Experiments and Enhancements for IP and ATM Integration: The IthACI Project

Ilias Andrikopoulos and George Pavlou, CCSR, University of Surrey
Panos Georgatsos, Nicholas Karatzas, and Kostas Kavidopoulos, Algonet S.A., Greece
Jürgen Röthig and Sibylle Schaller, CCRL, NEC Europe Ltd., Germany
Dirk Ooms, Alcatel Bell, Belgium
Pim Van Heuven, IMEC, University of Gent, Belgium

ABSTRACT

IthACI has been a European project in the ACTS framework concentrating on fast layer 2 forwarding methods for IP traffic based on labeled flow mechanisms. The approach is also known as IP switching and is considered promising for enhancing IP performance. Several flavors of IP switching have been proposed by various vendors (e.g., IP Switching by Ipsilon, Tag Switching by Cisco, ARIS by IBM, IPSOFACTO by NEC), all of them different and not interoperable. IP Switching has been adopted by the IETF under the umbrella of Multi-Protocol Label Switching (MPLS).[1] Although MPLS has made remarkable progress recently, a number of issues remain largely open for further investigation. The scope of the IthACI project was to address such issues and propose solutions. The issues addressed were multicast, QoS, resource management, and mobility support in a multicast environment. IthACI conducted both theoretical and experimental work. Three network islands, each based on a different flavor of IP switching, were set-up and the interoperability of these different IP switching/MPLS flavors were investigated and demonstrated.

INTRODUCTION

Since its inception around 1990, asynchronous transfer mode (ATM) network technology has been regarded as an antipode to existing Internet Protocol (IP) technology. ATM has started being deployed by traditional voice carriers (telephone companies), while IP is deployed by carriers of data traffic. ATM is considered to be fast, but complex, expensive, and ineffective for short-lived applications, mainly due to its connection-oriented nature. IP is regarded as simple, mature, well proven and accepted over the years, but missing QoS functionality and resulting in slow speeds due to the implementation routing functionality in software. Efficient methods for combining IP and ATM technology and transporting IP traffic over ATM backbone infrastructure have been considered. The result is known as "IP Switching" — a kind of IP router with IP protocol functionality that employs ATM hardware for efficient data forwarding.

Originally, various flavors of IP Switching were proposed: Ipsilon IP Switching, Cisco Tag Switching, IBM ARIS, Toshiba CSR, NEC IPSOFACTO, just to name a few. This prompted the Internet Engineering Task Force (IETF) to address a standardized approach through a working group on multiprotocol label switching (MPLS).

IthACI [1] (*I*nternet and *the* ATM: *C*onvergence and *I*ntegration) was a European Advanced Communications Technologies and Services (ACTS) project, which ran from March 1998 to Dec 1999 with the overall scope to evaluating and contributing to the different technologies that permit the efficient transport of IP traffic over, private or public, ATM backbone infrastructure. In this context, the project addressed the requirements for efficient IP multicasting, accommodation of QoS demands, mobility in a multicast environment, and resource management. It subsequently undertook enhancements of existing IP switching solutions with respect to the previous features, and generated recommendations based on experience gained from implementation and experimentation.

Besides the functional enhancements, the project's main goal was to influence the actual standardization process in the area of IP switching, and thus to work within and bring the project results to the IETF MPLS working group.

[1] *The terms MPLS and IP switching are used interchangeably throughout this article.*

0163-6804/01/$10.00 © 2001 IEEE

The article describes the multicast, QoS, and resource management enhancements developed in the IthACI project. These enhancements led to certain advances in these areas, fed back to the MPLS working group. The tests carried out for validating the developed enhancements are also presented, summarizing the produced results and demonstrations and highlighting the experience gained.

PROJECT OVERVIEW AND ADOPTED TECHNOLOGIES

The IETF established a working group on MPLS in early 1997 in order to consider methods for label swapping based forwarding (*label switching*) in conjunction with network layer routing. Although restricted to neither any layer 2 technology nor to a specific layer 3 routing protocol, IP over ATM[2] (IP switching) is implemented as an important incarnation of MPLS.

First, the IETF MPLS working group built up a general MPLS framework and architecture [1]. So far, the main focus of the MPLS working group was on interoperability aspects, especially the Label Distribution Protocol (LDP), which provides the signaling facilities among the label switching routers (LSRs) and coordinates the distribution of label bindings aiming at setting up label switched paths (LSPs). Recently, work on traffic engineering issues in MPLS also began.

There are a lot of issues in MPLS networks that require further investigation. Issues in the areas of multicast, QoS provisioning and resource management stimulated the IthACI project work [2]. Considering three existing IP switching technologies, the project undertook the design, implementation, and experimentation of a number of enhancements in these areas.

The various IP switching techniques used within the IthACI project originate from three different sources: Tag Switching (available as a product by Cisco), IPSOFACTO/LCATM (a prototype for IP switching developed by NEC), and YALSA (a new IP switching prototype specially designed for multicast within the IthACI project). A description of adopted IP switching technologies is given in the following sections.

In order to achieve its objectives, the project did set up three independent islands, each employing one of the above-mentioned existing (proprietary) IP switching technologies. The islands and technologies were NEC's IPSOFACTO, the "Green Island;" CISCO's Tag Switching, the "Blue Island;" and Alcatel's YALSA, the "Red Island."

Project work resulted in enhancing the considered technologies with features in the areas of multicast, QoS provisioning, resource management, and mobility in a multicast context, in order to provide added value to existing solutions. The islands were interconnected to demonstrate interoperability of the IP Switching techniques as well as the cooperation of the developed enhanced features.

It should be noted that although the developed enhancements were implemented in these technologies, the design was not limited to them. Project work followed the emerging specifications from the IETF MPLS working group, and contributed to the ongoing MPLS work. Moreover, the adopted baseline technologies themselves moved by their manufacturers to comply with the MPLS specifications.

IPSOFACTO/LCATM — IPSOFACTO [3] is a soft-state traffic-driven technique where each switch makes independent decisions about shortcuts for data flows.

A node that wants to send a packet selects an unused virtual circuit identifier (VCI) on the appropriate outgoing link. The receiving IPSOFACTO switch is configured to send all cells with an unassigned VCI to the switch controller. The switch controller reassembles the IP packet and then uses the IP routing table to decide how the packet should be forwarded. In this phase, the IPSOFACTO switch acts as a normal IP router. An unused VCI on the appropriate outgoing link is selected, and the shortcut path is created when the switch controller informs the underlying switch to shortcut the incoming VCI to the outgoing VCI (*route once, switch many*). It should be clear that IPSOFACTO uses implicit upstream allocation: the upstream node selects a new VCI from a pool of unused VCIs. No label distribution protocol is necessary. The upstream node sends the first packet on an unused VCI. The downstream node reassembles this first packet to discover the label semantics (e.g., destination address) attached to the incoming VCI. The receiving switch learns the label semantics by inspecting the first packet of the flow. The first packet of a flow paves the way for the following packets. Note that shortcut paths are never established for IP control messages.

A characteristic of IPSOFACTO is that it was designed in the context of multicasting and multicast routing right from the start, addressing the Distance Vector Multicast Routing Protocol (DVMRP) and especially Protocol Independent Multicast (PIM). The multicast routing database is used to establish multicast forwarding state in the switch controller. IPSOFACTO maps this state to a (number of) point-to-multipoint VC(s) within the switch. A large part of the specification of IPSOFACTO deals with mapping PIM sparse mode (PIM-SM) and dense mode (PIM-DM) to the ATM switching paradigm.

In the course of the IthACI project, the NEC IPSOFACTO implementation evolved to label switch controlled ATM (LCATM), a prototype that supports native IP multicast over MPLS as proposed by NEC to the IETF [4, 5]. LCATM employs the original IPSOFACTO ideas for multicast traffic, and uses the standard LDP protocol and methods for MPLS unicast traffic. In the rest of this article the enhanced IPSOFACTO prototype is referred to as LCATM.

YALSA — Yet Another Label Switching Architecture is a flexible, modular architecture that allows implementing and testing various label switching methods. YALSA uses a dedicated lightweight label distribution protocol, with signaling between neighboring LSRs to advertise

Project work resulted in enhancing the considered technologies with features in the areas of multicast, QoS provisioning, resource management, and mobility in a multicast context, in order to provide added value compared to existing solutions.

[2] *Not to be confused with Classical IP over ATM.*

Currently several multicast routing protocols are being implemented and standardized. These protocols expose different tree set-up and maintenance characteristics, which yield that some multicast routing protocols combine better with IP Switching than others.

the association between a forwarding equivalence class (FEC) and a label.

The LSRs in the YALSA Island can be configured to do label advertisement in either downstream or downstream-on-demand mode. The LSPs are set up in a distributed fashion — each LSR can independently decide to start a (partial) LSP — and the setup can be initiated by different triggers. For multicast LSPs the following triggers were tried: multicast routing protocol messages, changes to the multicast forwarding cache, and IGMP snooping. Consequently, YALSA can be configured to work in either traffic or topology-driven mode.

Tag Switching — Cisco's Tag Switching was one of the first commercial attempts for shortcutting IP traffic on layer 2 paths. Cisco's Tag Switching is topology-driven and uses downstream allocation on demand[3] for label assignment. For each entry in its routing table, the upstream node asks the downstream node to provide a tag. The downstream node can either first ask its own downstream node for a tag and then reply to the upstream node, or reply immediately and then ask its downstream node for a tag.

The Tag edge routers add tags (a synonym for labels) to packets and participate in the L3 routing protocols. Tag edge routers can also perform additional L3 functionality such as security. The Tag Switches or Tag switch routers (TSRs) are the core nodes which forward packets based on tag values and also participate in the L3 routing protocols. TSRs do not add or remove tags; they only switch tags (translate the incoming tags to outgoing tags). The Tag Distribution Protocol (TDP) is the control protocol used by Tag switches and edge routers to request or distribute tag bindings. When distributing tag bindings, TDP also includes a hop count so the Tag edge routers can decrement the time to live (TTL) before transmitting the packets through the Tag switching network. TDP requests can be initiated by any node at any time.

When the switch controller has communicated upstream bindings and received downstream bindings, it can instruct its switch to shortcut incoming tags to the appropriate outgoing tags.

Tag Switching supports traffic engineering by allowing the establishment of explicitly routed (traffic engineered) paths and the routing of incoming traffic on them based on extended filters on source, destination IP addresses and other information in the IP header.

MPLS IN A MULTICAST ENVIRONMENT

The multicast enhancements undertaken by the IthACI project aimed to apply MPLS-style shortcut techniques to IP multicast to optimize network performance without any modification to the end-user environment or to existing IP multicast protocols. The shortcut point-to-multipoint ATM connections follow dynamically the tree topology changes that are the result of IP hosts joining or leaving a multicast group.

At the time the IthACI project was defined, the IETF had just created a working group on MPLS. Within the working group, the focus was solely on unicast forwarding, with the application of label switching to IP multicast traffic not yet addressed. Within the project, network functional models and algorithms were studied to enable MPLS for multicast flows. A prototype for MPLS multicast was defined, designed, and implemented to show the feasibility and scalability of label-switched multicast flows. This new approach to IP/label switching advanced the IETF MPLS standardization process.

Currently several multicast routing protocols are being implemented and standardized: DVMRP, PIM-SM, PIM-DM, Multicast Open Shortest Path First (MOSPF), and Core Based Trees (CBT). These protocols expose different tree setup and maintenance characteristics, which indicate that some multicast routing protocols combine better with IP switching than others: for example, flood and prune protocols (DVMRP, PIM-DM) will create highly dynamic L2 trees, resulting in a lot of label advertisement signaling and label consumption. Also, some multicast routing protocols (e.g., PIM-SM, CBT) allow the use of a shared tree for all sources, which leads to less label consumption, but potential merging problems when ATM is the underlying layer.

For multicast, the project selected PIM-SM as the routing protocol. This protocol is designed to efficiently route IP multicast datagrams to sparsely distributed wide area groups. The selection of PIM-SM over, say, DVMRP or PIM-DM was mainly because of scalability issues. The project has designed means for mapping multicast trees/routes to shortcuts and has elaborated on multicast issues in MPLS networks in general. The multicast enhancements have been realized in the Green and Red Islands based on Alcatel's YALSA and NEC's IPSOFACTO/LCATM technologies.

Each of the Red and Green Islands used its own method to do the mapping of multicast streams onto layer 2 (L2) connections. IPSOFACTO/LCATM used an implicit label advertisement method, YALSA a dedicated protocol. A dedicated protocol (e.g., LDP) has the disadvantage of needing to be developed from scratch. Label advertisements piggybacked onto PIM join/prune messages (Tag Switching) have several disadvantages. The periodicity of the messages creates more signaling than required, only downstream mode is possible, the deployment of each new multicast routing protocol demands adaptations to allow piggybacking, the method cannot be used for flood and prune protocols, and finally, it limits the scope of the MPLS domain in mixed (LSR and non-LSR) multi-access networks.

Besides correct interworking of IP multicast and IP switching in the separate islands, another goal of the project was to demonstrate the interoperability between the different multicast solutions, either on layer 2 or, as a fallback solution, on layer 3. Interoperability was demonstrated for

[3] *That is, when ATM is used as the underlying switching technology. The non-ATM-specific part of Tag Switching allows for other allocation strategies too.*

the Alcatel and NEC (Red and Green Islands). Multicast shortcut paths were dynamically established between both prototypes, crossing the border between the two IP switching technologies on layer 2 (Fig. 1).

The multicast implementation in LCATM follows the traffic-driven methodology to switch IP multicast flows. When the first packet of a multicast flow is detected, LCATM sets up point-to-multipoint ATM VCs according to the corresponding entry in the multicast forwarding cache maintained by the PIM-SM routing software.

The LCATM prototype supports native IP multicast over MPLS as proposed by NEC to the IETF [4, 5]. Basically this scheme maps the IP tree onto a per-source point-to-multipoint ATM tree, even for a core-centered approach like PIM-SM, in a traffic-driven way. In order to avoid PIM piggybacking, two modes of label distribution are advocated: a downstream approach equivalent to the unicast case and an upstream implicit approach, derived from the NEC IPSOFACTO architecture [3], which speeds up the ATM tree establishment.

When a PIM join/prune message affects a multicast forwarding entry, all the corresponding ATM VCs are modified accordingly. Two special cases must be handled in a particular way by LCATM: the transition from a shared tree to a source specific tree and the rendezvous ponit (RP). The protocol used by the IP switching module to manage the ATM hardware conforms to the General Switch Management Protocol (GSMP).

When a new packet is received (interaction 1 in Fig. 2) and no multicast forwarding cache (MFC, containing per-source multicast entries) entry exists, the kernel (in our case it is a Linux kernel) asks the multicast routing daemon (2) about the route to assign to this packet, and stores the incoming packet in a kernel buffer. Eventually the kernel receives the daemon reply, and a new MFC entry is created (3). Subsequent packets will be forwarded without any request to the daemon.

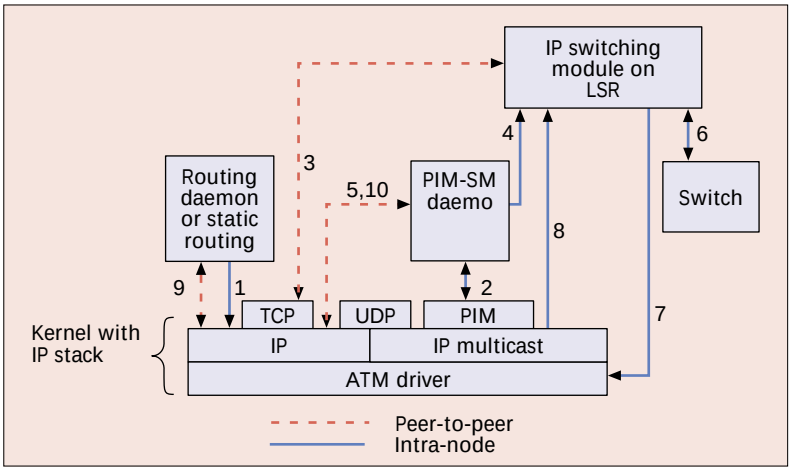

Figure 1. *Building blocks and interfaces on an LSR to support IP multicast.*

The LCATM IP multicast support provides an MPLS extension of the MFC: some hooks are introduced in the IP multicast forwarding path to store the multicast label information base (MLIB). The MLIB entry contains mainly the IP multicast flow parameters (IP source address, group address), some statistics provided by GSMP, the input/output ports and VPI/VCI and the GSMP priority as class of service (CoS) indicator. When a new MFC entry is created, the hook pre_forward hook is called (4) to create a new MLIB entry (5) and the Real-Time Transport Protocol (RTP) flow detection is applied to set-up the CoS indicator.

In the downstream mode, the packet is then forwarded on the routed path to all the next hops (6) and a post_forward hook is called afterward (7) to send a message to the upstream node indicating which VPI/VCI to use for the correspond-

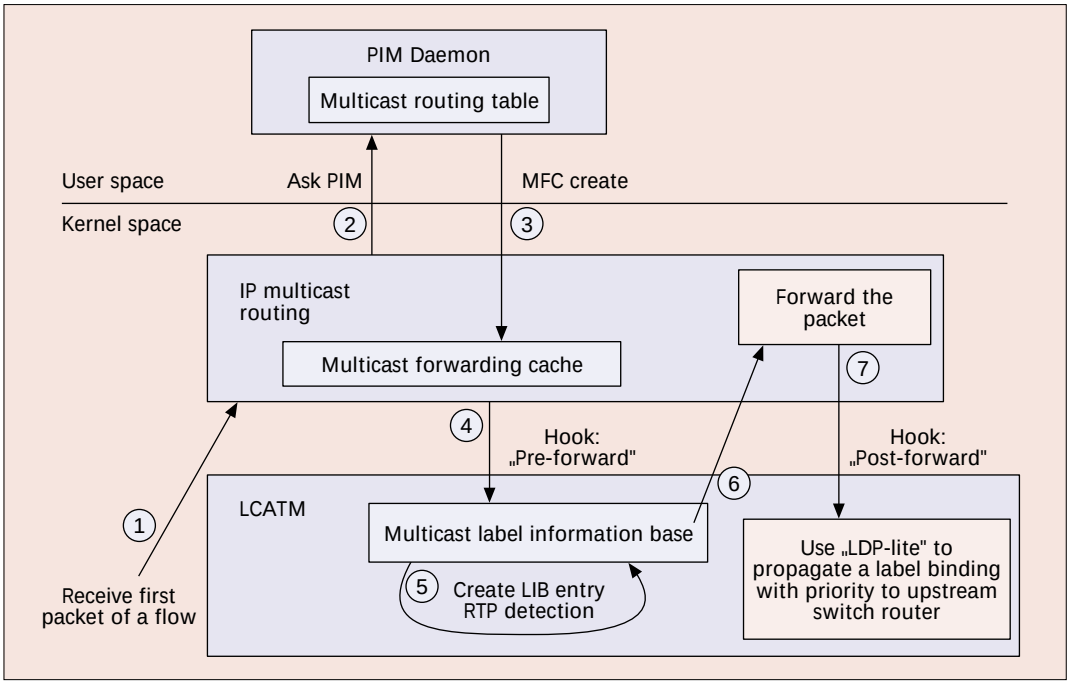

Figure 2. *LCATM and RTP flow detection (core router, downstream).*

Resource management enhancements aim at providing capabilities for using network resources efficiently, from both the network's and the user's perspective.

ing sender, group (S, G) traffic. The protocol used is a minimal LDP implementation called LDPLite. When an LDPLite mapping message containing an outgoing label for an outgoing port is received, a branch of the MFC entry, a branch is added to the corresponding ATM point-to-multipoint VC on the ATM switch. The subsequent packets will be switched in the hardware. Note that this implementation is flexible enough to support simultaneous layer 2 and layer 3 IP multicast forwarding, as defined in [5].

QoS and Resource Management in MPLS

The IthACI project, following the emerging trends and efforts in QoS provisioning and traffic engineering, specified, designed and implemented a number of QoS and resource management enhancements on top of IP switching technologies. IthACI looked at the issue of QoS, advancing the state of the art in the area in a number of aspects.

The project has looked at the issue of QoS and resource management from two viewpoints: from the network technology provider's (vendor's) viewpoint and from the service provider's viewpoint. To these ends, a number of enhancements were designed. Although these enhancements are valid for all IP switching technologies considered in the project, they are integrated and tested in separate islands according to the particular focus and characteristics of the islands. This is because the focus of the project is primarily to validate the concepts underlying a particular enhancement, not to specify common standards for their application in IP switching technologies (currently specified in IETF). Validated enhancements can then be ported in the context of other IP switching technologies (and in the emerging MPLS standards in general).

QoS and Resource Management from the Equipment Vendor's Perspective

From the vendor's viewpoint the project has designed enhancements in the core of network elements for facilitating QoS provisioning. Resource management enhancements aim at providing capabilities for using network resources efficiently, from both the network's and the user's perspective.

Quality of Service for RTP Flows — When the time the project started, there was little work in the area of QoS provisioning for multipoint multimedia applications. Providing QoS in the network is essential for applications like video, audio, and conferencing in the Internet of the future. The project focused on the detection of RTP flows in order to provide them with a special QoS [6]. RTP is the current standard for real-time transmission over the Internet and is expected to gain more importance with the growth of multimedia applications.

RTP flow detection was implemented in NEC's IP Switching Island (LCATM, Green Island). Rather than providing per-flow QoS, a task that can be done with Resource Reservation Protocol (RSVP) but incurs complexity and scalability problems, the primary goal is to divide traffic into CoSs, thus approximating a differentiated services (DiffServ) solution. The existing flow detection mechanisms in the IPSOFACTO architecture were further enhanced with an RTP flow detection mechanism. It should be noted here that RTP-based applications are very important because they are actually real-time and therefore requires better service than traditional data applications.

Furthermore, flow differentiation among the RTP flows was done using some clever assignment of flows with specific RTP profiles, which were mapped to the appropriate ATM switch service classes. This mechanism comprises two main sub-tasks:

- First, the detection of RTP flows is based on the header validity check described in [7]. This is a "weak test," and may therefore fail.
- Second, the provisioning of special QoS for that RTP traffic to be derived from the RTP profile information [8]. This may be, in the simplest approach, just a priority, but it is also possible to make a bandwidth reservation for very special flows according to the RTP header information.

Joint MPLS/IPSOFACTO Platform Architecture — A further goal within the Green Island was the development of a common prototype for the coexistence of a topology-driven unicast-oriented IP switching technique (compliant to MPLS specifications) and a flow-driven IP switching technique (LCATM proprietary). The basic idea of the combination of standard MPLS and LCATM was to use MPLS as a "standard" method for unicast traffic, whereas LCATM is used for multicast support and covering issues not yet addressed by MPLS specifications (e.g., QoS support for unicast). The LCATM-enhanced MPLS is able to communicate standard MPLS protocols (i.e., LDP) with any pure MPLS-capable device. At the same time it can use the additional functionality of LCATM for multicast and QoS support. Both protocols, MPLS/LDP and LCATM, run in parallel and independent of each other on the same controller. The common part to be used by both would be resource management, that is, the administration and management of VCs (MPLS labels).

The integration of standard MPLS unicast and the MPLS multicast modules (LCATM) requires a central, consistent management of the underlying ATM network resources, an ATM resource manager that can serve all modules that need ATM resources. To control the ATM switch hardware the resource manager uses the GSMP protocol currently being standardized in the IETF. To achieve this integration, the following enhancements were pursued:

- Incorporation of the LDP protocol mechanisms for label assignment and distribution, according to emerging IETF specification, in the LCATM architecture
- Intelligent allocation, administration, and management of VC identifiers for short-cutting (unicast and multicast) traffic in the LCATM platform

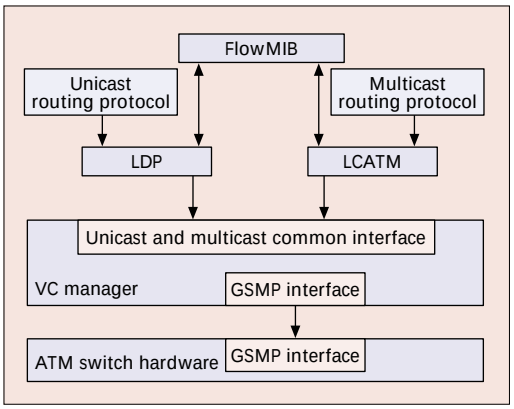

Figure 3. *The joint architecture with VC manager.*

- IP address resolution for the mapping of IP datagrams to ATM VCs (which VC has to be used for a given IP datagram)

The joint architecture is shown in Fig. 3. The unicast and multicast daemons are influenced by the respective routing protocols. The LDP functional block is responsible for sending/receiving LDP PDUs. The VC (label) Manager is a single point of control for managing the label table and the GSMP interface.

IPSOFACTO Flow MIB — Fine grain accounting of network traffic is common in plain old telephony service (POTS) or ATM networks. However, in IP networks this way of accounting is problematic, because there is no concept of connection in the network layer. The IETF Meter management information base (MIB) [9] is targeted at overcoming this lack in IP traffic accounting. Within the IthACI project the Meter MIB was enhanced by multicast capabilities and adapted to measuring multicast traffic. Special flow attributes for multicast data flows have been added.

A Flow MIB was developed for NEC's LCATM prototype. The Flow MIB is an extension of the Meter MIB proposed by the Real Time Flow Monitoring (RTFM) working group of the IETF [9]. The Meter MIB collects simple per-flow data at network elements. Flows may be aggregated to varying degrees of granularity, based on rule sets provided by the network manager. Among others, data on source and destination address (any layer), bytes forwarded, bytes dropped, creation time, last active time are collected.

The Flow MIB extends the standard Meter MIB in that it also records:

- Flow data such as bandwidth assigned, other QoS parameters, and IP switching label and routing information
- Statistical analysis of flow data such as flow rate frequency distribution, averages, deviations, consumed bandwidth as a function of time, and traffic matrices (indexed by sender/receiver, each element denoting average bandwidth on that connection)
- Prediction of future bandwidth needs
- Non-local information such as per-flow traffic volume between network elements (this is a concept from the RMON2 MIB)

The Flow MIB collects and provides information on past and present flows. This flow information can be used in many ways. Applications for a Flow MIB include:

- Statistics on the frequency of sender/receiver address pairs serves to pre-establish frequently used shortcuts.
- Information on current shortcuts (source/destination, shortcut ID) serves to aggregate flows into already existing shortcuts.
- Prediction of future bandwidth use, based on past flow data, serves to dynamically adapt bandwidth e.g., for RTP applications. This can also be extended to other QoS parameters.
- Data on current usage of the network serves to explicitly route and re-route flows (traffic engineering as defined by the MPLS working group).

QoS and Resource Management from the ISP's Perspective

Adopting the viewpoint of an ISP, the project has specified the so-called Explicit Services QoS provisioning framework for delivering end-to-end QoS in Tag-switching and MPLS networks.

Adopting the viewpoint of an ISP, the project has specified the so-called *Explicit Services QoS provisioning framework* for delivering end-to-end QoS in Tag Switching and MPLS in general, networks [10]. The main requirement underlying the design of the framework was to provide the appropriate flexibility and means to meet evolving requirements with respect to service provisioning since these are driven by market needs in a multiprovider competitive environment and specific QoS requirements of emerging services.

The proposed framework specifies that the QoS-based services offered by the network are predetermined, with each QoS parameter (e.g., bandwidth, loss) which can be managed by the network associated with an explicit arithmetic value (usually denoting a bound); hence the name Explicit Services QoS provisioning framework. In essence, the QoS space offered by the network is discretized into a finite set of explicitly defined QoS-based services, called and *explicit network CoS* (enCoS).

In this particular study, the performance of the enCoSs is characterized by the minimum guaranteed bandwidth; that is, enCoSs are considered to offer a connectivity service offering low-latency low-loss transfer at a minimum guaranteed rate. Drawing similarities with DiffServ networks, the considered enCoSs can be seen as EF-based classes that can be defined based on an EF per-hop behavior (PHB) per network node. However, in this study an MPLS-capable network was assumed with no DiffServ capabilities.

Under this model of QoS provisioning (i.e., by discretizing the QoS space), the network appropriately dimensions and manages its resources for offering users the enCoSs that best match their QoS requirements, rather than trying to dynamically meet users' arbitrary QoS requirements.

The traffic engineering approach operates on a virtual network topology imposed to the physical one, called *soft multilink network* (SMLN). It segregates the bandwidth per physical network interface into virtual links, called soft network links (SNLs). Explicit routes per enCoS from edge to edge are defined through the determined SNLs (Fig. 4). This approach enables

The traffic engineering logic builds on top of the QoS support capabilities that may be offered by the network elements in an MPLS network and is decomposed in a hierarchical functional model, offering the desired levels of adaptivity to changing workload conditions.

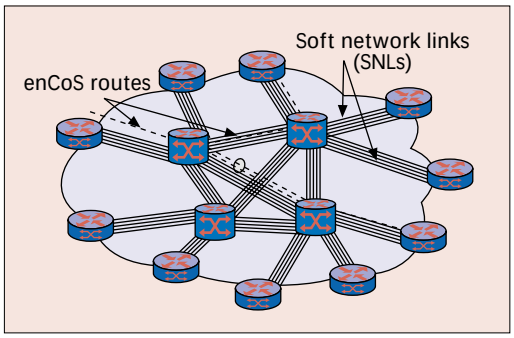

Figure 4. *The soft multilink network.*

increased and flexible control of access to and sharing of network resources between the supported enCoSs. Therefore, cost effectiveness in network operation and enforcement of business policies with respect to QoS provisioning can be achieved. If no segregation was in place, all enCoS streams would compete for the entire physical bandwidth on a first-come first-served basis, making it impossible to enforce business policies.

The traffic engineering logic builds on top of the QoS support capabilities that may be offered by the network elements in an MPLS network and is decomposed in a hierarchical functional model, offering the desired levels of adaptivity to changing workload conditions. Specifically, the functional model combines both centralized logic (for building the virtual network), the SNLs, and the explicit routes (network policy maker, NPM, component) and dynamic distributed (per network node) logic. The distributed logic manages the bandwidth allocated to the determined SNLs according to their usage level: the local policy making (LPM) component regulates incoming traffic should traffic conditions warrant; the traffic regulator (TR) component dynamically manages the routing of enCoS streams across the defined explicit routes (enCoS routing, enCoSR component). Actual usage statistics per link and/or SNL, required by the operation of the previous components, are captured by the *monitor metric* (*MM*) component through appropriate monitoring functions. This approach attains the merits of dynamic network operation schemes (as in the Internet today), and at the same time also has the benefits of cost-effective network planning and dimensioning. Figure 5 depicts the IthACI functional model in an example network configuration.

The specified approach presents one of the first attempts for policy-based network management. Specifically, a policy is decided by a centralized component (NPM) for sharing of network bandwidth among the competing traffic classes (enCoSs). The policy is substantiated in two bandwidth parameters associated with the determined SNLs: minimum available bandwidth (MAB), always to be provided to SNLs, and minimum bandwidth to be guaranteed (MBG) per SNL under congestion. The enforcement of the policy to the network is then delegated to appropriate network node components (LPMs), which based on current SNL usage statistics, manage SNL allocated bandwidth within the constraints of the determined optimal bandwidth sharing policy. The enforcement of the results of the policy in the network (determined nominal bandwidth per SNL) is not done every time a decision is made; instead, it is enforced to the network only when traffic conditions warrant (when the SNL used bandwidth is greater than its determined nominal bandwidth).

The specified functional approach relies heavily on the notion of explicit routing, supported by the emerging MPLS and traffic engineering specifications.

A point worth mentioning relates to the ongoing discussion in MPLS regarding whether reservations (and therefore connection admission control) must accompany the establishment of (explicitly routed) LSPs. As already outlined, the proposed framework does not rely on (and views unnecessary) LSPs with reserved bandwidth for delivering QoS. Instead, it relies on more liberal policy-based means for ensuring QoS delivery. Specifically, it relies on "soft" reservations of resources, which apply to aggregate rather than individual flows. The term soft denotes that the reservations are maintained at a virtual level (LPM component) and enforced to the network only when traffic conditions warrant (as previously

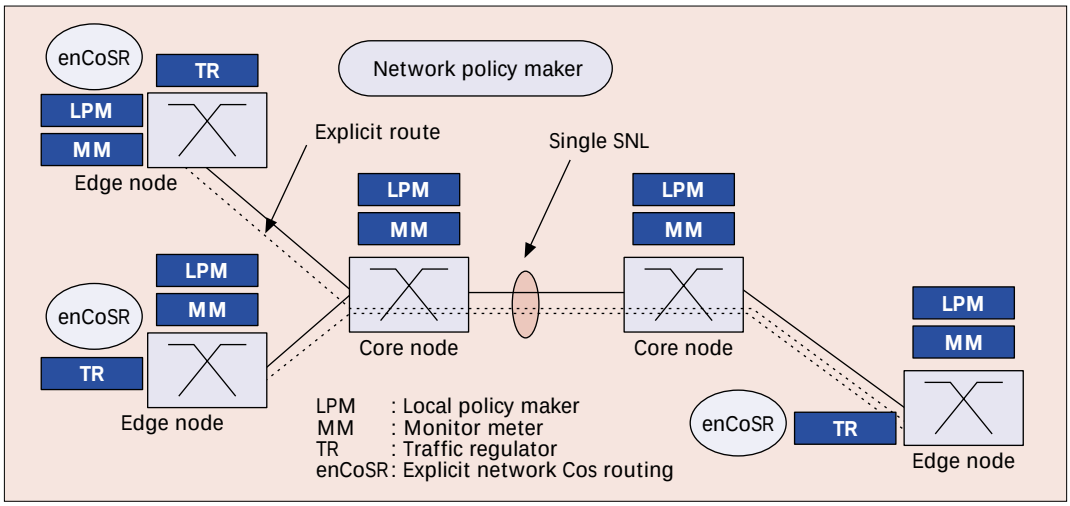

Figure 5. *Functional model for QoS provisioning in MPLS networks.*

discussed). In such cases, the enforcement of the reservations to the network is done by appropriately regulating incoming traffic through policing (TR component) mechanisms, and as such takes place only at the edges.

Finally, it should be noted that because of the aggregate traffic-based reservation scheme employed, QoS cannot be fully (100 percent) guaranteed throughout the lifetime of a flow, but only within statistical levels, depending on actual traffic conditions.

TESTS AND RESULTS

The enhancements described in the previous sections were implemented and tested in the project islands. The multicast enhancements were tested in the Red and Green islands, and their interoperability was assessed by interconnecting these two islands. The tests used the multicast routing protocols DVMRP and PIM-SM. With both routing protocols, LCATM and YALSA established layer 2 shortcuts for multicast flows, as expected. Hosts in the edge networks joined and left the multicast groups independently, and the data transmission worked fine. The QoS and resource management enhancements from the vendor's perspectives were tested in the Green Island, incorporating NEC's LCATM technology. The QoS and resource management enhancements from the ISP perspectives were tested in the Blue Island, based on a four-node testbed comprised of Cisco 7xxx series routers. In all islands, commercial and/or widely used videoconferencing applications were used to generate the required input traffic (e.g., RealServer G2, Mbone Vic).

It should be noted that the main goals of the testing work were:

- Proof of concept, for proving the validity of the design underlying the specified enhancements
- Demonstration of the benefits of the undertaken enhancements in typical operational scenarios

Based on the gained design and experimentation experience, a number of valuable conclusions were drawn on the applicability of IP switching technologies and their enhancements to support multicast and QoS, which are described in the following. More details on the undertaken experiments can be found in [11].

MULTICAST

♦The developed enhancements allow shortcuts to be established dynamically. Layer 2 branches (along the route given by the layer 3 multicast routing protocol) are established and deleted as multicast senders and receivers join or drop out.

♦The developed approach is independent of the multicast routing protocol. In dense mode environments, DVMRP can be used without any problems. In larger sparse scenarios, PIM-SM is the preferred solution due to its current better acceptance and deployment in backbones. The multicast solution was proven usable for both multicast routing protocols, and did not highlight any additional scalability issues.

♦The scalability supported in terms of number of groups and sources is roughly the same as for the multicast routing protocol, basically better for PIM-SM than for DVMRP. When using a UNIX-like system (e.g., Linux or FreeBSD), the scalability of PIM-SM is in general less than what the protocol specification promises. This is due to the use of a per-source (S, G) MFC, that is, an MFC entry always consists of a source, group pair, even when the multicast routing table aggregates different sources in the same (*, G) entry. Note also that our LCATM design uses this (S, G) MFC structure to provide per source LSP. So the maximum number of groups supported depends on the number of sources per group, and is ultimately limited by the kernel memory availability.

♦Our approach allows to have one label per (*, G) entry, which scales better than other methods which create an LSP per source even if the state is (*, G).

QOS PROVISIONING

♦RTP flow detection in IP-switching (and MPLS in general) networks is possible and feasible. Although no explicit signaling is required and only the first packets of the flows are routed, additional load from the detection algorithm is put on all nodes. Alternatively, flow detection could be done only at ingress points. But then other means for path setup must be used (explicit signaling, MPLS, DiffServ). Therefore, locating IP switches with application flow detection at the border between local and wide area networks appears to be an attractive approach.

♦The detection of RTP flows allows providing these flows with a better QoS than just best effort. The type of content/data carried in the data packets, or some policy rules or traffic engineering aspects may influence the choice of assigned QoS/CoS. The granularity for making QoS/CoS assignments to RTP flows can be coarse, just checking whether a packet is RTP or not, or more fine-grained, preferring only RTP flows with certain payload types. Although RTP flow detection is based on a weak test, we have found that RTP traffic is properly detected with a very high probability. RTP detection on the LCATM MPLS prototype works with a detection ratio above 90 percent.

♦In-band QoS provisioning, based on QoS detection methods such as RTP detection, is relatively simple to achieve compared to reservation-based architectures like the Internet integrated services (IntServ) model, achieving similar merits with QoS differentiation ala DiffServ architecture. Benefits of this approach are:

- Application QoS detection makes use of information which is already contained within the data stream.
- There is no need for special QoS signaling between the network elements; hence, no additional protocol overhead is introduced. No complex protocol is needed to communicate QoS between network nodes.
- Differentiation in this case is provided by assigning a different priority than the otherwise used default priority in the ATM switching hardware. The complexity of resource management (buffer, bandwidth, and queuing strategy) schemes may be minimized.

It should be noted that because of the aggregate traffic-based reservation scheme employed, QoS cannot be fully guaranteed throughout the lifetime of a flow, but only within statistical levels, depending on actual traffic conditions.

MPLS offers versatile grounds for providing end-to-end QoS. Its features to support explicitly routed paths and to utilize for routing more rich information than just the destination address, are particularly useful.

♦The level of QoS/CoS to be assigned to the different flows is influenced by the rules of the traffic engineering solution in the network. To gain some insight into the issue of RTP flow differentiation through different CoS level assignment (at the ATM switch), simulations were conducted. The main conclusions derived were:

- Control traffic, including RTCP traffic, should use the highest priority. It was shown that given its conservative traffic rate and small queue size, the control traffic does not significantly affect all other types of traffic, which have been assigned a lower priority.
- Data traffic, such as best-effort TCP/IP traffic, should use the lowest priority. The fact that TCP is responsive to network congestion provides the flexibility to allocate TCP-based traffic to the lowest priority queue.
- RTP flow differentiation/prioritization was indeed shown to be useful. Instead of multiplexing all RTP types in one queue, allocating them to different priority queues according to some rules had clear advantages. A basic rule of thumb that can be used to determine the priority of an RTP type is that the more delay-sensitive the RTP type, the higher the assigned priority should be, while high-throughput RTPs should be assigned lower priorities.
- Within a particular class, two loss priorities can be used. For example, in the case of MPEG2 video, control and motion information is more important than texture information.
- It is critical that appropriate admission control and policing mechanisms be used to restrict the service rates from all but the lowest priority level.

♦MPLS offers versatile grounds for providing end-to-end QoS. Its features to support explicitly routed paths and to utilize for routing more rich information than just the destination address are particularly useful. Our work proved that it is possible to deliver end-to-end QoS (service differentiation in terms of different levels of guaranteed throughput levels) in existing MPLS-capable IP networks. This was achieved by integrating value-added, intelligent network management functions (e.g., those specified in our design) on top of commercially available network equipment. The performed demonstrations showed that the network was able to differentiate its clients on the basis of the services to which they had subscribed as well as the cost effectiveness of the network operation. The proposed design takes into account business policies with respect to service provisioning; these are reflected through appropriate reservations of network resources. To increase network cost effectiveness, these reservations are dynamically managed on an actual demand basis. It was shown in the demonstrations that as long as QoS guaranteed services were not requested, some of their reserved bandwidth was awarded to best-effort traffic. As long as QoS guaranteed services were requested, bandwidth awarded to best effort services was reclaimed, ensuring that the network resources were shared between the supported services optimally according to the service provisioning policies of the ISP.

♦Of particular value for QoS provisioning is the support of MPLS for explicit routing. This has been advocated (and is justified by our experience) by the following reasons:

- Existing layer 3 routing protocols currently supported in network equipment are not QoS-aware.
- In a multiclass environment, the traffic commodities (ingress-egress-CoS) that contribute to the congestion of a specific network link (or area) can be known.

♦Considering a multiclass network environment, traffic engineering based on the virtual network topologies seems a viable solution. That way, and by means of explicit routing, network bandwidth is segregated among competing traffic classes, thereby contributing to network cost effectiveness and enabling the enforcement of business policies with respect to service provisioning.

♦Based on our experience, we see it as unnecessary to see MPLS LDP extended with reservation capabilities. Our work showed that reservations can be enforced through more lightweight means (e.g., traffic regulation at the edges as in our approach or through PHBs as in DiffServ). This is at the expense of hard guaranteed QoS, which we believe can be tolerated considering the benefits gained (scalability, reduced signaling protocol overhead and processing).

♦We see it as necessary to have MPLS combined with DiffServ. This would provide a powerful set of capabilities for enabling end-to-end QoS provisioning.

Resource Management

♦Unicast and multicast traffic can coexist in a MPLS environment, provided that efficient resource management schemes are in place to cater for the use of shared resources (e.g., connection identifiers).

♦Our experimentation has verified that MPLS increases transmission efficiency, reducing processing load.

♦MIBs for retrieving traffic-related information in an MPLS environment and configuration of related mechanisms (e.g., explicit routes) are required for the definition of traffic engineering solutions.

♦The project has developed a Flow MIB extending the IETF Meter MIB by new flow attributes specific to MPLS, multicast, and the traffic class. It can be customized to deliver fine-grained or coarse-grained traffic information by specifying traffic flows to be monitored.

Summary

The IthACI project addressed the issue of integration of IP and ATM through IP switching technologies and proposed several enhancements on such enabling technologies, which were validated through experimentation. Enhancements were pursued in the areas of multicast, QoS, resource management, and mobility in a multicast environment. Although the project experimented with proprietary solutions available at the time, the solutions proposed apply to MPLS in general.

The enhancements were tested through implementation and experimentation in individual and/or interconnected islands. Experimentation proved the validity of the concepts underlying the design of the enhancements and demonstrated the benefits of applying the developed enhancements in realistic network environments.

The enhanced features developed within the IthACI project advance the state of the art in MPLS networks. The work in MPLS multicasting described earlier was fed back to the MPLS working group. The work in QoS provisioning showed that MPLS offers versatile grounds for building effective traffic engineering solution for QoS delivery, especially through support for constraint-based routing.

Although the project experimented with proprietary solutions that were available at the time, the solutions that were proposed apply to MPLS in general.

ACKNOWLEDGMENTS

This article describes work undertaken in the context of the research project AC337 IthACI as part of the ACTS program, which was partially funded by the Commission of the European Union. The authors wish to thank all their project colleagues who contributed in many ways to the formation of the ideas described here.

REFERENCES

[1] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching Architecture," RFC 3031, Jan. 2001.
[2] http://www.algo.com.gr/acts/ithaci/
[3] A. Acharya, R. Dighe, and F. Ansari, "A Framework for IP Switching over Fast ATM Cell Transport (IPSOFACTO)," *Proc. SPIE Voice, Video and Data Commun.,* Nov. 1997.
[4] A. Acharya *et al.*, "IP Multicast Support in MPLS Networks," draft-acharya-ipsofacto-mpls-mcast-00.txt, Internet Draft, Aug. 1999.
[5] D. Ooms *et al.*, "Framework for IP Multicast in MPLS," draft-ietf-mpls-mcast-00.txt, Internet Draft, Jan. 2001.
[6] F. Griffoul *et al.*, "Layer 4 QoS Detection in Flow-based IP Switching," IDC '99, Sept. 1999.
[7] H. Schulzrinne *et al.*, "RTP: A Transport Protocol for Real-Time Applications," RFC 1889, Jan. 1996.
[8] H. Schulzrinne, "RTP Profile for Audio and Video Conferences with Minimal Control," RFC 1890, Jan. 1996.
[9] N. Brownlee, "Traffic Flow Measurement: Meter MIB," RFC2720, Oct. 1999.
[10] P. Georgatsos, "QoS Provisioning in MPLS Networks," MPLS '99, June 1999.
[11] AC337 IthACI Project, "Evaluation of IP Switching Enhanced Features and Trial Results," Deliverable D7, Jan. 2000.

ADDITIONAL READING

[1] H. Stuttgen, "ACTS IthACI - MPLS Extensions for Multicast and Quality of Service," *IEEE Globecom '99*, Dec. 1999.

BIOGRAPHIES

ILIAS ANDRIKOPOULOS (ilias@ieee.org) holds a Diploma in physics from the University of Athens, Greece, an M.Sc. in information technology from University College London (UCL), United Kingdom, and a Ph.D. in communication networking from the University of Surrey, United Kingdom. While studying for his Ph.D., he served as a research fellow in the Networks Research Group at the Center for Communication Systems Research (CCSR) of the University of Surrey working in EU and UK funded research projects. His main research interests include IP QoS, traffic management, Internet technologies, mobile and satellite networking, and network management.

GEORGE PAVLOU (G.Pavlou@eim.surrey.ac.uk) received a Diploma in electrical engineering from the National Technical University of Athens, Greece. He obtained M.Sc. and Ph.D. degrees in computer science from University College London, where he worked as a senior research fellow and lecturer. He is currently professor of information networking at the Center for Communication Systems Research at the University of Surrey, where he leads the activities of the Networks Research Group. His research interests include network planning an dimensioning, traffic engineering and management, programmable and active networking, multimedia service control, and distributed object-oriented platforms.

PANOS GEORGATSOS (pgeorgat@algo.com.gr) received a B.Sc. degree in mathematics from the National University of Athens, Greece, in 1985, and a Ph.D. degree in computer science from Bradford University, United Kingdom, in 1989. He is currently working at Algonet S.A., Athens, Greece, where he is responsible for the R&D Group in telecommunications. His research interests include service quality management, network routing, planning, resource dimensioning, analytical modeling, simulation, and architectures for distributed systems.

NICHOLAS KARATZAS (nikos@algo.com.gr) received his B.Sc. in mathematics from the University of Athens in 1980. His graduate studies were conducted at the University of Southwestern Louisiana, from where he received his M.Sc. in artificial intelligence in 1983 and concluded doctoral studies in software engineering in 1985. His research interests include, among others, service creation and deployment, accounting and billing systems, intelligent networks, and communications management. He is currently technical director of the Total Solutions and Advanced Services department of Algosystems S.A, Greece.

KOSTAS KAVIDOPOULOS (kavid@algo.com.gr) received his B.Sc. degree in electrical and computer engineering from the National Technical University of Athens (NTUA), Greece, in 1992, and his Ph.D. degree in network engineering also from NTUA in 1998. He is currently working at Algonet S.A., Athens, Greece, responsible for the Value Added Services Group. His research interests include resource and traffic management, value-added services on IP networks, network and service management, distributed systems, and Internet technologies.

JÜRGEN RÖTHIG (jroethig@gmx.de) studied computer science at the University of Karlsruhe, Germany, where he received his diploma in 1991 and his Ph.D. in 1994 with a thesis on resource management and load control in ATM networks. He then worked as a project manager with the German Aerospace Center in Cologne on behalf of the German Ministry of Education and Research, and next as research staff member and project manager at C&C Research Labs in Heidelberg, where he was responsible for the NEC part in the IthACI project. After NEC, he was for half a year with Lufthansa Systems, Kelsterbach, Germany, as a project manager responsible for the optimization of the worldwide data network of Lufthansa.

SIBYLLE SCHALLER (Sibylle.Schaller@ccrle.nec.de) received her diploma in computer science from the Technical University of Dresden in 1993. She worked for IBM at the European Networking Center in Heidelberg, Germany, and the IBM research laboratory in Haifa, Israel. In 1998, she joined the NEC C&C Research Laboratories in Heidelberg. She was a member of the NEC team in the ACTS IthACI project.

DIRK OOMS (Dirk.Ooms@alcatel.be) graduated from the University of Leuven in 1989 in electrical engineering and in 1992 from the University of Gent in physics. He did software development for ATM switches from Alcatel, Siemens and Newbridge. Since 1997 he has worked in the Network Architecture Group of Alcatel. His main interests are currently IP multicast and IPv6. He is an active participant in the IETF and the author of multiple publications in international conferences and journals.

PIM VAN HEUVEN (pim.vanheuven@intec.rug.ac.be) graduated in computer science from Ghent University in 1998. At the same university he joined the Integrated Broadband Communications Networks Group (IBCN) in 1998; the latter is headed by Prof. Piet Demeester (demeester@intec.rug.ac.be). He is working toward a Ph.D. degree and worked previously in the ACTS IthACI project. His research interests include MPLS, network resilience, QoS and traffic engineering, and Linux-based routers.