Providing Quality of Service across Multiple Providers: The Case of European Research and Academic Space

Christos Bouras, University of Patras, Greece

Apostolos Gkamas, University of Patras, Greece

Kostas Stamos, University of Patras, Greece

ABSTRACT

In this chapter, the authors present some of the latest developments related to the provisioning of Quality of Service (QoS) in today's networks and the associated network management structures that are or will be deployed to support them. They first give a brief overview of the most important Quality of Service proposals in the areas of Layer 2 (L2) and Layer 3 (L3) QoS provisioning in backbone networks, and they discuss the network management structures and brokers that have been proposed in order to implement these services. As a case study, they describe the pan-european research and academic network, which is supported centrally by GEANT and which encompasses multiple independent NRENs (National Research and Education Networks). In the last few years, GEANT has developed and deployed a number of production and pilot services meant for the delivery of quality network services to the end users across Europe.

Keywords: Quality of Service, Bandwidth on Demand, Network provisioning, Network monitoring, Bandwidth Broker, GEANT2, AutoBAHN, Pathfinding, AMPS

INTRODUCTION

The GN2 European project GÉANT2 (2009) encompasses a range of research activities to advance both networking and user services in Europe. Central to this project, is the goal of providing high-quality services from one end user to another over multiple interconnected networks. The GÉANT2 (2009) network connects 34 countries via 30 national research and education networks (NRENs), using multiple 10Gbps wavelengths. GÉANT2 also connects to worldwide NRENs and the public Internet to ensure a global Gigabit-per-second connectivity for all users.

Quality of Service has been developed as a concept for several years and has reached maturity especially related to L3 implementations, although user demand and provider uptake has not always been as high as expected. The emergence of L2-based network architectures that try to avoid the high costs of high-end routing capabilities and take advantage of direct administration of optical circuits by organizations such as European NRENs, has led to the need for L2 QoS and Bandwidth on Demand services.

GEANT has deployed services in two main areas: The provisioning of L3 QoS based on DiffServ architecture, and the provisioning of Bandwidth on Demand (BoD) based on dynamic allocation of L2 circuits. The GN2 activity that has specified and is now prototyping a Bandwidth on Demand service intended to operate in a multi-domain environment using heterogeneous transmission technologies is called AutoBAHN, while the GN2 activity that has developed a L3 QoS provisioning framework is called AMPS. In addition GN2 has developed a monitoring system for the AutoBAHN service, which has proved in itself a complex and challenging task.

This chapter presents some of the latest developments related to the provisioning of Quality of Service (QoS) in today's networks and the associated network management structures that are or will be

deployed to support them. The remaining of this chapter is structured as follows: The next section presents the international experience in the area of L2 and L3 QoS**.** Section 3 presents the efforts of GEANT in order to implement and deploy L2 and L3 QoS services. Section 4 presents the future trends in the area. Finally, Section 5 concludes this chapter.

BACKGROUND

QoS architectures at Layer 3

IP networks are built around the idea of best effort networking, which makes no guarantees regarding the delivery, the speed and the accuracy of the transmitted data. While this model is suitable for a large number of applications and works well for almost all applications when the network load is low (and therefore there is no congestion), there are two main factors that combine to lead to the need for an additional capability of quality of service guarantees. One factor is that the amount of real-time and other multimedia data transmitted over the Internet increases, and this type of data have stricter service requirements. The other factor is that Internet usage in general is steadily increasing, and although the network infrastructure is often also updated, it is not always certain that network resources offering will be ahead of network usage demand. Several researchers for example, argue that there are signs that Internet demand is outstripping capacity (Nemertes, 2007). Furthermore, several providers are implementing usage caps to alleviate the problem.

The two main architectures that have been proposed for Quality of Service are IntServ and DiffServ. They follow different philosophy as they approach the topic of Quality of Service from different point of views.

The IntServ architecture tries to provide absolute guarantees via resource reservations across the paths that the traffic class follows. The main protocol that works with this architecture is the Reservation Protocol (RSVP). However, its operation is quite complicated and it also inserts significant network overhead. On the other hand, DiffServ architecture is more flexible and efficient as it tries to provide Quality of Service via a different approach. It classifies all the network traffic into classes and tries to treat each class differently, according to the level of QoS guarantees that each class needs. In the DiffServ architecture, 2 different types (per hop behaviours, Nichols, 2001) have been proposed, the expedited forwarding (Jacobson et al., 1999) and the assured forwarding (Heinanen et al., 1999), and their difference is on the packet forwarding behaviour. Expedited forwarding (EF) aims at providing QoS for the class by minimizing the jitter and is generally focused on providing stricter guarantees. This type tries to simulate the virtual leased lines and its policy profile should be very tight. Assured forwarding (AF) inserts at most 4 classes with at most 3 levels of dropping packets. Every time the traffic of each class exceeds the policy criteria then it is marked as lower level QoS class.

The operation of DiffServ architecture is based on several mechanisms such as packet classification, packet marking, metering, shaping and queue management. The classification is done via marking the DSCP (Differentiated Service CodePoint) field.

Although QoS provisioning mechanisms have been extensively tested and deployed in several networks, the biggest hurdle in wider application of the idea has been the fact that the Internet is fragmented in separate administrative sections (domains). Even if a domain implements some form of QoS provisioning and guarantees, there is no guarantee that traffic will receive the same treatment end to end, because there is no widespread availability and deployment of multi-domain automated provisioning systems. The research effort in GEANT2 has largely been devoted in changing this situation.

IP Premium service

The IP Premium service has been defined in the framework of the SEQUIN project (Bouras et al., 2003) and aims at providing absolute bandwidth guarantees and minimum delay and jitter to a subset

of the overall network traffic. Its main characteristic is that it follows the classic DiffServ architecture. It classifies the packets using the DSCP values for admitted and downgraded packets. The policing is performed at the edge of the network and high priority queuing is applied in the core and access routers at the outgoing interfaces.

QoS at Layer 2 and Bandwidth on Demand

In recent years, rapid technological developments combined with strong growth in demand for transmission capacity has encouraged network carriers to invest heavily in optical network infrastructure. The following paragraphs give an overview of the standardization efforts that have taken place in this area, and specifically the related technologies and standards proposed by the Internet Engineering Task Force (IETF), the International Telecommunication Union (ITU-T) and the Optical Internetworking Forum (OIF).

Generalised Multi Protocol Label Switching (GMPLS)

Generalised Multi Protocol Label Switching (GMPLS) is a technological framework proposed by the Internet Engineering Task Force (IETF) and targeted at enabling dynamic provisioning capabilities in optical networks. The approach followed by IETF has been to extend the well-known Multi Protocol Label Switching (MPLS) technological framework to encompass devices used for building optical networks.

According to the Generalised MPLS (GMPLS) framework RFC3945 (2004), the MPLS Traffic Engineering (TE) control plane is extended to include network elements such as Add-Drop Multiplexers (ADMs) and Optical Cross-Connects. The MPLS framework RFC3031 (2001) was designed for network elements capable of recognizing packet or cell boundaries. In contrast, the GMPLS framework has been proposed for network elements that can also recognise time-slots, lambdas or ranges of lambdas and fibres. More specifically, according to the GMPLS architecture the following interfaces are defined:

1. *Packet Switch Capable (PSC)* interfaces are able to recognise packet boundaries and forward data based on the content of a specific field in the packet header. Examples of this type of interface include router interfaces.

2. *Layer-2 Switch Capable (L2SC)* interfaces are able to recognise frame or cell boundaries and switch data based either on the value of the MAC header for Ethernet frames or on the value of the VPI/VCI pair for ATM cells. Examples of this type of interface include Ethernet and ATM switches interfaces.

3. *Time-Division Multiplex Capable (TDM)* interfaces switch data based on time-slots. An example of this type of interface is an SDH interface.

4. *Lambda Switch Capable (LSC)* interfaces switch data based on the wavelength on which incoming signal is modulated. These kinds of interfaces may also be able to switch contiguous groups of wavelengths (waveband switching). An example is the Optical Add Drop Multiplexer (OADM).

5. *Fibre-Switch Capable (FSC)* interfaces switch data based on the fibre or fibres that they are using. An example is the interface of a photonic cross connect.

Automatically Switched Optical Network (ASON)

To overcome the limitations of centralised manual provisioning, ITU-T Study Group 15 started, following a top down approach, the development of complete definition of the operation of an Automatically Switched Transport Network G.807 (2001). Automatically Switched Optical Network

(ASON) G.8080 (2001) is not a protocol or collection of protocols. It is a framework that defines the components in an optical control plane and the interactions between these components. An Automatically Switched Optical Network is an optical transport network that is capable of dynamically adding and removing connections. This capability is accomplished by using a control plane that performs the call and connection control functions in real time.

ASON can be thought of as an improved optical transport network (OTN) that adds sufficient intelligence to the optical nodes to permit dynamic provisioning that can respond to changing traffic patterns. ASON is an architecture that defines the components of an optical control plane and the interactions between those components. In itself, it does not define any protocols. A key principle of ASON is to explicitly build a framework that supports legacy network equipment.

The ASON architecture is based on the assumption that a network's design and segmentation is dictated by the operator's decisions and criteria (e.g. geography, administration, technology). Network subdivisions are defined in ASON as 'routing areas'. Recommendation G.8080 defines a routing area as a set of subnetworks. A routing area contains smaller routing areas interconnected by Subnetwork Termination Point Pool (SNPP) links. Routing uses a hierarchical structure based on a decomposition of the network into a subnetwork hierarchy. Each subnetwork has its own dynamic connection control, which knows its own topology but does not know the topology of other subnetworks belonging to either the same hierarchical level or different levels.

Optical Internetworking Forum (OIF)

The Optical Internetworking Forum (OIF) has as its main objective to foster the development of a low-cost and scalable internet using optical technologies. In order to achieve this, OIF brings together the architectures and requirements defined by ITU-T as well as the protocols defined by IETF into a complete working solution. OIF has defined two interfaces:

- User to Network Interface (UNI): The UNI 1.0 recommendation was defined by OIF in December 2001, the current release was published in February 2004. The OIF UNI 1.0 enables clients to establish optical connections dynamically using signalling procedures compatible with GMPLS signalling. The UNI 2.0 specification finalized in 2008. It extends UNI 1.0 by adding various functionalities such as the separation of call and connection control and non disruptive connection modification.

- External Network to Network Interface (E-NNI): The purpose of E-NNI is to support deployment of an optical control plane in a heterogeneous environment. The support for such technically heterogeneous networks is achieved by introducing the concept of control domains. Signalling and routing information exchange between those domains is performed over the E-NNI. Thus the OIF specifies the E-NNI to be a interface for signalling messages, attributes and flows for the creation of transport connections across multiple heterogeneous domains.

Other research approaches

Several other research networks have dealt with similar issues and devised their own approaches. For example, the DRAGON project (Leung et al., 2006) has also conducted research and developed technologies to enable dynamic provisioning of network resources on an interdomain basis across heterogeneous network technologies. The OSCARS/BRUW project (Guok, 2005) focuses on L3 MPLS QoS and adopts SNMP queries to the routers for monitoring LSP teardown and usage. The UCLP community project and the related Argia commercial product enable users to control and manage network elements for the purposes of establishing End-to-End (E2E) lightpaths (Wu et al., 2003, Argia Web site, 2007). The MUPBED project attempts to integrate and validate, in the context of user-driven large-scale testbeds, ASON/GMPLS technology and network solutions (Cavazzoni, 2007). MUPBED takes into account that for an optical network operated according to OTN, specific

fields of the frame are reserved to user's monitoring of a connection. VIOLA (VIOLA Web site, 2007) is another related project for the development and test of software tools for the user-driven dynamical provision of bandwidth but does not focus on the development of tools for the monitoring of the provisioned resources. VIOLA is focused instead on the area of network technology and application development as part of a testbed environment.

IETF has worked on the concept of Path Computation Element (PCE) which aims to separate routing decisions from the packet forwarding procedures (RFC4655, 2006). GEANT2 approach has been similar to the one proposed by IETF, in that path finding decisions are taken at an overlay control layer. Because of the multi-domain requirements of GEANT2, it has placed greater focus on the abstraction and limited information exchange, and less focus on optimal path selection mechanisms.

THE CASE OF EUROPEAN RESEARCH AND ACADEMIC SPACE

Description of GEANT2 activities

Within the GN2 project, a number of service activities (SAs) and joint research activities (JRAs) are being pursued. Service activities represent more mature services that are considered to be ready for production deployment, while research activities lead to pilot deployments of services. Since the issue of Quality of Service in layer 3 had been researched and numerous experimental deployments had taken place, it was decided that such a service should be available in a mature level within GEANT. It has also been understood that in the research and academic environment, there are applications and research fields (such as radioastronomy, high-energy physics and general Grid applications) with strict demands for the provisioning of guaranteed and dedicated capacity. For this reason, the GN2 project has also developed the AutoBAHN (Bandwidth on Demand – BoD) Joint Research Activity 3.

The main purpose of the relevant GEANT2 activities was to design, implement and deploy systems for the interoperation of provisioning mechanisms between different domains, so that traffic from one end of Europe could reach another end at a potentially different country with a network managed by a different entity, while at the same time receiving guaranteed service (in the case of Layer 3 QoS provisioning) or even a dedicated circuit (in the case of Bandwidth on Demand service). Several research and academic networks across Europe had already developed their own solutions for their inner network, however these solutions were not interoperable and could therefore only guarantee service until the boundary of the specific domain. The GEANT2 design of these services has therefore gone to great length in order to make sure that the deployed systems are modular enough so that pre-existing solutions can "plug-in" and interoperate in an international context. Although this design choice adds some complexity to the system, a monolithic solution would be unacceptable as few, if any, network domains would agree to give up on their installed and tested provisioning systems for a completely new solution. Furthermore, GEANT2 research activities have designed the provisioning systems in such a way that as little information as possible is shared between domains, since separate network management entities are not willing to freely share detailed information about their inner network.

Since both the AMPS and AutoBAHN systems faced the above requirements, they have taken a similar approach in general system design and interoperability, as will be seen in more detail in the following paragraphs.

GN2 L3 QoS service (AMPS)

AMPS (ADS, 2006) implements the recommendations of the SEQUIN (Bouras et al., 2003) project for providing premium IP service across multiple, independent networks. The purpose of AMPS is to regulate the maximum allowed amount of prioritised IP traffic in a given domain, thus ensuring that approved traffic continues to receive a top-quality service (low delay, low packet loss) in the event of network congestion. It does this by evaluating each new request for Premium IP (PIP) against existing, approved PIP requests and the total available network resources, and then only approving those new

requests which will not adversely affect any existing reservations. AMPS also includes a communication module which enables it work with neighbouring peers in order to establish a PIP service that crosses multiple domains.

AMPS has been developed for IP-based domains, and is based on the assumption that they are either over-provisioned, or use the DiffServ architecture. Currently, AMPS does not always directly configure the network devices (routers), although several AMPS deployments, such as the one in the Greek NREN GRNET do support automatic router configuration.

The overall AMPS architecture is shown in Figure 1. Communication between the 4 subsystems takes place through web services so that they can be easily upgraded or switched with equivalent modules developed independently (Bouras et al., 2007).

AMPS, similarly to the AutoBAHN system described later, take the approach of developing reference implementations for all their subsystems, but with maximum modularity so that they can be replaced by modules already existing in an organization and so that they can easily be upgraded to support more technologies.

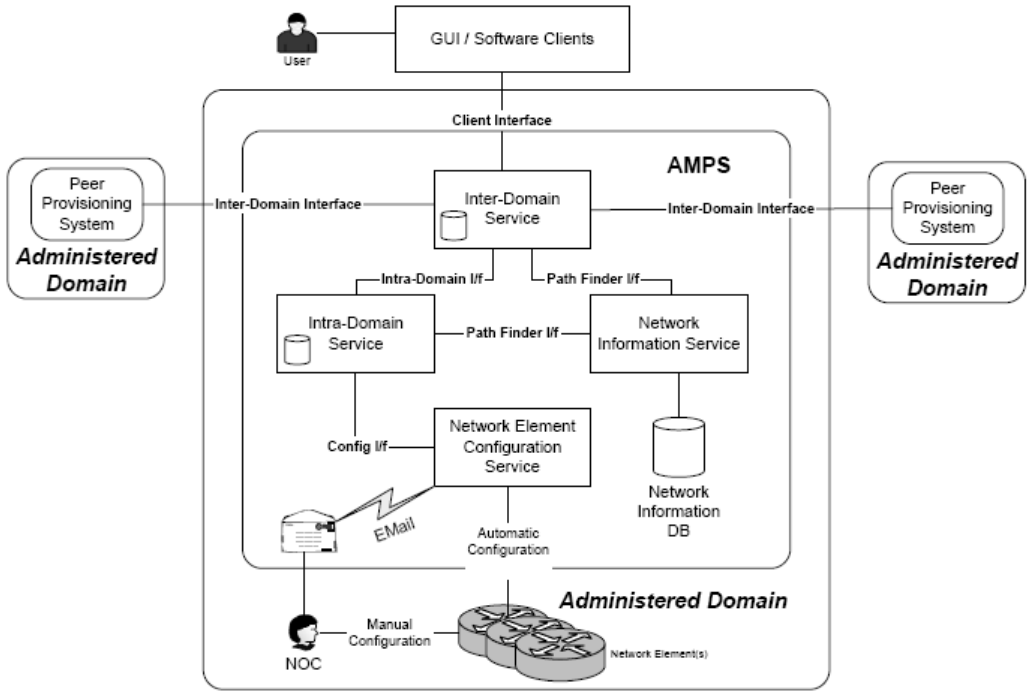

Figure 1. AMPS architecture (ADS, 2006)

Figure 1 displays 3 AMPS instances in 3 neighbouring domains, with the middle one enlarged so that its subsystems and their interactions are visible. A user is also displayed as an example connecting to the middle AMPS instance.

Inter-Domain subsystem

The Inter-Domain subsystem is the core of the AMPS multi-domain functionality, and the module that is expected to be deployed in most, if not all, deployments of the AMPS system. It is responsible for receiving messages from and sending messages to external systems as well as handling transactions. These external systems may be either an end-client or the AMPS system in an adjacent domain. Through the Inter-Domain subsystem a reservation request is relayed from one end of the path to the

other end and the reservation decision is negotiated before being announced to the end user. AMPS implements a chain communication model, where an AMPS Inter-Domain system cannot directly communicate with another Inter-Domain system in a domain to which it is not directly connected, but such messages can be relayed to the distant Inter-Domain system via intermediate Inter-Domain systems.

Intra-Domain subsystem

The Intra-Domain subsystem is responsible for its own domain and it keeps a record of all approved PIP reservations that pass though its domain. Requests for new reservations are passed to it by the Inter-Domain subsystem. The reservation includes basic parameters such as the ID of the user submitting the reservation, source and destination end-point, required capacity, start and stop time. The Intra-Domain subsystem will first check that the request is valid according to several sanity and policy tests (for example, whether the capacity and time period requested are within the rights of the user). This is done by querying its Policy Module. If a request is within policy limits then the Intra-Domain subsystem requests the intra-domain path from the cNIS database. The reservation database is then checked to see if that path has sufficient spare PIP capacity for the requested period. If so, the reservation database is updated with the new information and the Inter-Domain subsystem is informed of a successful reservation.

Reservation status and transaction manager

Since a reservation requests might have to traverse multiple remote domains, and since resources might not be automatically allocated in all domains, AMPS has to implement a method of gradual updates to the end user regarding a submitted reservation's status. Therefore, when the Inter-Domain subsystem receives a request for a new PIP reservation it immediately responds to the client with a service id and status as "pending". The domain where a request is submitted is called the home domain. The Inter-Domain subsystem of the home domain then, acting as a transaction manager, starts a new transaction. It then contacts cNIS using the pathfinder interface to check which is the next domain (if any) and what is the current domain's egress interface and next domain's ingress interface. It then sends the reservation request to the Intra-Domain subsystem. Once the Intra-Domain subsystem reports that reservation request is successful then the request is forwarded to the Provisioning System of the next domain along the path. Assuming that the reservation is successful in the next domain this chain process continues until the last domain along the path is reached. Once the request is successful in the last domain; a message indicating that the request is successful is relayed back to the originating domain. This end-to-end chaining process completes the transaction and the status is updated as "accepted". If any domain in the path rejects the request, a message indicating the request is unsuccessful is relayed back to the originating domain. The transaction manager is then responsible to send a rollback message to all domains along the path where a reservation may have been made to cancel the reservation. Once all domains have rolled back their reservation the transaction manager is notified. This completes the transaction and the request status is changed to 'rejected', with a reason given for rejection. The transaction manager will keep sending the rollback until it receives the acknowledgement (or will notify some human administrator). If any service along the path does not respond, a timeout mechanism will be triggered on the transaction manager and a rollback process will be started automatically. Once all domains have rolled back their reservations the transaction manager is notified. This completes the transaction and status is changed to "rejected", and a reason given for rejection.

The client can periodically query the status of the request. Once the transaction is complete the status will be updated from "pending" to "accepted" or "rejected". If the reservation is submitted to a domain but it starts in another (in other words, if the home domain is different from the source domain), the request will be forwarded to the source domain. Even though the reservation starts in the source domain, the home domain will act as the transaction manager. In all other respects AMPS behaves as described above.

cNIS database

The Network Information Service subsystem uses the Network Information database to calculate what route a given flow will take across the network. The database must therefore have a record of all the links in the network, it must know each link's metric, and it must be informed (by a client, using the NIS's PathFinder interface) the start and end points of the flow. The NIS will then return to the client (using the same PathFinder interface), the link by link path that the specified flow will follow.

The cNIS database started from the AMPS activity but its use has been generalized in the GN2 project, and can therefore store information about layers 1, 2, and 3 and is also used by the AutoBAHN service described below.

GN2 BoD service (AutoBAHN)

Similarly to AMPS, the architecture of AutoBAHN system which implements the GN2 BoD service has been designed to meet the fundamental requirement for operation in a multi-domain, multi-technology environment. The BoD service was defined so that it could provide end-to-end, multi-domain, point-to-point symmetric (in terms of bandwidth capacity and path selection) guarantees. Point-to-multipoint services may be realized as a set of point-to-point ones. Furthermore, a BoD service instance may be requested in advance, and a reservation is expected to last from days to years. A minimum time period is required between the request and the actual provisioning of the service, which is due to the level of automation in the resource allocation process. The provisioned paths can be either unprotected, partially or fully protected. All services with full protection require the set-up of two completely separate paths from source to destination, including the physical layer, so as to survive failures even in the case of events such as fibre cuts.

The GN2 BoD system is composed of the modules presented in Figure 2. Its architecture is modularized in order for individual modules to be easily upgraded or replaced, and the communication between these modules is heavily based on web services interfaces.

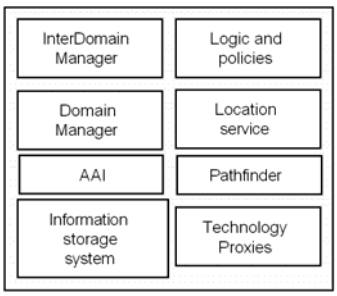

Figure 2. AutoBAHN system Modules (Campanella et al., 2006)

Inter-domain Manager (IDM)

The IDM module, similarly to the AMPS Inter-Domain subsystem is the key component of the AutoBAHN system. As the only ingress point to the system, it is responsible for receiving BoD service requests directly from a user or application, or indirectly by another domain and is then responsible for the admission and instantiation of these requests. The IDM functionality is also independent of the underlying individual per-domain implementations and can apply to both automated and manual per-domain BoD provisioning.

When service end points are in different domains, the IDMs involved will cooperate on a peer-to-peer basis to create the requested end-to-end path. Each domain can independently choose the policies and technologies for the BoD service. The peering model also allows scaling to a large number of domains and eases the synchronization constraints of the implementation.

Upon receiving a reservation request, the IDM module uses the Inter-domain pathfinder to contact the next domain and establish an end-to-end path. In order to commit to the reservation request, IDMs use a chain model, similar to the one used by RSVP (Resource ReSerVation Protocol).

The IDM relies on the local DM to implement the service requested in the form of a provisioned circuit. The DM deals, through the Technology Proxy, with the physical details of the particular network domains and the different technologies used to implement the BoD circuit. For more details refer to DJ3.3.1 (2008) and FSIDM (2007).

If needed, the IDM module also communicates with the AAI service, in order to authenticate the AutoBAHN user and the associated privileges for the BoD service, always applying the local domain rules and policies.

Domain Manager (DM)

The Domain Manager (DM) module is responsible for instantiating the reservations within the single domain where it has been deployed. It has detailed knowledge of the topology of its domain and participates in the inter-domain pathfinding process by examining the feasibility of providing an end-to-end path within its local domain. For the actual, technology-specific configuration to take place, the DM contacts the Technology Proxy module.

Technology Proxy (and Resource Modelling) module

The Technology Proxy module performs the translation of requests received by the DM (in abstract network language) into vendor- or equipment-specific configurations. It is the AutoBAHN module that is expected to be more frequently upgraded and replaced for various deployments depending on the existing low-level technologies in each domain. The proxy module may configure the network via an existing Network Management System (NMS) or act as a GMPLS agent; consequently it does not act directly upon the network, but relies on an intermediate control layer.

Other modules

The *Policy module* contains all the rules and policies which are available for use by other modules when inspecting and elaborating on a request. The rules are collected in a single module for easy maintenance, modification and to enforce coherence.

The *Pathfinder* module contains the algorithms and the logic to search for a path that satisfies each BoD reservation request according to specific sets of constraints, algorithms and policies. The module actually consists of two independent blocks, one for inter-domain pathfinding and one for intra-domain pathfinding, which are respectively incorporated in the IDM and DM modules.

The *Information Storage System* and the *Location Service* function mainly offer support to the other modules. The Information Storage System is responsible for providing storage, archival and database functionalities for data explicitly relevant to the BoD system, while the Location Service locates the addresses of all type of services and modules.

The Inter-domain Pathfinder

The inter-domain Pathfinder module is responsible for producing a list of paths that satisfy the reservation requests. It receives a set of parameters from each reservation request, takes into account the local policies and the information received by other IDMs and computes a list of candidate end-to-end paths over which the request can potentially be implemented.

While in a standard routing protocol, signaling, topology discovery and update, routing policies and route computation are tightly coupled, in the BoD architecture, these functionalities are decoupled and placed in different modules.

- Inter-domain topology signaling is placed in the IDM.
- Inter-domain topology discovery and update is performed at the IDM, which receives updates from all other IDMs and creates an abstracted inter-domain topology.
- The routing algorithms operate on the abstracted topology produced by the Pathfinder.
- Policies are applied both in the IDM and in the Pathfinder. Various set of policies exist, such as signaling policies (filtering) and request handling policies (which includes user access policies) in the IDM and path computation policies (algorithm parameter values) in the Pathfinder.

The Pathfinder module, located in the source domain, computes the complete path from the source domain to the destination domain. Then, the source domain IDM contacts the next domain in the path, which in its turn contacts the next one using the announced path (or using its own routing computation in case of a mismatch) in a chained model.

The source domain's view of the abstracted inter-domain BoD topology is built from the announcements from other IDMs. For each specific request, the Pathfinder routing algorithm uses the abstracted topology to return a list of feasible paths based on the parameters of the reservation request. In order to return multiple paths, k-shortest path algorithms (Shier (1979), Eppstein (1994)) are used.

Since the Pathfinder uses a chained model, it is preferable to have the source domain compute the path and announce it to the subsequent domains. Even if all domains may re-compute the same end-to-end path, announcement of the path is useful in order to reduce load and ensure consistency. In case the chained reservation process fails at some point, the source domain will use a different path (either from the set already computed by the Pathfinder or by calling the Pathfinder module again) and start again.

The Pathfinder implementation is based on the OSPF protocol to distribute information about the links and build the database necessary for path computation. In order to carry Traffic Engineering information using OSPF, the Pathfinder module will use Opaque LSAs of type 9, 10, 11 according to RFC 2370 (Coltum, 1998).

Reservations

A service request is submitted at the source domain (Home Domain) and is examined using a chain communication model, so that each domain is only in contact with its direct neighbour. When a reservation request is submitted, the Home Domain performs a local validation procedure and executes inter-domain pathfinding in order to identify a list of feasible paths. If it has sufficient resources available to perform the reservation, it propagates the request forward to the next domain of the selected inter-domain reservation path. Attached to the propagated request are the reservation constraints, which are subsequently propagated all the way to the last domain on the reservation path. A reverse propagation of the status of the request then takes place, so that domains are informed whether the request was accepted by the subsequent domains or not.

Implementing a brokering and monitoring service

An integral part of the AutoBAHN service is the monitoring of the provisioned end-to-end connections. This section describes in more detail the architectural design and decisions taken in order to implement this module of the system (Bouras et al., 2008).

The basic concepts for the End to End (E2E) Monitoring were:

- *Intradomain link:* Intradomain link is a link between two nodes (routers or switches) that both belong to the same domain.

- *Virtual link:* Virtual link is the conceptual link which connects two border nodes (routers or switches) that both belong to the same domain. A virtual link is the result of the conjunction of many Intradomain links. Virtual links are the kind of information that a domain presents to the outside world about its internal state: it does not propagate detailed information about its internal links; instead it provides aggregated information about Edge to Edge connections represented by Virtual links.

- *Interdomain link:* Interdomain link is a link between two border nodes (routers or switches) of two different domains. Interdomain links are typically monitored by both domains they connect to, but the case where only one domain monitors an interdomain link can also be handled.

- *E2E (End to End) link:* E2E link is the conceptual link which connects two end points from different domains. An E2E link is the result of the conjunction of many Interdomain and Virtual links.

Each domain provides information about the links it monitors (these links can be either Virtual links or Interdomain links). The E2E Monitoring system computes the overall status of an E2E link by aggregating the status information from the involved domains.

The overall architecture of the E2E monitoring system consists of the following modules:

- *Low-level reference implementation:* This module can be a secure scripting server, if the underlying network is based on Ethernet technology, or it can be a technology proxy to the NMS (Network Management System) if the underlying network is based on SDH. It is responsible for the low-level, technology specific communication with the network devices of the domain in order to check the status of the physical links.

- *DM monitoring module:* The monitoring module of the DM is responsible for monitoring the status of both Virtual and Interdomain links and provide this information to the Visualisation server via the Database.

- *Database:* The database functions as an intermediary level between the DM monitoring module and the visualization server. It stores persistent information that can be asynchronously retrieved by the upper layers, thereby insulating low-level functions from user requests.

- *Visualisation server:* The visualisation server collects monitoring information from the DMs (through the corresponding monitoring module) of the domains involved in the E2E monitored link and presents monitoring information to the user/administrator through a graphical interface.

As displayed in Figure 3. the modular architecture in combination with the usage of standards such as SOAP for the implementation of the interfaces allows modules to be easily and transparently changed or upgraded, allowing the full or partial deployment of the monitoring system in various NREN environments.

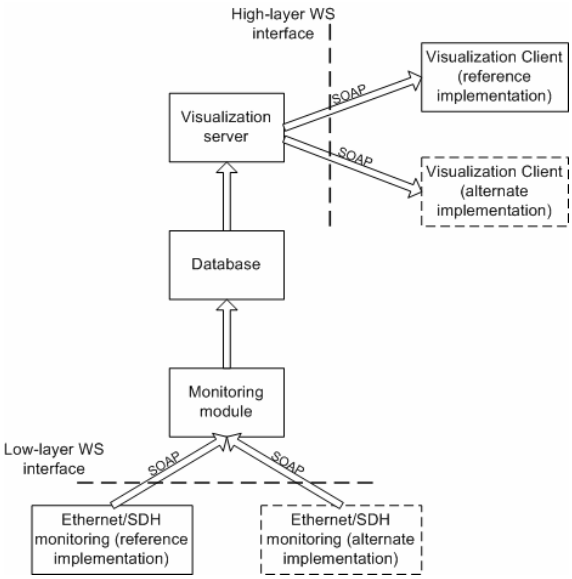

Figure 3. Information flow for the monitoring module

During the monitoring of an end-to-end Bandwidth on Demand circuit over Ethernet Infrastructure the following steps take place:

The IDMs and DMs of the involved domains cooperate, as described in the relevant section above, in order to establish an end-to-end Bandwidth on Demand circuit over the existing infrastructure. The results of this establishment will be an end-to-end circuit in the abstracted topology of the involved domains. This end-to-end Bandwidth on Demand circuit or E2E link for shortness is the input for the E2E monitoring system which is responsible to monitor the status of the E2E link. The E2E monitoring system extracts from the abstracted topology the domains and the corresponding interdomain and virtual links which comprise the E2E link. In repeated time intervals the E2E monitoring system polls the DM monitoring modules of the involved domains and aggregates all the status information. Each DM monitoring module reports aggregate status about the links that it monitors to the E2E monitoring system. In order for the DM monitoring module to check the status of an interdomain link or an intradomain link, it has to communicate (via web services) with the low-level technology specific implementation. For example, a reference implementation has been developed for Ethernet, where communication is done using a Secure Scripting Server. This communication for security reasons in based on SSH. The Secure Scripting Server communicates with the corresponding routers / switches in order to check the status of the physical link that corresponds to an intradomain or interdomain link. An interdomain link can be monitored by either one of the domains connected by this interdomain link or by both connected domains. In the latter case the E2E monitoring system is responsible for combining the monitoring information from both involved domains into a unified status for the link. The E2E monitoring system computes the overall status of an E2E link by aggregating the status information of the involved domains and presents the overall status of an E2E link through the visualization server. The communication between the visualization server and the DM monitoring system is also based on web services.

The E2E monitoring system in its current version provides information for two monitoring parameters: The Operational State, which is derived from the operational state of the involved physical devices, and the Administrative State, which reflects the management processes performed by the domains.

Demonstrations

The AutoBahn team has made various demonstrations of the AutoBahn BoD service during the last years. During the GEANT2 Workshop in January 2008 the following demonstration took place: Multiple parallel connections between end points were supported by data plane, while domains' resource management was fully automated and performed without administrator attention. The following figures show the AutoBahn GUI and topology of the demonstration.

Figure 4. AutoBahn GUI

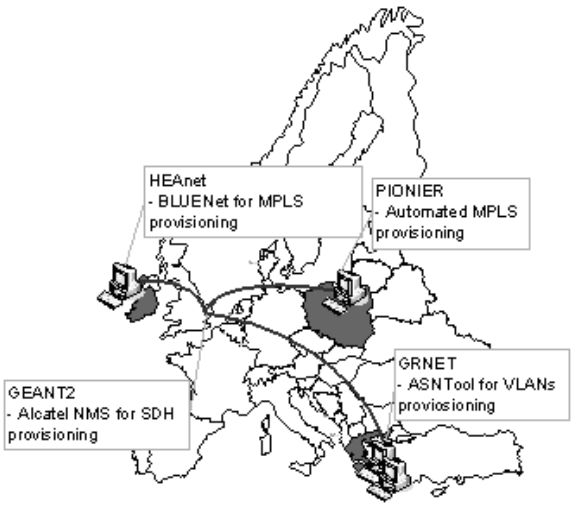

Figure 5. Demonstration topology (GEANT2 Workshop in January 2008)

In addition in November 2007, AutoBAHN was presented at the SuperComputing'07 event, where the European test environment was interconnected with a similar research environment in the USA. Four European domains were involved and collaborated with the USA system. The following figure shows the topology of the demonstration and some performance graphs.

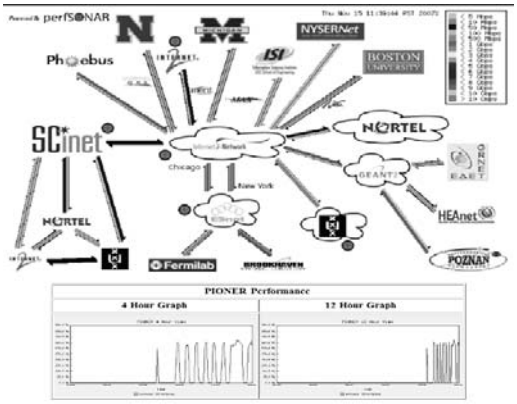

Figure 6. Demonstration topology (SuperComputing'07 event)

FUTURE RESEARCH DIRECTIONS

GEANT and the European NREN community intend to continue working on and expand the described services in the future, and in particular in the context of the upcoming research projects.

One of the most emerging issues in the area of L2 QoS is the standardization and interoperability of services like the AutoBAHN service of GEANT. This will allow the simplification of the deployment of the end to end L2 BoD services and will reduce the cost of such service deployment both in terms of equipment cost and map power cost. Standardization and interoperability issues will have to be studied in two directions:

- Standardization and interoperability in the interface between similar L2 QoS services: In this direction protocols must be standardized for the communication of IDM entities of various L2 QoS services.

- Standardization and interoperability in the interface between the service and the domains participating to the L2 QoS cloud. In this direction protocols must be standardized for the communication of DM entities and the technology proxy of a participating domain to the L2 QoS cloud.

The current experience suggests the standardization of the above interfaces to be based on XML. This will allow the easy implementation of the relative protocols and will increase the interoperability among different implementation platforms.

Furthermore, GEANT intends to further investigate L3 QoS provisioning as a production-level service and streamline the reservation procedure so that the users can more seamlessly take advantage of dynamic network resource allocation.

CONCLUSION

Designing and implementing the services for Quality of Service at layer 3 and Bandwidth on Demand in such a large scale for GEANT has proven to be a complex and challenging task. At the completion of the GEANT project, both services were operating at a functional level, with multiple deployments across European NRENs. For the AutoBAHN system, the network abstraction and a network description language, technology stitching and inter-domain pathfinding were the most significant areas where research and development effort was required. The AMPS system spawned the creation of a generalized persistency layer in the form of cNIS database and was successful in interoperating with several pre-existing single domain tools, which it incorporated in a multiple federation domain setting. Furthermore, the introduction of a complex monitoring solution across multiple heterogeneous

domains was achieved by modularizing the system using several levels of abstraction and by isolating compartments of the overall task.

It is important to note that in each case, a domain can adopt the parts of the AMPS or AutoBAHN and the accompanying monitoring module system that it sees fit, and it can put in place its own custom components in the remaining places by obeying to the standardized provided interfaces.

Current development and deployment has focused on specific popular technologies such as IP, ethernet and SDH. However the design concepts are absolutely valid for a heterogeneous technology environment, and the design has been general enough to accommodate heterogeneous technology environments from the start.

REFERENCES

ADS. (2006). *GEANT2, deliverable GN2-04-153v4 "AMPS – Design Specification."*

Bouras, C., Campanella, M., Przybylski, M., & Sevasti, A. (2003, February). QoS and SLA aspects across multiple management domains: the SEQUIN approach. *Future Generation Computer Systems archive, 19*(2), 313 – 326.

Bouras, C., Haniotakis, V., Primpas, D., Stamos, K. & Varvitsiotis, A. (2007). AMPS - ANStool: Interoperability of automated tools for the provisioning of QoS services. *TERENA Networking Conference 2007*, Lyngby, Denmark, 21 - 24 May 2007.

Bouras, C., Gkamas, A., & Stamos, K. (2008). Monitoring End to End Bandwidth on Demand Circuits over Ethernet Infrastructure. In *Seventh International Network Conference – INC 2008*, Plymouth, UK, 8 - 10 July 2008.

Campanella, M., Krzywania, R., Reijs, V., Sevasti, A., Stamos, K., Tziouvaras, C. & Wilson, D. (2006). Bandwidth on Demand Services for European Research and Education Networks. In *1st IEEE International Workshop on Bandwidth on Demand*, 27 Nov 2006, San Francisco.

Coltum, R. (1998 July). The OSPF Opaque LSA Option. *RFC 2370*.

DJ3.3.1. (2008). *Definition of Bandwidth on Demand Framework and General Architecture (DJ3.3.1).*

DJ3.2.2.3. (2007). *Third review of Bandwidth on Demand related technologies (DJ3.2.2.3).*

Eppstein, D. (1994). Finding the k shortest paths. In *35th IEEE Symp. Foundations of Comp. Science, Santa Fe*, (pp. 154-165).

FSIDM. (2007). *Functional Specification of GÉANT2 Inter-domain Manager (IDM) Prototype.*

G.807. (2001 July). ITU-T Rec. G.807/Y.1302. *Requirements for Automatic Transport Networks (ASTN).*

G.8080 (2001). ITU-T Rec. G.8080/Y.1304. In *Architecture for the Automatically Switched Optical Networks, November 2001 and Amendment 1, March 2003*.

GÉANT2. (2009). *GÉANT2: The pan-European R&E network*. Retrieved from http://www.geant2.net/

Heinanen, J., Baker, F., Weiss, W. & Wroclawski, J. (June 1999). *Assured Forwarding PHB Group, RFC 2597.*

Jacobson, V., Nichols, K. & Poduri, K. (1999 June). *An Expedited Forwarding PHB, RFC 2598.*

Lehman, T. (2006). InterDomain Peering and Provisioning via GMPLS and Web Services. In *6th Global Lambda Workshop*, Tokyo, Sep. 1-13, 2006.

Nichols, K. & Carpenter, B. (2001 April). *Definition of Differentiated Services Per Domain Behaviors and Rules for their Specification, RFC 3086.*

RFC3945 & Mannie, E., (2004 October). IETF RFC 3945. *Generalised Multi-Protocol Label Switching (GMPLS) Architecture*.

RFC3031 & Rosen, E., et al (2001, January). *IETF RFC 3031, Multiprotocol Label Switching Architecture.*

Shier, D. (1979). On algorithms for finding the k shortest paths in a network. *Networks, 9*(3), 195-214.

The Internet Singularity, Delayed: Why Limits in Internet Capacity Will Stifle Innovation on the Web. (2007, November). Mokena, IL: Nemertes Research.

Leung, F., Flidr, J., Tracy, C., Yang, X., Lehman, T., Jabbari, B., Riley, D. & Sobieski, J. (2006). The DRAGON Project and Application Specific Topologies. In *Broadnets 2006*, San Jose, CA.

Guok, C. (2005). ESnet On-Demand Secure Circuits and Advance Reservation System (OSCARS). In *Internet2 Joint Techs Workshop*, Salt Lake City, Utah, February 15, 2005.

Formatted: English (U.K.)

Wu, J., Campbell, S., Savoie, J. M., Zhang, H., Bochmann, G. v. & St.Arnaud, B. (2003, Sept. 7-11). User-managed end-to-end lightpath provisioning over CA*net 4. In *Proceedings of the National Fiber Optic Engineers Conference (NFOEC)*, Orlando, FL, (pp. 275-282).

Inocybe technologies Inc. (n.d.). *Argia product*. Retrieved from http://www.inocybe.ca/

Cavazzoni, C. (2007). MUPBED Overview and Architecture. In *TERENA Networking Conference 2007*, Copenhagen, Denmark.

VIOLA project - Vertically Integrated Optical Testbed for Large Applications in DFN. (n.d.). Retrieved from http://www.viola-testbed.de/

RFC4655, Farrel, A., & Vasseur, J.-P. (2006 August). A Path Computation Element (PCE)-Based Architecture. *IETF RFC 4655*.

Formatted: French (France)