ELSEVIER

Future Generation Computer Systems 19 (2003) 313–326

FGCS
FUTURE GENERATION COMPUTER SYSTEMS

www.elsevier.com/locate/future

QoS and SLA aspects across multiple management domains: the SEQUIN approach

Christos Bouras [a,b,*], Mauro Campanella [c], Michal Przybylski [d], Afrodite Sevasti [a,b,e]

[a] *Department of Computer Engineering and Informatics, University of Patras, 26500 Rion, Patras, Greece*
[b] *Computer Technology Institute, 61 Riga Feraiou Str., 262-21 Patras, Greece*
[c] *INFN-GARR, Via Celoria 16, 20133 Milan, Italy*
[d] *Poznan Supercomputing and Networking Center, ul. Noskowskiego 10, 61-704, Poznan, Poland*
[e] *Greek Research and Technology Network, 56 Mesogion Av., 11574 Athens, Greece*

Abstract

The aim of this work is to define and implement an end-to-end approach to quality of service (QoS), operating across multiple management domains and exploiting a combination of link layer technologies. The architecture for the Premium IP service is presented, which aims at offering the equivalent of an end-to-end virtual leased line service at the IP layer across multiple domains. Also, the results of the initial testing performed for the validation of the service and the provisioning model for Premium IP are described. The work presented has been carried out in the framework of SEQUIN, a European Commission-funded research project.

Keywords: Quality of service; Premium IP service; Differentiated services; Service level agreement; Monitoring

1. Introduction

The need to support differentiated quality of service (QoS) has been recognized since the creation of the previous generation backbone networks and has been delivered, until recently, using ATM technology. However, with the advent of gigabit networks, ATM alone is no longer available to deliver end-to-end service differentiation and QoS. The availability of high-speed transmission media and networking equipment, as well as the evolution of quality-demanding applications operating over the Internet protocol (IP) have raised interest in the provision of advanced quality services at the network layer, in addition to the traditional best-effort service of the Internet.

In the research field, a number of alternatives for service differentiation and QoS provision have been proposed and standardized, but in the case of backbone networks, the Differentiated Services [2] architecture has prevailed, due to its scalability and deployment feasibility. The DiffServ framework stands out by attempting to provide service differentiation to traffic in a scalable manner, featuring the aggregation of individual application flows with similar quality needs. It introduces the definition of different service classes to

* Corresponding author. Present address: CTI, PO Box 1122, Patras 26110, Greece. Tel.: +61-960375; fax: +61-996314.
E-mail addresses: bouras@cti.gr (C. Bouras), mauro.campanella@garr.it (M. Campanella), michalp@man.poznan.pl (M. Przybylski), sevastia@cti.gr (A. Sevasti).

0167-739X/02/$ – see front matter
PII: S0167-739X(02)00156-5

which such aggregates are appointed and the implementation of mechanisms for differential treatment by network elements (Per Hop Behaviour–PHB) of the packets belonging to each service class. A PHB describes the treatment of aggregated traffic in a manner that ensures the quality guarantees required by the corresponding service class.

SEQUIN [16], an acronym standing for 'Service Quality across Independently Managed Networks', is the name of a project involving eight partners in seven countries and co-funded by the European Commission under the Information Society Technologies (IST) Programme. SEQUIN has adopted the principles of the DiffServ framework in order to define an end-to-end approach to QoS over consecutive interconnected domains. As a result, a complete and scalable service model for a service named 'Premium IP' has emerged, offering quality such as that of a virtual leased line at the IP layer. The proposed architecture is targeted at the GÉANT network (the pan-European Gigabit Research Network) and is applicable to each connected National Research and Education Network (NREN) across Europe and any local DiffServ domains.

After the definition of the service model and architecture, SEQUIN recognized the need to have an initial implementation of the architecture by implementing a 'proof of concept' test-bed as a precursor to a full set of field trials involving user groups. The goal of this test-bed has been to have access to a controlled environment composed of a variety of hardware and underlying technology to verify the functionality of each component required to implement Premium IP.

This paper provides an overview of the general framework and implementation architecture for provisioning DiffServ-based QoS in the form of the end-to-end, IP-based, qualitative Premium IP service. It also describes the efforts to check the feasibility of the proposed Premium IP implementation in production networks, with the strong emphasis on end-to-end QoS delivery. Finally, it presents the basic principles for the deployment of the service level agreements (SLAs) on which Premium IP provisioning is based both in a bilateral fashion (between peering domains) and in an end-to-end fashion (between end-users), together with guidelines for the deployment of a monitoring infrastructure that will verify performance goals and the SLAs.

2. QoS definition approach

The service architecture proposed with the introduction of Premium IP service benefits from the convergence of a bottom-up approach that defines QoS parameters and a top-down approach, which starts from users' requirements. The two approaches converge to a common set of QoS parameters, which will be used on the definition of the proposed service. The architecture aims at delivering a production service in a short timescale, and to accomplish this, it takes a pragmatic approach to balancing configuration complexity, available technology, generality, benefits and implementation timescale.

The proposed service is limited in scope to network QoS. Nonetheless, the quality of an end-to-end application is the result of a combination of network, operating system and application behavior. All these components must be capable of providing QoS guarantees to obtain the needed quality. A more detailed description and in depth analysis can be found in SEQUIN deliverables and in particular, in [4–6,17].

IETF and ITU-T have already defined a list of QoS parameters that can be chosen to quantify QoS services. The two organizations agree on the list of parameters that can be used to gauge the performance of an IP link, although with small differences. The more significant difference is that the ITU-T proposal adopts a statistical definition of QoS parameters, while IETF allows for more than one measurement procedure for each parameter.

We propose the following list of parameters as the basic set used to quantify any QoS service:

- One-way delay (OWD),
- Instantaneous packet delay variation (IPDV),
- One-way packet loss (OWPL),
- Capacity.

The naming and meaning of these QoS parameters will follow IETF IP Performance Metrics group guidelines and framework [15].

Some basic characteristics of network behavior (such as physical, data link layer and routing stability, negligible packet reordering or duplication, overall network hardware performance, etc.) greatly influence the overall quality of any service. Such conditions are supposed to be met in a well-behaved network and

Table 1
QoS requirements grouped in service classes

QoS service	OWD	IPDV	OWPL	Capacity
BE	Wide	Wide	Medium	Wide
Premium IP	Medium	Short	Short	According to SLA
Prioritised bandwidth	Medium	Medium	Medium	According to SLA
Guaranteed bandwidth	Medium	Medium	Short	Single value

in following discussion of IP-based QoS we take this assumption for granted.

As part of the top-down approach, a set of research groups of users in Europe was interviewed through a questionnaire in order to assess their perception and requirements for QoS in the network. The groups were chosen in such a way as to provide a non-homogeneous sample, and range from large Universities to projects on network research. The only requirement was usage of the network. A full description of results can be found in [5].

Overall, the users showed medium knowledge of their QoS needs and QoS techniques, but unanimously requested it, as a way to have a better service from network for their work. It is worth noticing that a major cause of present difficulties is attributed to congestion, and that willingness to pay is proportional to the real benefits, granularity of the service, provisioning time and flexibility as well as behavior of Best Effort (BE) traffic.

Table 1 groups the users' QoS requirements in classes and identifies possible services. These classes of service are characterized by the width of the value range for each QoS parameter.

The results of the interviews show that we can reduce the number of QoS services to three, by merging Prioritised Bandwidth and Guaranteed Bandwidth services into a new IP+ service. The interviews also provided indicative value ranges for the QoS parameters, which are listed in Table 2.

Table 2 shows that the implementation of a Premium IP service like the one described in this paper can satisfy, in practice, all of users' QoS requirements.

3. Premium IP specification and architecture

In order to satisfy users' requirements, the Premium IP service must be able to provide a bounded OWD, minimal IPDV and null or insignificant OWPL. The service is thus similar to the 'virtual leased line' original proposal of Nichols et al. in [13] from which the name is inherited. A summary of the specifications (see also [6]) follows.

In addition to the aforementioned characteristics, the Premium IP service has to take into account additional requirements.

- It must be applicable to a network composed of multiple connected domains.
- The only protocol to be used in the operational network immediately is IPv4. Nonetheless, it would be an advantage to define a service that can be applied to IPv6 flows with minimal or no modifications.
- The service should be modular, highly scalable and adapt easily to network modifications, such as link upgrades and additions.
- The service should be based on state-of-the-art hardware and software and should be designed so that it can be adopted on a production infrastructure.

Table 2
Indicative value ranges for the QoS parameters

QoS service	OWD	IPDV	OWPL	Capacity
BE	Unspecified	Unspecified	<5%	Unspecified
Premium IP	Distance delay + 50 ms	<25 ms	Negligible	According to SLA
IP+	Distance delay + 100 ms	<25–50 ms	<2%	According to SLA

It must not require a separate physical infrastructure.
- The service model and implementation should remain compliant to the IETF standard track, and may be revised in the future if required by the ongoing work at IETF.
- The service must be independent from the data link layer transport technology.
- The activation of the Premium IP service should not generate any side effect on the whole of the BE traffic, except for a reduction in capacity available to BE when Premium IP packets are present. In particular it should not starve the BE traffic. Starvation of the BE traffic should be considered as an indication of the need of a modification of the IP SLAs in effect, or a capacity upgrade.
- The activation of the Premium IP service must not forbid the activation of other QoS services, when needed, according to hardware and software support.

According to the above requirements, the DiffServ architecture and more particularly the expedited forwarding PHB (EF PHB) [9] are chosen as the framework where the Premium IP service will be defined.

The definition for the Premium IP service so far is valid for a single DiffServ domain and specifies the behavior of the domain at each hop traversed by eligible traffic. The implementation of the Premium IP service on an end-to end scale implies traffic, in general, that crosses multiple domains. We will adopt the following requirements to build an end-to-end Premium IP service:

- All the QoS domains involved must implement the DiffServ architecture and map the Premium IP traffic to the EF PHB.
- An interface specification is agreed between the various domains to correctly map Premium IP traffic between them. The interface specification may contain mapping between DSCP values, policing rules, capacity assurances and all the parameters needed to ensure a correct propagation of the service.
- The interface should be defined in such a way that when packets cross management boundaries, the packet treatment is compliant to the EF PHB.

In this way, peering domains are free to have different physical implementation of the Premium IP service. The interface specification has to assure that Premium IP service traffic flows to and from two peering domains in a seamless way, without any packet loss and with minimum delay and delay variation. The effectiveness and sufficiency of this approach is subject to experimental validation.

3.1. Premium IP implementation

In engineering the implementation of the Premium IP service, additional choices have been made, always with the objective of simplifying the general structure. Main issues and principles are considered in the following sections.

3.1.1. Flow shaping

Flow shaping is the foundation of the correct behavior of the whole service. It ensures that the flows do not face packet losses due to traffic conditioning and minimizes creation of burstiness due to aggregation between different flows. Last but not least, shaping each flow ensures a fair sharing of the services between elastic and inelastic transport protocols like TCP and UDP. For these reasons, the architecture mandates that the sending host or source shapes the flows sent according to its allowed sending rate. This is required in order to avoid initial packet loss due to traffic conditioning when entering the DiffServ domain, and to ensure a fair share of the aggregated Premium IP capacity amongst all its simultaneous flows, since the network is not responsible for such fair sharing. The service will not shape per flow anywhere along the path to the final destination, being based on an aggregation model, and the network will not apply any additional aggregated shaping, both in ingress and egress points.

3.1.2. Admission control and classification

With respect to the service model, two different types of user requirements have emerged:

- A ‘virtual leased line service’ identical in functionality to a point-to-point link.
- A service in which selected packets should get preferential treatment, independent of the destination, up to a contractual ingress capacity.

The first service allows a precise dimensioning of the resource requirements in the network. At each node

it is possible to estimate the maximum capacity to be transported, whilst knowledge of the forwarding path allows measurement of the OWD and of the IPDV under normal operation. This type of service, denoted as 'destination aware', implies that the admission control is based on the mandatory pair of (*IP source, IP destination*) prefixes of IP packets. The second request implies that the traffic can be shaped and policed at the ingress according to an SLA, but then the traffic path cannot be predicted. The latter type of service is denoted as 'destination un-aware'. In a destination un-aware situation, it is easy to find cases in which the SLAs are violated due to the lack of information in the service set-up. For this reason, the Premium IP service is mandated to be destination aware, requiring an admission control rule that will be analyzed in the sequel and based on the destination prefix of packets, in addition to their origin and traffic conditioning rules.

More specifically, admission to the Premium IP service will be based, at the border nearest to the source, on (*IP source, IP destination*) prefixes and appropriate traffic conditioning rules. Packets with Premium IP-eligible prefix pairs but exceeding the agreed traffic conditioning will be discarded. Packets admitted to the Premium IP service will be marked with a DSCP or IP Precedence value that is strongly recommended to be equal in all involved domains. Between peering domains, packets will be served according to the QoS tag (DSCP or IP Precedence), 'trusting' the ingress domain. The admission control can be also based on other parameters, as defined case by case. In a particular case, the source is capable of DSCP-tagging of packets and admission is then granted only when the tag is present. However, this is discouraged due to security concerns.

Admission control and classification must be enabled on all border routers in the form of a general 'deny-unless-explicitly allowed' rule. The general 'deny' rule must be active before the service is started. It is also suggested that each domain builds a matrix to compute and account the IP Premium rate subscribed between each pair of its border links, but more details on this are provided in Section 3.1.4.

3.1.3. Maximum premium IP traffic capacity

There is a limit to the amount of capacity to devote to Premium IP, due to:

- The type of service, which does not allow loss after initial traffic conditioning.
- The choice of never starving the BE traffic.

Moreover, it has been shown in [7] that in a network with aggregate FIFO scheduling, for sufficiently low enough utilization factors, deterministic delay bounds can be obtained as a function of the bound on utilization of every link and the maximum hop count of any flow.

It is thus suggested that the amount of Premium IP capacity subscribed does not exceed 5% of the core link speed. The computation should take into account the link speed between domains, and total Premium IP rate may vary between each link. The premium capacity can be larger nearest to the user. The choice minimizes the probability of instantaneous burstiness at aggregation nodes, which leads to packet loss. This minimal percentage also ensures that in case of re-routing, the service will continue to work without packet loss, albeit the OWD and IPDV will be different from base values.

3.1.4. Traffic conditioning

Policing as a means of traffic conditioning is a fundamental component of the proposed Premium IP architecture. Policing will be performed by means of a token bucket. In brief, the proposed service will not police or shape per flow, being based on an aggregation model. However, it may police aggregates according to their destination domain as a safety measure.

Micro-flow traffic conditioning with the form of policing should only be done as close as possible to the source of the flow, in the first DiffServ domain, using a contracted (via an SLA) Premium IP rate. Packets exceeding the allowed sending rate will be discarded. When crossing successive DiffServ domains, the policing functions are only based on aggregated Premium IP capacity and rules on capacity can be slightly relaxed, according to the relevant agreements. Policing is also enabled at ingress of core domain borders based on packets' DSCP tagging and the aggregated Premium IP capacity per each pair of source and destination peering domains. This implies a simpler set of rules and higher scalability. The core domain, in this case, does not need to know the addresses of participating Premium IP end nodes, but needs to maintain a global matrix of agreed aggregated

Premium IP rates between each pair of peering domains. The matrix does not need to be symmetric.

As far as the value of the policing token bucket rate is concerned, it is suggested that the rule in the core enforces a rate limit between each pair of peering domains that is 20% greater than the sum of all the contracted values between each pair, as computed from the Premium IP traffic matrix. If no traffic rate is agreed between a particular peering domains' pair, but packets marked with the Premium IP DSCP are encountered, the packets should be remarked to BE, a solution preferred to dropping. In the case of two domains connected by more than one link, the rules have to be applied at every ingress interface. It is suggested that the rule sets are identical, so that, in the event of routing failure, the Premium IP traffic is not affected. If the rules are based only on the Premium IP DSCP value, the sum of the allowed premium rates on the two links will be twice the agreed value if the rules set are identical. Appropriate values for the rates have to be investigated case by case, according to routing patterns.

In the case of a network device that has multiple interfaces, each one carrying Premium IP flows, the possibility of a collision of Premium IP packets coming from different interfaces on the same egress bucket exists, even if the links are unloaded. Moreover, if the interfaces do not have the same speed, for example, when a packet flows from a higher-speed link to a lower speed, a small burstiness in the high-speed part might cause packet discard in the lower-speed link. Besides, the proposed implementation requires policing only on ingress flows in selected nodes, usually only at the border, and never at egress. Hence, packet loss due to egress policing is automatically avoided, although the decision to avoid both shaping and egress policing might increase burstiness.

For these reasons, a depth of one full MTU for the token bucket policers in the core of Premium IP enabled architecture is considered insufficient. Experimental tests in [20] support this conclusion. It is suggested that the depth of the policing token bucket per Premium IP aggregate be progressively increased when moving further away from the source, at the price of a small increase of delay variation. Initial values can be set at two MTUs near the source and five MTUs at the core domains; larger values

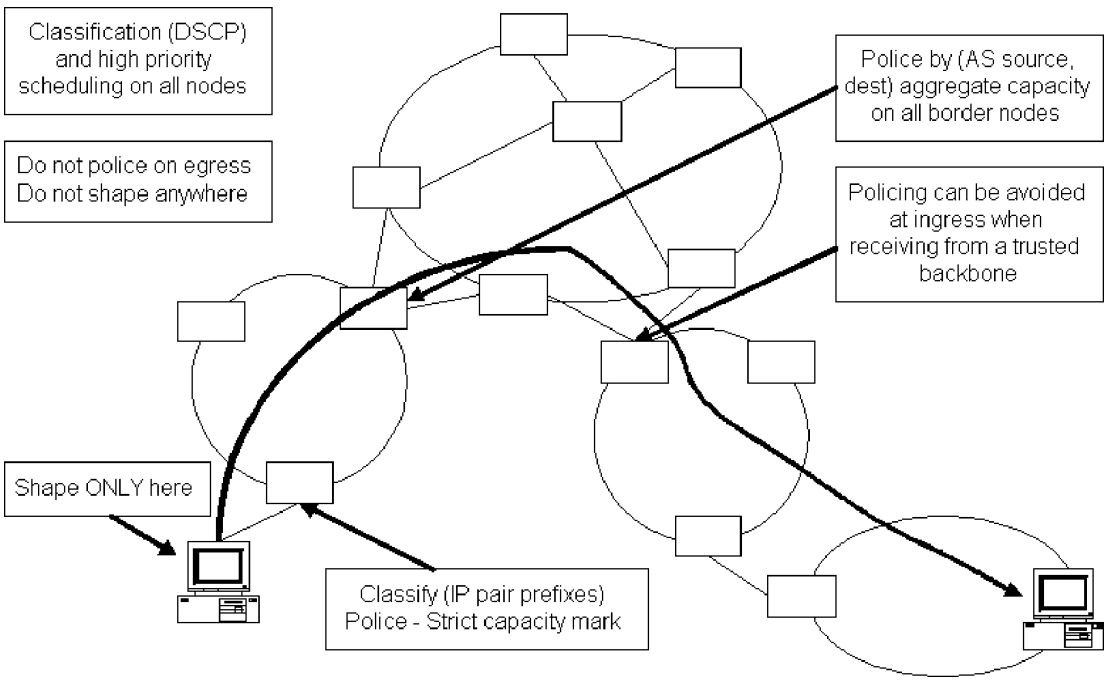

Fig. 1. Premium IP functionality on the end-to-end path.

can be configured when the number of hops in the path becomes large. It has to be stressed that the total depth of the policing bucket is used only when needed and, provided a correct, limited configuration on the amount of Premium IP capacity; it should be completely used only in very rare cases.

3.1.5. Summarizing overview

After this thorough description of the Premium IP architecture, Fig. 1 summarizes the required actions in each node of the end-to-end path.

4. Premium IP testing over production networks

The implementation model of Premium IP has been validated using production networks because such an approach has been considered more likely to provide fundamental feedback on the feasibility of the proposed architecture. Five of the NRENs (GARR from Italy, GRNET from Greece, GWiN from Germany, RENATER from France and SWITCH from Switzerland) participating in the SEQUIN project, together with the GÉANT network, have conducted test cases involving external users. It was important to select a group of users with good understanding of QoS requirements and implementation. Therefore, users from the TERENA TF-STREAM community [19] were selected because this task force performs research on and tests of real usage and scalability of audio/video streaming conferencing tools and techniques. This section describes the efforts to check the feasibility of the proposed Premium IP implementation over a production environment infrastructure for applications of H.323 videoconferencing.

4.1. Multi-domain test-bed topology

The test-case involving end-users has been designed in a way to reflect the complexity of a multi-domain heterogeneous pan-European network. It is composed of five high- and lower-speed national networks connected via the GÉANT backbone, connecting six testing locations. This composition helps to investigate the issues concerning the interaction between different types of networks, to validate the QoS techniques for use with different technologies and to check the behavior of those QoS techniques on different platforms.

The core network (GÉANT) is built with 10 and 2.5 Gbit/s POS technology and Juniper routers. Access networks connect to the GÉANT with 2.5 Gbit/s POS links, except for GRNET, which connects with 2×155 Mbit/s ATM links. The detailed test-bed topology has been depicted in Fig. 2.

A wide range of H.323 videoconferencing equipment has been used, allowing users to check the behavior of Premium IP service under different end-system conditions. Additionally, a dedicated computer for active measurements has been installed in each location, while end-point equipment was always connected via dedicated LAN to the nearest router.

4.2. Measurement techniques

For the metric measurements of the quantitative parameters of Premium IP (OWD, IPDV, OWPL and capacity), a sample traffic pattern has been prepared, based on the H.323 traffic recording from a videoconferencing unit. On the other hand, the end-user satisfaction was also be expressed by subjective measurements or evaluation, i.e., for the videoconference transmission, the users assessed the picture and sound quality. According to these assumptions the following measurement techniques have been adopted [11,14].

For the subjective measurements, the users assessed the quality of the long distance videoconferencing by comparing it with local videoconference results. The assessment has been given in the form of the simple number ranging from 1 to 6.

- IPDV [8] has been measured with the use of active measurement techniques. For the metric measurements of IPDV, artificial traffic was generated with the RUDE/CRUDE tool set [10]. The default traffic pattern was intended to simulate H.323 traffic by sending variable-sized packets irregularly spaced, replaying a sample of actual H.323 traffic. Equal packet size traffic has also been used.
- OWPL [1] has been measured with active measurement techniques using ICMP Ping. The measurements have been performed with 1 s Ping run in the background of the test stream for the transmission time.

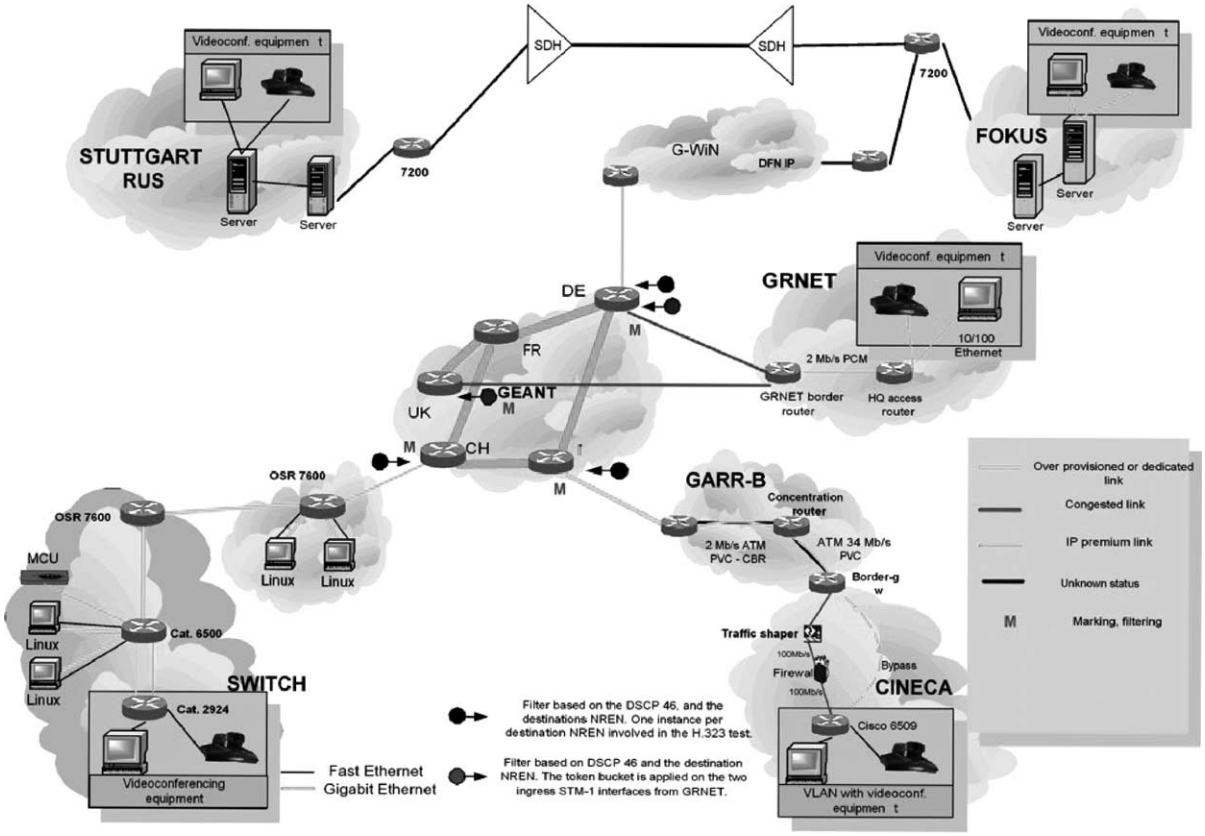

Fig. 2. The premium IP ‘proof-of-concept’ testing topology.

- Round trip time (RTT)—Due to limitations of the measurement infrastructure (requirements for fine GPS time synchronization on both endpoints), the use of ICMP Ping has been accepted for the RTT tests. In the same way as for packet loss, this measurement has been done with 1 s ping for the transmission time being.
- The capacity available for the stream [12] has been measured as the maximum IP-level throughput between endpoints. For this test, the Netperf UDP stream was used.

4.3. Test results

This paragraph shows preliminary qualitative results on the aforementioned test infrastructure. More extensive tests are in preparation. Figs. 3 and 4 show the different behavior of traffic flowing from FOKUS to GRNET as a function of its class of service. One can easily notice the positive effect in IPDV reduction obtained by the Premium IP service. No packet loss has been recorded during the test.

The set of tables below show the qualitative results gathered during multiple international measurement sessions for Premium IP. The perceived audio/video quality (Tables 3 and 4), capacity (Table 5), OWPL (Table 6) and RTT values (Table 7) are presented for all cases of end-to-end videoconferencing. The tables are reported here just as an example. Some of the tests have been performed on a network partially configured for Premium IP and partially over-provisioned. This explains why a DoS attack sometimes caused a perturbation to the service.

For Tables 3 and 4, tests with an (MCU) notation were performed with the use of an MCU unit, due to interoperability problems in videoconference

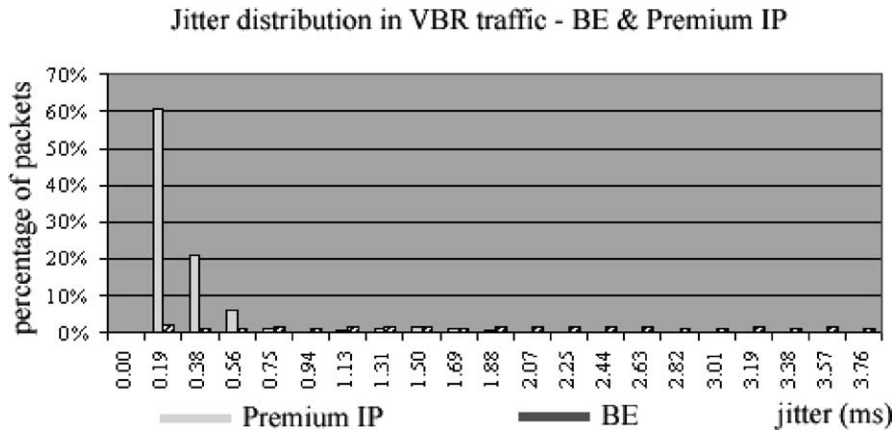

Fig. 3. Comparison of average jitter (IPDV) for different packet sizes of Premium IP and BE traffic.

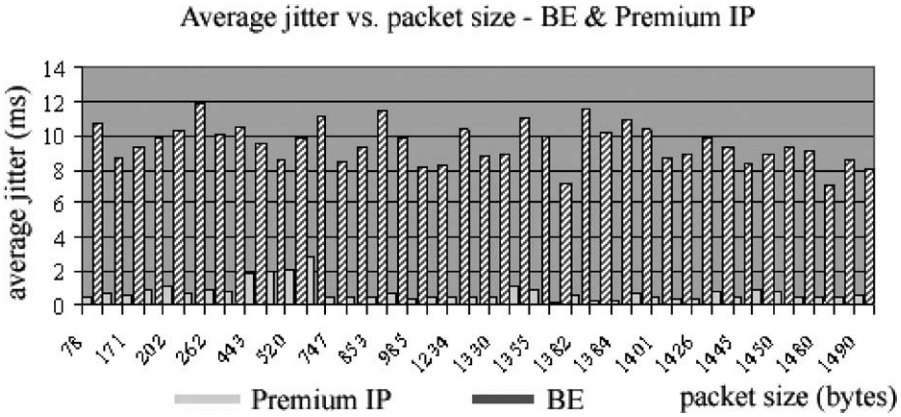

Fig. 4. Distribution of jitter (IPDV) values for Premium IP and BE traffic.

equipment. Furthermore, low results for audio transmission were mainly considered to be caused by low audio volume. For Table 5, the test with the (*) notation was performed during a DoS attack, while for all tests, guaranteed Premium IP capacity was configured at 2000 kbit/s.

In conclusion, the tests showed the importance of a spread measurement system, not available at the time,

Table 3
Perceived audio quality for videoconference transmission using Premium IP

	From				
	SWITCH	FOKUS	RUS	GRNET	CINECA
SWITCH		3 (MCU)	4–5	6	6
FOKUS	3.6		6	3	6
RUS	3.6	6		6	6
GRNET	5.4	3 (MCU)	5		6
CINECA	6	6	5	6	

Table 4
Perceived video quality of videoconference transmission using Premium IP

	From				
	SWITCH	FOKUS	RUS	GRNET	CINECA
SWITCH		6 (MCU)	5	6	6
FOKUS	4.8		6	5	6
RUS	4.8	6		4	6
GRNET	5.4	5 (MCU)	5		5
CINECA	5.4	6	5	5	

Table 5
Available bandwidth (capacity, 10^3 bit/s) for Premium IP service

	From				
	SWITCH	FOKUS	RUS	GRNET	CINECA
SWITCH		3307.87	1909.83	870.00	1816.73
FOKUS	1910.00		8725.30	910.00	1825.09
RUS	1910.00	8895.45		830.00	1835.18
GRNET	1910.00	853.41*	1909.02		1839.94
CINECA	1751.46	1944.39	1844.84	910.00	

Table 6
OWPL (%) for premium IP service

	From				
	SWITCH	FOKUS	RUS	GRNET	CINECA
SWITCH		0.00	0.00	0.02	0.00
FOKUS	0.00		0.00	0.01	0.00
RUS	0.00	0.00		0.02	0.00
GRNET	0.00	0.00	0.00		0.00
CINECA	0.00	3.07	2.70	0.25	

to analyze and quantify the QoS service behavior and pinpoint and solve problems in a complex environment.

5. Service provisioning

Service provisioning for QoS-enabled networks comprises a process where intensive testing and probing of the available infrastructure has to take place before the QoS offering can be quantified, including concrete parameters and values in the agreement. Also, during the operation of the service, monitoring of its behavior is crucial. This section outlines the approach followed in the case of the provisioning procedures for Premium IP.

The bilateral 'IP Premium' SLA specification between a Premium IP enabled domain and each one if its peers is proposed to comprise of two parts (see also [3]):

- The administrative/legal part.
- The service level specification (SLS) part, defining the set of parameters and their values, for the provision of IP Premium service to a traffic aggregate by a DiffServ domain.

The administrative/legal part of the SLA is suggested to comprise of a number of fields that will define the procedures and framework for the provision of the service for which that the SLA is established. As such, it contains fields like the administrative and technical parties involved, the SLA duration in time, SLA availability guarantees, monitoring procedures, response times by the provider in cases of client requests for adjustment of the SLA, fault handling-trouble ticket procedures, quality and performance of support and helpdesk, pricing of the contracted service and a general description of the provided service, describing qualitatively its characteristics (in terms of e.g. delay, packet loss, throughput) and operation.

The SLS part of the SLA is proposed to contain the following fields:

(i) *Scope*. The recommended field is: *ingress interface of upstream domain, set of ingress interfaces of downstream domains*.
(ii) *Flow description*. The IP Premium definition under consideration supports aggregated policing according to the packets' destination domain, and therefore classification of IP Premium packets

Table 7
RTT loss (%) for premium IP service

	From														
	SWITCH			FOKUS			RUS			GRNET			CINECA		
	Min.	Avg.	Max.	Min.	Avg.	Max.	Min.	Avg.	Max.	Min.	Avg.	Max.	Min.	Avg.	Max.
SWITCH				37.0	37.0	41.0	50.68	51.31	55.43	112.22	114.29	124.14	17.04	19.91	19.97
FOKUS	30.0	38.00	60.00				14.66	17.30	414.66	109.67	110.49	167.59	17.80	20.50	40.00
RUS	50.0	50.00	61.00	10.0	13.0	480.0				186.94	229.82	313.69	29.95	39.62	49.96
GRNET	110.0	114.00	190.00	117.0	119.0	141.00	186.90	230.20	254.80				119.80	120.04	127.82
CINECA	25.1	27.67	48.41	27.0	30.0	82.0	39.93	42.01	81.85	119.82	120.05	127.82			

must be extended to further granularity among different policers. Thus, the flow description field is: *QoS tag attribute, [source attribute], [destination attribute]*.

(iii) *Performance guarantees*. The suggested performance parameters for in-profile traffic in the case of IP Premium and their respective values are:
- *OWD*. It is suggested to be guaranteed as the maximum packet transfer delay between the scope-defined points measured. Indicative values are the distance delay plus 50 ms.
- *IPDV*. It is suggested to be guaranteed as the maximum packet transfer delay variation measured between the scope-defined points. Indicative values are equal or less than 25 ms.
- *OWPL*. It is suggested to be guaranteed as the ratio of lost in-profile packets between the scope endpoints and the injected in-profile packets at the ingress, defined by the scope field. Indicative value is 10^{-4}.
- *Bandwidth*. It is defined as the rate measured at the set of egress points (defined by the scope field) of all packets identified by the flow descriptor. As already mentioned, a suggested value for the IP Premium aggregate is 5% of ingress capacity. It is suggested that this capacity is distributed to a guaranteed throughput matrix of values corresponding to traffic from each upstream peer to each downstream peer for a Premium IP enabled domain.
- *MTU*. It is the largest physical packet size in bytes that the SLS guarantees to be transmitted without being fragmented. The suggested value for a WAN is 4470 bytes.

(iv) *Traffic envelope and traffic conformance*. The traffic conformance algorithm adopted is that of token bucket with b as the depth and r as the capacity parameters. In the particular case considered here, that is, SLSs between an NREN and GÉANT, the following values are suggested:

$$b = f(\textit{number of router interfaces on the same router that are part of the service}, \textit{distance from the source}),$$
$$r = \{1.2, \ldots, 1.5\} \times r_C$$

where r_C is the contracted capacity as defined in the 'performance guarantees' field of the SLS.

(v) *Excess treatment*. For the purposes of IP Premium dropping of out-of-profile packets is suggested.

(vi) *Service schedule*. It indicates the start time and end time of the period for which the service is provided.

(vii) *Reliability*. Reliability should define allowed mean downtime per year (MDT) and maximum allowed time to repair (TTR) in case of breakdown for the provision of the service described by the SLS.

The fact that SEQUIN deals with QoS provisioning across independently managed networks, reinforces the need for a tight mechanism for the direction of the establishment of end-to-end SLAs, based on bilateral SLAs and provisioning methodology. SLA definition between two peers is the structural unit for the establishment of end-to-end services.

However, end-to-end configuration and seamless provisioning of QoS has a number of peculiarities that must be dealt with. An end-to-end SLA (e2e SLA) is essential in co-ordinating the service's provision across multiple independently managed domains so that end-users perceive a stable and predictable service with predefined quality guarantees, regardless of the domains and bilateral SLAs involved. As depicted in Fig. 5, in order for the e2e SLA to be established, a chain of bilateral SLAs must exist in advance. The individual bilateral SLAs must be defined in a consistent manner, in such a way that no part of the end-to-end path is left uncovered. The aim of each bilateral SLA between domain D1 and domain D2 is to define the procedural and qualitative guarantees provided as D2 carries the IP Premium traffic of D1 across D2. Instead, the aim of the e2e SLA is to define the guarantees provided to IP Premium traffic originating from end-user's A premises up to the end-user's B equipment.

In order to verify an e2e SLA based on a chain of bilateral SLAs, a monitoring infrastructure has to be defined. This monitoring infrastructure should consist of:

- Monitoring equipment/functionality placed in intermediate positions along the end-to-end path from end-user A to end-user B (Fig. 5), referred to as Service Providers' Monitoring Equipment (SPME) from now on.

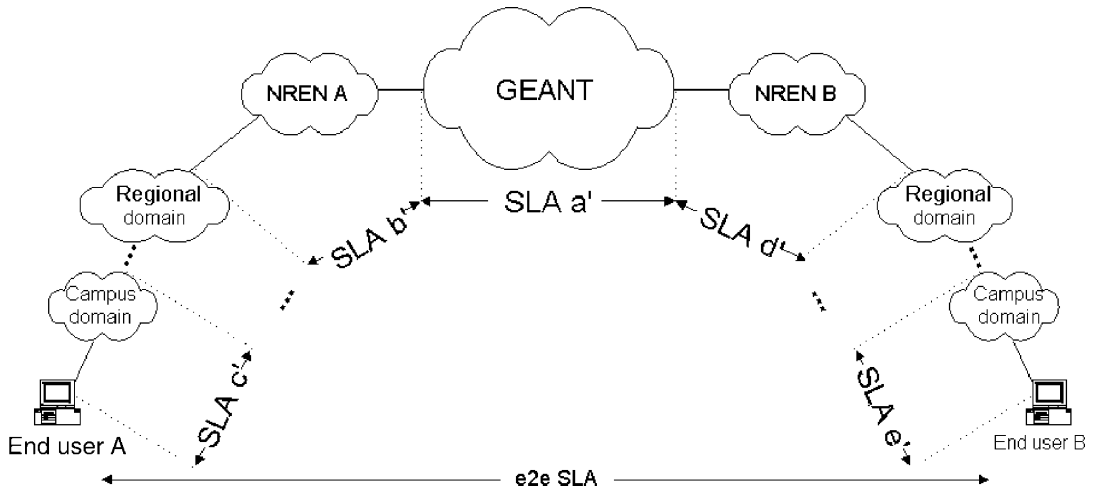

Fig. 5. End-to-end SLA establishment topology.

- Monitoring equipment/functionality located at the premises of each end-user, referred to as End-users' Monitoring Equipment (EME) from now.

As already explained, bilateral SLAs signed between service providers tend to be of a more permanent nature than e2e SLAs between end-users. Therefore, the existence of SPME is primarily essential for the establishment and monitoring of bilateral SLAs. SPME has to be located in critical positions of the domains involved in a bilateral SLA, in order to constantly monitor performance of the service provided and indicate possible causes and origins of a service malfunction.

For the case of a bilateral SLA, SPME must exist on all interfaces included in the scope field of the SLA. For example, in the case of a bilateral SLA for IP Premium connectivity between Domain 1 and Domain 2 (Fig. 6), SPME should exist on all interfaces A, B, C, D and E in order for the SLA to be properly monitored.

For its own purposes or for the purpose of monitoring bilateral SLAs with upstream domains, Domain 1 of Fig. 6 might also choose to place SPME on interfaces A′ and B′. Furthermore, each domain might choose to deploy a monitoring infrastructure within its administrative borders. This infrastructure, although not directly involved in the bilateral SLA monitoring procedure, might help in isolating deficiencies in the service provision within a domain. The latter will be particularly useful when monitoring between edge interfaces (e.g. A and C) results in violation of the bilateral SLA guarantees. Provided that bilateral SLAs along the end-to-end path between two end-users are monitored as already outlined, then the quality guarantees of each individual bilateral SLA are

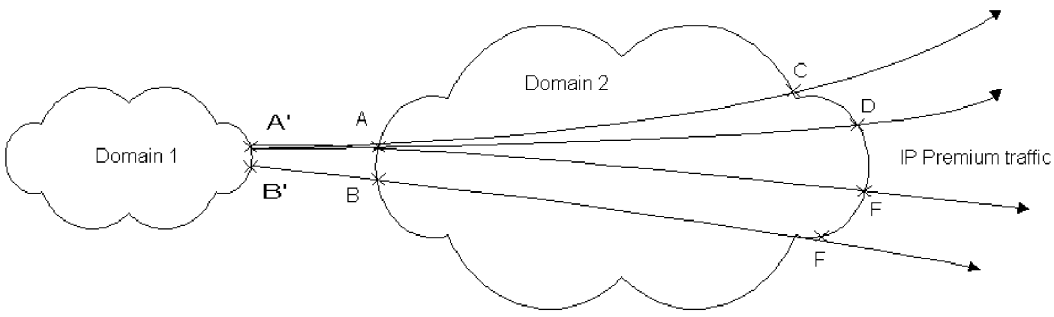

Fig. 6. Suggested locations for monitoring infrastructure supporting a bilateral SLA.

constantly monitored. They can therefore be used in the e2e SLA establishment process, in order to derive the end-to-end guarantees that can be achieved.

However, after the establishment of the e2e SLA, the end-users must also be provided with tools (EME) to verify the quality and quantity of throughput provided by the service. Due to the nature of e2e SLAs, which are of a less permanent nature than bilateral SLAs between domains, EME cannot be based on hardware and complex procedures. Therefore, it is suggested that end-users be provided with a set of software-based, active monitoring tools, referred to as software management tools (SMTs) from now on, allowing them to observe the performance of the provided service at regular intervals. SMTs are also strongly suggested because they do not require synchronization between the end-users' equipment and are therefore easier to deploy. SMTs provided to end-users must be accompanied by a set of scripts for processing the logs created during the SMTs' operation and guidelines for a set of parameters that need to be configured for each SMT's operation. An indicative selection of SMTs is the one that has already been used in the testing, carried out within SEQUIN and presented in Section 4.2 of this paper.

6. Conclusions and future work

The Premium IP service is now enabled and being tested on the GÉANT network to further validate the implementation architecture and to tune various parameters, like the token bucket depth and the increase in rate policing values at intermediate borders. The modularity of the service and its active configuration on the core backbone will allow it to become useful in a short time, even if its adoption is not yet ubiquitous, because Premium IP domains can be configured where needed. Testing has proved its effectiveness in providing an IP-based equivalent to virtual-leased-lines. Although it has not yet been possible to look at all the details, it has been shown that user requirements for QoS can be achieved based on SLAs negotiation. Future work should also concentrate on service provisioning with users in real-life scenarios as well as on the deployment of a thorough monitoring infrastructure to support Premium IP provision as well as bilateral and end-to-end SLAs.

Acknowledgements

The work presented here has received essential input from the effort of the TF-NGN [18] and SEQUIN [16] projects and the authors wish to thank all of their colleagues in these projects for their contribution. The authors would also like to acknowledge the work of H.323 users, participating in the testing phases: Alain Bidaud, Lars Burgstahler, Franca Fiumana, Ernst Heiri, Hyung-Woo Kim, Maria von Kaenel, Simon Leinen, Lutz Mark, Juergen Rauschenbach, Rudolf Roth for their work on traffic tests and active support during results analyzing. This work was also supported by SPIRENT Communications that gave its SmartBits for tests purposes.

References

[1] G. Almes, S. Kalidindi, M. Zekauskas, A one-way packet loss metric for IPPM, IETF RFC 2680, 1999.
[2] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, An architecture for differentiated services, IETF RFC 2475, 1998.
[3] C. Bouras, M. Campanella, A. Sevasti, SLA definition for the provision of an EF-based service, in: Proceedings of the 16th International Workshop on Communications Quality and Reliability (CQR 2002), Okinawa, Japan, 2002, pp. 17–21.
[4] M. Campanella, GN1 (GÉANT) Deliverable D9.1-Addendum 1: Implementation architecture specification for the Premium IP service, GÉANT, 2001. Available at: http://www.switch.ch/lan/sequin/GEA-D9.1-Addendum-1-v05.pdf.
[5] M. Campanella, P. Chivalier, A. Sevasti, N. Simar, SEQUIN D2.1: Quality of service definition, Project SEQUIN (IST-1999-20841), 2001. Available at: http://www.dante.net/sequin/QoS-def-Apr01.pdf.
[6] M. Campanella, T. Ferrari, S. Leinen, R. Sabatino, V. Reijs, GN1 (GÉANT) Deliverable D9.1: Specification and implementation plan for a Premium IP service, GÉANT, 2001. Available at: http://www.dante.net/tf-ngn/GEA-01-032.pdf.
[7] A. Charny, J.Y. Le Boudec, Delay bounds in a network with aggregate scheduling, in: Proceedings of the First International Workshop of Quality of future Internet Services (QofIS'2000), Berlin, Germany, 2000, pp. 1–13.
[8] C. Demichelis, P. Chimento, Instantaneous packet delay variation metric for IPPM, IPPM Internet Draft, 2000.
[9] V. Jacobson, et al., An expedited forwarding PHB, IETF RFC 2598, 1999.
[10] J. Laine, S. Saaristo, R. Prior, RUDE and CRUDE tools. Available at: http://rude.sourceforge.net/.
[11] S. Leinen, M. Przybylski, V. Reijs, Sz. Trocha, GN1 (GÉANT) Deliverable D9.4: Testing of traffic measurement tools, GÉANT, 2001. Available at: http://www.dante.net/tf-ngn/D9.4v2.pdf.

[12] M. Mathis, M. Allman, A framework for defining empirical bulk transfer capacity metrics, IETF RFC 3148, 2001.
[13] K. Nichols, V. Jacobson, L. Zhang, A two-bit differentiated services architecture for the internet, IETF Internet Draft, 1997.
[14] R. Oliviera, M. Przybylski, Measurements for premium IP traffic between Poland and Switzerland national research networks over TEN-155, Test Report, 2001. Available at: http://qos.man.poznan.pl/files/psnc-unibe.pdf.
[15] V. Paxson, G. Almes, J. Mahdavi, M. Mathis, Framework for IP performance metrics, IETF RFC 2330, 1998.
[16] SEQUIN Consortium, SEQUIN-service quality over interconnected networks, Project SEQUIN (IST-1999-20841). See also: http://www.dante.net/sequin.
[17] A. Sevasti, M. Campanella, SEQUIN D2.1-Addendum 2: Service level agreements specification for IP premium service, Project SEQUIN (IST-1999-20841), 2001. Available at: http://www.switch.ch/lan/sequin/SEQ-01-043.pdf.
[18] TF-NGN Task force. See: http://www.dante.net/tf-ngn.
[19] TF-STREAM Task force. See: http://www.terena.info/task-forces/tf-stream/.
[20] T. Ferrari, et al., Quantum deliverable D6.2: Report on results of quantum test programme, QUANTUM Project EP 29212, 2000. Available at: http://www.dante.net/quantum/qtp/final-report.pdf.

Christos Bouras obtained his Diploma and PhD from the Computer Science and Engineering Department of Patras University (Greece). He is currently an Assistant Professor in the above department. Also he is a scientific advisor of Research Unit 6 and director of Networking Technologies Sector, in Computer Technology Institute (CTI), Patras, Greece. He has extended professional experience in Design and Analysis of Networks, Protocols, Telematics and New Services, Distance Learning and Education, Electronic Publishing, High Speed Networks and Interactive Hypermedia and Multimedia. He has published 100 papers in various well-known refereed conferences and journals. He is a co-author of five books in Greek. He has been a PC member and referee in various international journals and conferences. Also he is member of team of experts in the Greek Research and Technology Network (GRNET), ACM, IEEE, EDEN, AACE and New York Academy of Sciences.

Mauro Campanella is presently working for the National Institute for Nuclear Physics (INFN) in Italy. His main activity is related to the Italian Research Network (GARR) as senior engineer. He is actively involved in the engineering of the new generation of the Italian network and he has been one of the author of the project of the present one. He is also working on testing and definition of advanced services for the GÉANT European backbone lately in the QoS and high speed area. He holds a laurea in physics, has spent 1 year at CISCO systems between the years 1999 and 2000 and lives north of Milan in Italy.

Michal Przybylski was born in 1976. He obtained his Engineer degree from Poznan University of Technology in 2000. He works at Poznan Supercomputing and Networking Center, where he is responsible for implementation and introduction of advanced networking technologies into Polish National Research Network. His responsibilities also include coordination of networking projects in PSNC.

Afrodite Sevasti obtained her Diploma and her Master's degree (MSc) from the Computer Engineering and Informatics Department of Patras University (Greece). She is currently an R&D Computer Engineer in Research Unit 6 at the Computer Technology Institute, Patras, Greece. She is also attending the second year of studies to obtain her PhD degree at the Computer Engineering and Informatics Department of Patras University. She has published 16 papers in well-known refereed conferences and journals. She has participated in several R&D projects among which is the project SEQUIN—"service quality across independently managed networks" (IST-1999-20841) and the project "design and implementation of a QoS architecture over GRNet" of the Greek Research and Technology Network (GRNET).