Contract Number: IST-2000-26417
Project Title: GN1 (GÉANT)

Deliverable D9.1 - Addendum 2

Service Level Agreements specification for IP Premium Service

Contractual Date:	
Actual Date:	
Work Package:	WP2
Work Item:	WI2.1
Nature of Deliverable:	R - Report
Dissemination Level:	PU - Public

Authors: Afrodite Sevasti GRNET
Mauro Campanella INFN/GARR

Abstract:

The document elaborates a template for the Service Level Agreement for the Premium IP service. The template is initially targeted for the Service Level Agreements between an NREN and the GÉANT network, but is general enough to be applied between any communicating DiffServ domains in order for them to provide the IP Premium service. It states also the requirements to provide an end to end service on the multi-domain european network.

Table of Contents

EXECUTIVE SUMMARY

The Service Level Agreement for the Premium IP service is a corner stone in defining and implementing the service. The document elaborates a template that can be tailored to each instantiation between the involved domains for the IP Premium service as defined in the framework of the GEANT (IST-2000-26417) and SEQUIN (IST-1999-20841) projects and more specifically in [GN-D9.1] and [SEQ-D2.1].
The document elaborates further on the implications for the SLA of the goal to provide an end to end service on the multi-domain european network.

Due to the complex structure of the global network there is a strong need for standardisation of the SLA and associated procedures.

SLA specification for QoS enabled networks aims at providing positive quality guarantees and setting out the limits of the services provided. Such SLAs move one step forward in the direction of traditional ones, in the sense that they do not only have to specify availability, security, quantity of allocated resources and a number of other quantitative values but also have to specify the values of appropriate quality parameters.

In networks where QoS is inherently supported (such as ATM) the provision of SLAs comes as a natural delimitation of the relevant parameters. However, in IP networks where best-effort traffic has no quality guarantees, the introduction of QoS and associated services requires a thorough and accurate engineering of QoS metrics in the SLA specification on top of the guarantees for availability and characteristics of the transport medium, security, fault handling etc.

The analytical computation of such metrics is extremely complex taking into consideration the extensive level of aggregation and more generally the nature of traffic flowing in large interconnection domains. Usually only upper bounds for the relevant parameters can be defined.

Therefore, SLA specification for QoS enabled networks becomes a process where intensive testing and probing of the available infrastructure has to take place, before being able to quantify the QoS offering and include concrete parameters and values in the agreement. Also, during the operation of the service monitoring its behaviour is crucial.

This deliverable has been produced by individuals working on the GÉANT and SEQUIN projects and is therefore submitted for both projects. It is submitted as D9.1 Addendum 2 for GÉANT and as D2.1 Addendum 2 for SEQUIN.

1 INTRODUCTION

The provisioning of services over an IP network with an assured level of quality is today often associated with the negotiation of service contracts between customer and provider. Until today, the instantiation of Service Level Agreements (SLAs) between customers and providers has been a rather static and labour-intensive task. The procedures involved in this process are proprietary to the provider, and, in many cases these procedures are invoked on a low frequency basis. By its proprietary nature, such a process does not allow for an open service architecture to be built upon an IP network.
It is to be understood that standardisation of the technical parts of the basic process may allow for a highly developed level of automation and dynamic negotiation of Service Level Specifications (SLSs) between customers and providers. This automation may prove helpful in providing customers (as well as providers) the technical means for the dynamic provisioning of quality of service guaranteed transport services.

There exist several efforts towards the standardisation of definition of SLAs and their instantiation in QoS enabled networks ([QBONE-BB], [TEQ-D1.1], [SOME-SLS], [AQU-SLS]). This document makes an attempt to define an initial template for the IP Premium service as defined in the framework of the GEANT (IST-2000-26417) and SEQUIN (IST-1999-20841) projects and more specifically in [GN-D9.1] and [SEQ-D2.1].

GÉANT and its connected networks could substantially benefit from the application of such mechanisms over the network infrastructure under consideration, since SLAs (and the corresponding SLSs) are the essential mechanisms for agreeing, configuring, delivering, guaranteeing, and evaluating the obtained QoS from one end-user to another across multiple domains. Therefore, SLAs are a means to provide a service to users (for example for GEANT to provide the IP Premium service to NRENs) and are particularly useful in defining the framework for the provision of the service: the provider's commitments, the user's obligations, what happens when the rules are broken.

SLA definition between two peers is also the structural unit for the establishment of end-to-end services. Provided that SLAs are properly defined all the way from the desired origin to the desired destination, proper mechanisms (such as the Bandwidth Brokers, see also [QBONE-BB2], [QBONE-BB2], [Frankhauser99], [RFC 2638]) can evaluate all connections between consecutive peers and determine the resources (according to the SLAs) that are available for servicing requests for the specific service. This procedure can successively conclude with a valid outcome on whether the end-to-end service can be provided or not, based on the individual SLAs, and which are the specific quality features of the service provided.
Based on this outcome, an end-to-end SLA will have to be established, defining all related parameters to SLA specification between the two points (origin and destination) regardless of the underlying infrastructure and intermediate domains. The establishment of the SLA can range from static to fully dynamic (see example below). For the purpose of end-to-end service provision, it is essential that each SLA is carefully designed and engineered to effectively reflect the QoS capabilities a single domain can offer to its peers, by containing specific traffic shaping and policing parameters, description of the flows being serviced according to it, performance guarantees etc.
Until fully dynamic mechanisms for SLA establishment and negotiation are standardised and fully deployed, the provision of end-to-end services will be based on investigation of the SLAs along the path and specified as the worst-case scenario. In cases where SLAs negotiation is possible, provision of end-to-end QoS services can be more flexible, involving better utilisation of resources by dynamic adjustment of SLAs.
For the early stages of the IP Premium service, as specified by SEQUIN, the static scenario will be followed. More specifically and in accordance with the Phase 0 of the QBONE architecture, after SLSs are negotiated and established according to a matrix that specifies the amount of uni-directional IP Premium aggregate traffic from one NREN to another through GEANT, edge routers in peer

domains that share an SLS will be configured to service packets (police, mark, queue, schedule, shape, etc.) according to the information in the SLSs. There will be no inter-domain communication (performed by BBs in QBONE), so the SLS negotiation will be performed manually (telephone, email, etc.) by the network operators of the two domains.
This document will initially attempt to define a suggested SLA (and consecutively SLS) template between GEANT and the NRENs attached to it.
Since a Service Level Agreement (SLA) is a combination of technical and non-technical parameters agreed by a customer and a provider related to the quality of a 'service', when acquiring/selling the service, the SLA specification is proposed to be comprised of two parts:

- The administrative/legal part
- The SLS part, defining the set of parameters (SLS template) and their values, which together define the service offered to a traffic stream by a DiffServ domain. A Traffic Conditioning Specification (TCS) will also be an integral element of an SLS. A TCS is a set of parameters and their values, which together specify a set of classifier rules and a traffic profile.

After the definition of the SLA specification, several instantiations of it will have to be produced, one for each of the several peering AS couples involved in the GEANT-NRENs architecture. Next steps will be to define the mechanisms for SLA negotiation and, of course, for the establishment of end-to-end services based in the individual SLAs.
Each instantiation of a SLS will comprise a so-called Service Level Object (SLO) and will contain the parameters and their values that describe the transport service a specified flow is to receive over the transport domain.
Bi-directional services will also be possible by the combination of two Service Level Objects that will be taken atomically when negotiating a service pertain to two flows, one at each direction. These SLOs will comprise a Transport Service, which will be part one SLA defined between the two AS, among which the bi-directional service is established. Figure 1 displays an SLA template and several instantiations of it, bringing the aforementioned individual SLA components together. The SLA instantiation on the left is an example of a bi-directional SLA containing two uni-directional SLOs.

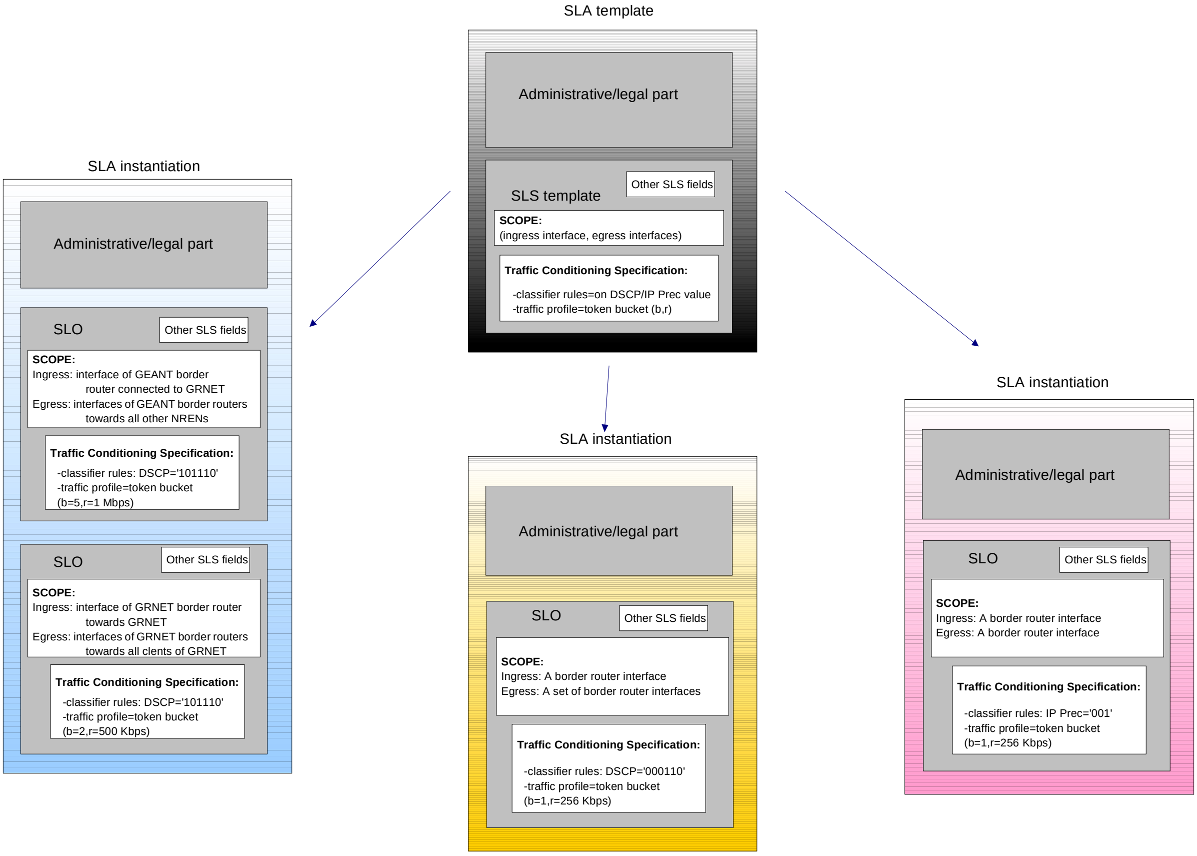

Figure 1. SLA template, SLA instantiations, SLS and SLOs

2 THE ADMINISTRATIVE/LEGAL PART OF THE SLA

This part is suggested to comprise of a number of fields that will define the procedures and framework for the provision of the service that the SLA is established for. Each of the fields should define one or more of the parameters ‘scope’, ‘process’, ‘remedy’ that describe the scope of the current field, the process for dealing with it and, in cases where the SLA is violated or faults occur, how will recovery take place or how will the problem be solved:

- Administrative and technical parties involved (scope): This section should contain at least one administrative and one technical contact from each of the two sides participating in the SLA
- Duration in time (scope): This section should contain the period for which the SLA is valid. This period can differ from the period defined at the ‘service schedule’ field of the SLS part of the SLA, but the value of the ‘service schedule’ field has to be a period WITHIN the period defined at this section of the SLA. The ‘service schedule’ is a set of time periods for which the service is active, while the SLA duration is a time period for which the SLA for the service’s provision is valid.
- Availability guarantees (scope, process, remedy): This sections should define the calculation of the service’s availability figures and how these will be derived (e.g. from the trouble ticketing system). The section should also provide a service availability ratio according to the SLA’s duration in time in comparison, an Unavailable Time Limit (UTL) and formulas for the calculation of compensation for unavailability
- Monitoring (scope, process): This section should specify how and when (constantly vs. periodical) will the SLA be monitored. It should specify the points of network topology where monitoring equipment is installed or where measurements are retrieved from. It should also specify the SLS metrics that will be visible to the client (see also section 4 of this document) and how the client will have access to this monitoring data.
- Response times (scope): This section concerns the overall response times guaranteed by the provider in cases of client requests for adjustment of the SLA (and/or SLS) and for necessary configuration of the relevant devices
- Fault handling-trouble ticket procedures (scope, process, remedy): This section should specify the actions taken by the provider when faults concerning the delivery of the service defined in the SLS occur and the reaction times.
- Quality and performance of support and helpdesk (scope, process): This section should thoroughly specify the contracted service’s support infrastructure
- Pricing of the contracted service: Pricing of the service provided is a crucial part of a SLA between a client and a provider of network services. There is extensive on-going research work on the issue of pricing for differentiated services ([Parris92], [Kelly94], [Semret98], [Marbach01], [Courcoubetis99], [Cocchi91], [Bailey00] etc.) and IP Premium certainly falls into this category. In order for a differentiated services pricing scheme that will efficiently reflect the service value and will maximise or meet several criteria (client revenue, efficient resource allocation, accepted service requests etc.) a very thorough and interdependent with SLAs, monitoring and accounting infrastructures’ study has to be made. This study, however, is not within the scope of the SEQUIN project.
- Description of the service: A general description of the provided service, describing qualitatively its characteristics (in terms of e.g. delay, packet loss, throughput) and operation

3 THE SLS

The standardisation of a template for Service Level Specifications should follow similar directions when it comes to SLSs between

- an end-user and the domain he is attached to
- two transport domains

Especially in the case of SEQUIN, where one of the transport domains plays the role of the interconnection domain (GEANT), the same SLS template can be used to define different instantiation of services between:

- an end-user and the domain he is attached to
- two transport domains or a transport domain and the interconnection domain

This document makes an attempt to provide a specification for a SLS template for IP Premium traffic between a transport domain and the interconnection domain initially, that could be later also used for the creation of SLOs along the transmission path from one end to another.

Following the recommendation of relevant bibliography, SLSs must be established in an uni-directional manner. Since the definition of IP Premium on which this document is based anticipates for aggregate policing of traffic from a single NREN at the ingress of GEANT, according to the destination NREN, an SLO (belonging to the SLA between an NREN and GEANT) that describes how IP Premium traffic is served should feature:

- a wide scope (covering all possible destinations of traffic exiting a NREN domain)
- a unique flow identification field (all Premium IP traffic from an NREN towards GEANT should deserve the same DSCP) and
- a vector of policer instantiations, one for each destination NREN.

In other words, the SLA between GEANT and NREN, will contain one or more SLOs (according to how many ingress point to GEANT each NREN has, see more on this below), the TCS of which, will contain a vector specifying how the aggregate Premium IP traffic from the NREN towards GEANT is policed depending on which NREN is it heading towards.

This is the basic structure of the proposed SLS that is further analysed in the sequel.

3.1 Proposed template for SLS between an NREN and GEANT

3.1.1 Scope

The scope should define the topological region to which the IP Premium service defined at the SLS will be provided. This field, according to [RFC 2475], must specify where packets conforming to the SLS are entering and exiting a DiffServ domain (in our case GEANT).

Recommended for the case of SEQUIN is the field:

(ingress interface of upstream domain, set of ingress interfaces of downstream domains)

where the 'ingress interface of upstream domain' will specify the interface of a GEANT border router to which NREN X is connected and the 'set of ingress interfaces of downstream domains' will specify the set of ingress interfaces where packets injected to GEANT from X can enter other NRENs.

It is recommended that at the first phase, all SLOs are constructed based on the hypothesis that there is only one point of connection between GEANT and each NREN at interface e.g. y. In this case, the SLA between each NREN and GEANT will contain only one SLO for IP premium, with the y as value for the 'ingress interface of upstream domain' part of the scope field.

The 'set of egress interfaces of downstream domains' part has to be further discussed, since it should not in any case be interpreted as some attempt to define policing rules at the packets exiting from the specified interfaces. It can be thought of as an attempt to set the geographical limits for the provision of the specified by the SLS service. Suggested value of 'set of egress interfaces of downstream domains' for the SLS between NREN X and GEANT (unless policing or the routing scheme enforce something different) is the set of ingress interfaces of all other NRENs (except for NREN X) where traffic originating from NREN X can enter.

Of course, if the hypothesis for one interface per NREN is removed, then the definition of the scope field, as far as the 'set of ingress interfaces of downstream domains' is concerned will have to be reconsidered and the 'set of ingress interfaces of downstream domains' part might even have to take value 'none' (destination un-aware SLS).

3.1.2 Flow description

A SLA can contain more than one SLSs (and consecutively SLOs) but each one of them has to concern one strictly specified flow in one direction. The flow description field will indicate for which IP packets the QoS guarantees of the specific SLS is to be enforced or in other words, which packets will receive the PHB treatment resulting in the QoS guarantees of the SLS.

The flow descriptor will be the IP source and destination prefix pair in the first domain and will be the DSCP or IP Precedence value for the remaining domains. The options here are either to mark packets with the same value of the QoS tag on all DiffServ domains or not, with the first choice being highly recommended.

From a first point of view, the DSCP or IP precedence value is all that is needed to include in the flow description field when it comes to an SLS between an NREN and GEANT, since it can uniquely identify the packets to receive IP Premium treatment among all packets injected from the NREN to GEANT and the packets have already been through a control admission procedure in the NREN.

However, in order for the interconnecting network (GEANT) to be able to distinguish between packets from different clients (NRENs) (for monitoring, accounting, policing etc. reasons), the flow description should also contain the sending NREN's identification (still, if we have one access interface per NREN and no routing backups between NRENs, the source NREN's id will be redundant in the flow description field).

The flow descriptor plays the role of the classifier for IP packets at a DS boundary node. As such, it has already been mentioned that the QoS tag may be more than enough to distinguish IP Premium packets from non-IP Premium ones. However, the IP Premium definition under consideration supports aggregation policing according to the destination DS, and therefore classification of IP Premium packets must be extended to further granularity among different policers. In order for this to be feasible, the destination NREN's identity has to be included as part of the flow identification field. This attribute should specify which of the multiple policers defined in the SLO is to be used for each IP Premium packet entering GEANT from a specific NREN.

The general form of the flow description field should then be:

(QoS tag attribute, source attribute, destination attribute)

with the source attribute being a unique value representing the NREN involved in the SLS and the destination attribute being a set of values (*s*) representing the possible destination NRENs. Each packet will then have contained the unique value for the source attribute and another unique value for the destination attribute (belonging to the SLS-defined set *s*), with the latter serving as an index to the throughput vector of the performance guarantees of the SLS. The different options for the QoS tag attribute are described in the following section.

The form of the description field proposed here suggests a Multi-Field (MF) classification which, as it has already been analysed, is necessary for supporting of aggregate macro-flows based on the destination NREN id (e.g. network prefix).

3.1.2.1 Options for the QoS tag attribute

As it has already been mentioned, the QoS tag for IP Premium throughout the network structure under consideration can be:

(i) either a global DSCP (e.g. the 'EF' DSCP), or

(ii) a value being mapped from one DSCP to another as traffic crosses through domains, with the consecutive mappings being explicitly specified in consecutive SLSs.

In the **first case**, the first border (ingress from the end user domain) checks all packets to see if they have a valid (IP source, IP destination address) pair. If they do, they are marked with EF DSCP, policed according to the SLS and either dispatched to the high-priority queue or dropped. Packets with an EF DSCP but no valid IP address pairs are suggested to be remarked as BE. In all other ingress borders, only packets with the 'EF' DSCP are considered and policed according to the current SLS. From these packets, conforming ones are dispatched to the high priority queue, while non-conforming ones are dropped. In this case, the QoS tag value (*q*) is globally the same and therefore can either be included in all SLAs or not. The flow descriptor field therefore will be:

(q, source NREN id, set of destination NREN ids)

In the **second case**, the first border checks all packets to see if they have a valid IP address pair. If they do, they are marked with the EF DSCP, policed according to the SLS and dispatched to the high-priority queue while the violating ones are dropped. If they have a correct EF DSCP tag, but do not have a correct IP source and destination address, they are suggested to be re-marked as BE. In all other borders, the border is policing all packets with the contracted DSCP from the previous domain ($DSCP_n$) according to the SLS. From these packets, all conforming ones are remarked with the next valid DSCP ($DSCP_{n+1}$) and dispatched to the high-priority queue, while non-conforming ones are dropped. Packets arriving with $DSCP_{n+1}$ are remarked as BE
In this second case, it is essential for the SLS to define the value of $DSCP_{n,}$ so as to distinguish the IP Premium flow from the NREN towards GEANT. It should be pointed out though, ([TEQ-D1.1]) that the $DSCP_n$ value specified in the SLS is independent of the DSCP re-marking of packets to $DSCP_{n+1}$. The latter (actually can be more than one values) is used for differentiating packets according to PHBs inside the downstream DS and is not used in classification of packets at ingress, therefore does not have to be part of the flow description field of the SLS.
The flow descriptor field therefore will be:

($DSCP_n$, source NREN id, set of destination NREN ids)

REMARK: As pointed out in [TEQ-D1.1], the IP routing scheme MAY put restrictions on combining scope and flow description within an SLS. It might not be possible for a flow described by the flow description field to be serviced within the scope defined in the scope field e.g. traffic from NREN *a* to NREN *b* is routed via some specific egress interfaces of GEANT that MUST be included in the scope otherwise support of IP Premium traffic between the two NRENs will not be possible.

3.1.3 Performance guarantees

The performance guarantees field depicts the guarantees that the network offers to the customer for the packet stream described by the flow descriptor over the topological extent given by the scope value.
In the case of SLS between an NREN and GEANT, this field provides seamless quality guarantees for the aggregate IP Premium flow injected from the NREN to GEANT and differentiates only in the capacity value guaranteed among the macro-flows resulting from the aforementioned MF classification of IP Premium marked packets according to the destination NREN.
The suggested performance parameters for in-profile traffic in the case of IP Premium and their respective values are:

- One-way delay: It is suggested to be guaranteed as the maximum packet transfer delay between the scope-defined points measured. Indicative values are the distance delay + 50ms. The distance delay can be roughly computed using a signal speed of about 7 us/Km. A quintile could also be optionally defined to specify the delay guarantee in 99.5% of the cases, since users might find the worst-case figure misleading.
- IPDV: It is suggested to be guaranteed as the maximum packet transfer delay variation measured between the scope-defined points. Indicative values are <25ms. Again a quintile could also be optionally defined to specify the IPDV guarantee in the majority of cases.
- One-way packet loss: 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. It is suggested that the appropriate numbers will be based on the actual contracted values for the transmission lines and modified (increased) to take into account the service-induced figures.
- Rate: 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. A suggested value for the IP Premium aggregate is

5% of ingress capacity. For the NREN-GEANT SLS it is suggested that this capacity is distributed to a guaranteed throughput vector of values corresponding to traffic from the NREN under consideration towards all other NRENs. The value of each vector item is the throughput guaranteed by the SLS for traffic originating at the NREN under consideration and terminating at a different NREN each time, through the interconnecting network. Each one of these values is used to calculate the rate parameter of each macro-aggregate policer at the ingress interface (see next section).

- 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.

Of course the above values are suggestions/estimations that will be refined and adjusted to reflect as well as possible the guarantees provided. More accurate values will be approximated during the testing period and later on during the real use and monitoring of the IP Premium operation.

3.1.4 Traffic Envelope and Traffic Conformance

The traffic envelope is a set of traffic conformance parameters describing how the stream of traffic from the NREN towards GEANT should look like in order to get the guarantees indicated by the performance parameters of the SLS. The traffic conformance algorithm itself is part of the SLS, describes how is traffic examined against the targeted/contracted behaviour and has as its input the traffic conformance parameters. It is possible to have either a binary-based or a multi-level based TC algorithm, but in the case of IP Premium, a binary-based algorithm identifying packets as either 'in-profile' or 'out-of-profile' is appropriate.
The IP Premium service aims at offering the equivalent of an end-to-end virtual leased line at the IP layer. Therefore, the conformance parameters are conformance to a shape and a limit of throughput/capacity. The traffic conformance algorithm adopted is that of token bucket with b as the depth and r as the rate parameters.
Within the SLS framework, the following specification for the traffic envelope and conformance field is proposed:

- Conformance parameters = (b, r)
- Conformance algorithm = the (b,r) token bucket

In the particular case considered here, that is SLSs between an NREN and GEANT, the following values are suggested:

$$b = f(\text{number of router interfaces on the same router that are part of the service, distance from the source})$$
$$r = 1.5*c$$

where c is the contracted capacity as defined in the performance guarantees field of the SLS.
However, as it has already been mentioned, policing of traffic in an NREN-GEANT SLS is not performed uniformly to the IP Premium flow aggregate injected by the NREN to the GEANT ingress. Instead, the entire flow aggregate is divided into several macro-flow aggregates according to the NREN the packets are heading towards and each macro-flow is policed by a different policer obeying to the conformance parameters and algorithm already specified, but with a different c value. Actually the c of each individual policer is derived from the throughput vector defined at the performance guarantees' field of the SLS.
In that way, the suggested policing per each ingress NREN-egress NREN pair using a capacity value greater than the sum of contracted values between the two NRENs at the specific ingress (between 1.2 and 2 times larger) and a token bucket with depth of at least 5 packets or more, will hopefully result in minimum packet loss.

3.1.5 Excess treatment

This attribute specifies how excess traffic (or out-of-profile traffic, according to the profile described by the traffic envelope and traffic conformance field) is treated. For the purposes of IP Premium dropping of out-of-profile packets is suggested.

3.1.6 Service schedule

Indicates the start time and end time of the period for which the service is provided. It is suggested to be of month range, either a single month or a group of sequential months.

3.1.7 Reliability

Reliability should define
- allowed mean downtime per year (MDT)
- maximum allowed time to repair (TTR) in case of breakdown

for the provision of the service described by the SLS. The values of these guarantee parameters must be compliant with the guarantees provided via the administrative part of the SLA.

4 USER VISIBLE SLS METRICS

Recent results of the TF-NGN on QoS monitoring and SLS auditing work point out more or less the same performance metrics as those defined in the IP Premium SLS are useful and important when it comes to user perceived QoS. This work suggests
- Available bandwidth
- One-way packet loss
- Burstiness of one-way packet loss
- IPDV and
- One-way delay

as the metrics that interest users most and depict realistically the QoS provided to them. It is obvious that in the later stages of IP Premium deployment and using an efficient measurement infrastructure, these user visible metrics will be fine-tuned as the SLS fields also will, in order to provide a clear picture of the provided QoS, both between peer domains and from one end to another.

5 REFERENCES

[AQU-SLS] S. Salsano, F. Ricciato, M. Winter, G. Eichler, A. Thomas, F. Fuenfstueck, T. Ziegler, C. Brandauer, "Definition and usage of SLSs in the AQUILA consortium", draft-salsano-aquila-sls-00.txt, Internet Draft, November, 2000

[Bailey00] J. Bailey, J. Boroumand, "The Economics of Internet Differentiated Services: Creating Markets for Leaky Token Buckets", Fifth INFORMS Telecommunications Conference, March 5-8, 2000, Boca Raton, Florida

[Cocchi91] "A study of priority pricing in multiple service class networks", R. Cocchi, D. Estrin, S. Shenker, L. Zhang, ACM SIGCOMM 91

[Courcoubetis99] "Managing and Pricing Service Level Agreements for Differentiated Services", C. Courcoubetis and V. Siris, In Proc. of IEEE/IFIP IWQoS'99, UCL, London, UK, May 31-June 4, 1999

[Frankhauser99] G. Fankhauser, B. Plattner, "Diffserv Bandwidth Brokers as Mini-Markets", Workshop on Internet Service Quality Economics, MIT, US, Dec. 2-3, 1999

[GN-D9.1] M. Campanella, "Implementation architecture specification for an Premium IP service", Deliverable D9.1 - Addendum 1, GEANT Project (IST-2000-26417), July 2001

[Kelly94] F.P. Kelly, “On tariffs, policing and admission control for multiservice networks”, Operations Research Letters, 15:1--9, 1994.

[Marbach01] “Pricing Differentiated Services Networks: Bursty Traffic”, P. Marbach, in Proc. of INFOCOM 2001

[Parris92] C. Parris, S. Keshav, and D. Ferrari, “A framework for the study of pricing in integrated networks” Technical Report TR-92-016, International Computer Science Institute, Berkeley, CA, March 1992.

[QBONE-BB] "QBone Bandwidth Broker Architecture”, Work in progress, http://qbone.internet2.edu/bb/bboutline2.html

[QBONE-BB2] R. Neilson, J. Wheeler, F. Reichmeyer, S. Hares (editors), “A Discussion of Bandwidth Broker Requirements for Internet2 Qbone Deployment”, Internet2 Qbone BB Advisory Council August 1999, Version 0.7

[RFC 2475] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, An Architecture for Differentiated Services, December 1998.

[RFC 2638] K.Nichols, V.Jacobson, L.Zhang, A two-bit differentiated services architecture for the Internet, July1999.

[Semret98] “Market Pricing of Differentiated Internet Services”, N. Semret, R. R.-F. Liao, A. T. Campbell and A. A. Lazar, Technical Report CU/CTR/TR 503-98-37, Dept. of. Electrical Engineering, Columbia University, 1998

[SEQ-D2.1] M. Campanella, P. Chivalier, A. Sevasti, N. Simar, “Quality of Service Definition”, Deliverable D2.1, SEQUIN Project (IST-1999-20841), 31 March 2001

[SOME-SLS] R. Rajan, E. Celenti, S. Dutta, “Service Level Specification for Inter-domain QoS Negotiation”, draft-somefolks-sls-00.txt, Internet Draft, November 2000

[TEQ-D1.1] D. Goderis, G. Cristallo, Y. T’Joens, P. Georgatsos, L. Georgiadis, C. Duret, J.P. Garbisu, C. Jacquenet, S. v.d. B., P. V. Heuven, D. Giannakopoulos, E. Mykoniati, G. Memenios, H. Asgari, R. Egan, D. Griffin, L. Sacks, C. F. Cavalcanti, A. Liotta, G. Pavlou, I. Andrikopoulos, P. Flegkas, P. Trimintzios, “D1.1: Functional Architecture Definition and Top Level Design”, TEQUILA Project “Traffic Engineering for Quality of Service in the Internet, at Large Scale” (IST-1999-11253), 11 September 2000

APPENDIX: AN SLA INSTANTIATION EXAMPLE (GRNET <-> GEANT IP PREMIUM SERVICE)

Part A: The administrative/legal part

- Administrative and technical parties involved:
  - The GRNET NOC and administration contact info
  - The GEANT IP Premium technical responsible and administration contact info

- Duration in time:
  - The year of 2002
- Availability guarantees:
  - Service availability is calculated by projecting the failure trouble ticketing incidents to the total of service requests
  - Guaranteed service availability 99.9%,
  - Unavailable Time Limit (UTL) = 44 minutes/calendar month
  - Reimbursement for anavailability

    R={(UT-UTL)/8400} *T

    where UT: unavailable time in minutes/calendar month

    T: service tarrif per month
- Monitoring:
  - Monitoring will be performed on a per minute (??) basis between points defined at the scope field of the SLS
  - Monitoring will be performed for the values of one-way delay, packet loss etc.
  - The monitoring metrics available to the GRNET will be one-way delay, packet loss etc. and will be displayed to him via a user interface
- Response times (scope):
  - Changes in the SLA (and/or SLS) have to be requested or advertised one week in advance, which is the maximum time for enforcement of changes (configuration of the relevant devices)
- Fault handling-trouble ticket procedures:
  - A help desk via telephone, mail and fax will be available 99.9% of the time.
  - In case of any kind of failure, a trouble ticket will be issued within 15 minutes of the incident being reported and sent to the GRNET NOC
- Quality and performance of support and helpdesk:

- This section should thoroughly specify the GEANT's support infrastructure

- Pricing of the contracted service: ??
- Description of the service
  - The Premium IP service aims at offering the equivalent of an end to end virtual leased line service (guaranteed throughput, bounded delay, negligible packet loss) at the IP layer across multiple domains.

Part B: The SLS

- Scope: (ingress interface of GEANT border router towards GRNET, set of ingress interfaces of all other NREN border routers attached to GEANT)
- Flow identification: (DSCP=101110, $GRNET_{id}$, {$NREN1_{id}$, $NREN2_{id}$, ….})
- Performance guarantees
  - x ms delay in 99.5% of the cases
  - y ms IPDV
  - z % lost packets
  - vector of guaranteed throughput from GRNET towards all other NRENs via GEANT

	$NREN1_{id}$	$NREN2_{id}$	…	$NREN(n\text{-}1)_{id}$
Throughput	10 Mbps	20 Mbps	…	30 Mbps

  - MTU=4470 bytes
- Traffic Envelope and Traffic Conformance
  - Conformance algorithm = the (b,r) token bucket
  - Conformance parameters = (x, w) where w = f(the total number of DiffServ ingress interfaces at the GEANT border router towards GRNET, mean of distance-in hops-of each transmission source until the border of GEANT where GRNET is attached to) and w as defined by the following vector

	$NREN1_{id}$	$NREN2_{id}$	…	$NREN(n\text{-}1)_{id}$
w	15 Mbps	30 Mbps	…	45 Mbps

- Excess treatment
  - Excess traffic (or out-of-profile traffic) is dropped
- Service schedule
  - The year of 2002
- Reliability
  - allowed mean downtime per year (MDT) = 528 min
  - maximum allowed time to repair (TTR) in case of breakdown= 60 min