the DOEGrids certificate repository and used as though it had been passed as part of a user signed token through the SOAP protocol. This will let us proceed in the design and implementation as though the user had authenticated originally using a certificate and then passed a proxy cert through to the AAAS. (This is what an application would do since there might not be a live user involved.)

This work-around is necessitated by the lack of a standard way to pass Grid proxy certificates through a web browser. Other communities are investigating several solutions to this problem. The most common solution is to use a MyProxy [35] server to store proxies and allow the WBUI to retrieve a proxy for the user from the credential store. This credential can then be used to

authenticate to the AAAS in a standard way. As these solutions mature, this project will take advantage of standard mechanisms.

To authorize the request, the AAAS will pass the authenticated ID to a PDP. Initially the PDP will be implemented as an access-list (ACL) in the RM. This will evolve into an external policy engine (e.g. VOMS).

If a request is authorized, it is handed to the BBS (through an interface currently being defined by the GHNP-WG).

4.3.3 BSS

The BSS will start as a simple calendaring system. Each link along the end-to-end path (determined by a traceroute) will have a predetermined bandwidth value associated with it. This value (which will dictate the QoS values) represents the maximum cumulative bandwidth (on that link) that can be allocated for on-demand bandwidth guaranteed paths. As the end-to-end paths are reserved (most likely via a web-page), the corresponding link values are decremented accordingly.

The next evolution of BBS will enable it to reschedule link reservations in the event of a planned network outage. (Active LSPs disrupted by unplanned events will be handled by MPLS and RSVP.) Knowledge of forthcoming (planned) outages will be mined from the ESnet Planned Maintenance Calendar system. A complication that arises from having to reschedule, is that of knowing what the alternate path will be (a traceroute prior to the outage would obviously not be useful). A simple solution is to deny the request if any of the links are affected by the planned outage. An alternative is to have a static map indicating alternate paths between strategic points in the network. This approach however is not scalable.

A more general solution to this re-routing problem will be investigated in this project.

Conflicts in resource reservation will be handled by the same PDP and PEP mechanisms as used by the AAAS.

The design of the BSS will be cognizant of the fact that if this service is successful, then in a production mode in the future it will undoubtedly have to be managed by an allocation management system as the service is a scarce resource.

4.3.4 PPS

The PSS (which is internal to the ESnet network) will instantiate an LSP by changing the configurations on the start-end of the path. For Juniper routers, this can be done using JUNOScript [22] (an Extensible Markup Language (XML) based application that Juniper Networks routers use to exchange information with client applications). For Cisco routers, the Really Awesome New Cisco config Differ (RANCID) [36] application will access the router and make changes via the command-line interface.

Prior to setting up an LSP, the PSS must perform two crucial tasks.

- 1) Validate the route the LSP will use. In the event of an unplanned outage prior to setting up the LSP, the path may deviate from the initial reservation. The PSS must feedback route differences to the BSS in-order to resolve scheduling conflicts (if applicable).

- 2) Query routers (using SNMP) along the path check for active LSPs. This is done to verify that the reserved resources are available for use. If an unscheduled LSP is present, the appropriate personnel will be notified and proper actions taken.

The success or failure to setup the LSP will be passed back to the RM and disclosed to the user by e-mail and/or a status page on a web-server.

4.4 Monitoring

We will modify our existing network statistics system to poll the routers (using SNMP) for LSP data. Both Cisco and Juniper have proprietary MIBs to address LSP queries. Tools such as RRDtool [24] may be used to store data and generate usage graphs.

To ensure that the guaranteed bandwidth mechanisms are operating correctly, bandwidth measurement tools (such as Iperf) can be run on a regular basis. By comparing the empirical throughput with the guaranteed bandwidth, we can validate quality assurance.

5 Interoperability

A goal of this project is to use current and emerging standards to advance interoperability between heterogeneous networks. Using widely accepted mechanisms such as MPLS, RSVP, and QoS, will allow us to annex similar services from other projects such as the Demonstration of Advance Reservation and Services by DataTAG [25], the Hybrid Optical Packet Infrastructure [26], and USNet. It will also position us to integrate emerging protocols such as the user-to-network (I-UNI) and inter-carrier (I-ICI) signaling solutions proposed by the Infranet Initiative [27 – 28].

6 Work Schedule

6.1 Year 1

- Test and deploy MPLS and RSVP.
- Develop and deploy the Resource Manager.
- Implement web-base user interface.
- Implement basic access-control security for AAAS.
- Develop simple scheduling algorithms for BSS.
- Test and implement access methods for PSS.
- Test and deploy QoS (with appropriate priority/drop characteristics).
- Test at least one user-level application using the QoS service.

6.2 Year 2

- Create tools to monitor LSP setup/teardown and bandwidth usage.
- Test and deploy DOEGrids certificate authentication for AAAS.
- Evaluate the AAAS with a user community.

6.3 Year 3

- Test and deploy authorization and auditing mechanisms for AAAS.
- Develop rescheduling algorithms for BSS to address network changes during a reservation.
- Evaluate the BSS with a user community.
- Test and develop PDP and PEP for AAAS and BSS.
- Test and deploy Generalized MPLS (GMPLS) to include optical cross connect equipment if applicable.

7 References

- [1] High Performance Network Planning Workshop, August 2002:
<http://www.doecollaboratory.org/meetings/hpnpw/index.html>
- [2] DOE Science Networking Challenge: Roadmap to 2008, June 2003:
<http://www.es.net/hypertext/welcome/pr/Roadmap/index.html>
- [3] Network Availability Management and Reporting, Mike O'Connor, ESCC, January 2004:
<http://www.es.net/pub/esnet-doc/ESCC-Jan-04-Hawaii-Mtg/moc-ESCC.pdf>
- [4] Advance Reservation/Quality of Service Work: <http://www.dsd.lbl.gov/QoS/>
- [5] RFC2990: Next Steps for the IP QoS Architecture. G. Huston. IETF RFC, November 2000.
- [6] RFC2597: Assured Forwarding PHB Group. J. Heinanen, F. Baker, W. Weiss, J. Wroclawski, IETF RFC, June 1999.
- [7] RFC2598: An Expedited Forwarding PHB. V. Jacobson, K. Nichols, K. Poduri, IETF RFC, June 1999.
- [8] QBSS: <http://qbone.internet2.edu/qbss/>
- [9] RFC3209: RSVP-TE: Extensions to RSVP for LSP Tunnels. D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, G. Swallow, IETF RFC, December 2001.
- [10] RFC3036: LDP Specification. L. Andersson, P. Doolan, N. Feldman, A. Fredette, B. Thomas, IETF RFC, January 2001.
- [11] RFC3212: Constraint-Based LSP Setup using LDP. B. Jamoussi, Ed., L. Andersson, R. Callon, R. Dantu, L. Wu, P. Doolan, T. Worster, N. Feldman, A. Fredette, M. Girish, E. Gray, J. Heinanen, T. Kilty, A. Malis, IETF RFC, January 2002.
- [12] RFC3031: Multiprotocol Label Switching Architecture. E. Rosen, A. Viswanathan, R. Callon, IETF RFC, January 2001.
- [13] RFC2211: Specification of the Controlled-Load Network Element Service. J. Wroclawski, IETF RFC, September 1997.
- [14] RFC2210: The Use of RSVP with IETF Integrated Services. J. Wroclawski, IETF RFC, September 1997.
- [15] Cisco Systems Inc., "Advanced QoS Services for the Intelligent Internet," White Paper, July 1998.
- [17] DOEGrids Certificate Service: <http://www.doegrids.org>
- [18] VOMS: <http://hep-project-grid-scg.web.cern.ch/hep-project-grid-scg/voms.html>
- [19] US CMS VO Project: http://www.uscms.org/s&c/VO/cms_vo_home.html
- [20] JUNOS Internet Software Configuration Guide:
<http://www.juniper.net/techpubs/software/junos/junos62/swconfig62-mpls-apps/html/rsvp-config9.html#1032498>
- [21] GARA: <http://www-fp.mcs.anl.gov/qos/>
- [22] JUNOScript: <http://www.juniper.net/support/junoscript/>
- [23] Iperf: <http://dast.nlanr.net/Projects/Iperf>
- [24] RRDtool: <http://people.ee.ethz.ch/~oetiker/webtools/rrdtool/>
- [25] Demonstration of Advance Reservation and Services, DataTAG Deliverable DataTAG-D2.5- 1.3, Feb 2004.
- [26] Developing a Hybrid Optical/Packet Infrastructure (HOPI) in the U.S.:
<http://www.internet2.edu/presentations/20031121-ISC-Corbato.ppt>
- [27] The Infranet Initiative: http://www.juniper.net/solutions/literature/infranet_initiative.pdf
- [28] Infranet Inter-Carrier Capabilities Solution Brief:
<http://www.juniper.net/solutions/literature/solutionbriefs/351030.pdf>

- [29] Simple Object Access Protocol (SOAP): <http://www.w3.org/TR/SOAP/>
- [30] Web Services Description Language (WSDL): <http://www.w3.org/TR/wsdl>
- [31] Grid High-Performance Networking: <http://forge.gridforum.org/projects/ghpn-rg/>
- [32] From Open Grid Services Infrastructure to WS-Resource Framework: Refactoring & Evolution: <http://www-106.ibm.com/developerworks/library/ws-resource/gr-ogsitowsrf.html>
- [33] OASIS Web Services Resource Framework TC: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsrf
- [34] G. Della-Libera et. Al. Web Services Secure Conversation Language (WS-SecureConversation). Version 1.0. December 18, 2002. Available: <http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnglobspec/html/ws-secureconversation.asp>
- [35] J. Novotny, S. Tuecke, and V. Welch. [An Online Credential Repository for the Grid: MyProxy](#). Proceedings of the Tenth International Symposium on High Performance Distributed Computing (HPDC-10), IEEE Press, August 2001. MyProxy Online Credential Repository. <http://grid.ncsa.uiuc.edu/myproxy/>
- [36] RANCID – Really Awesome New Cisco confIg Differ: <http://www.shrubbery.net/rancid/>