

17.03.05

## Deliverable D.S.3.9.1: Policy for allocation of Premium IP



### Deliverable D.S.3.9.1

Contractual Date:	30/02/05
Actual Date:	17/03/05
Contract Number:	511082
Instrument type:	Integrated Infrastructure Initiative (I3)
Activity:	SA3
Work Item:	9 (Nine)
Nature of Deliverable:	R - Report
Dissemination Level	RE - Restricted
Lead Partner	GRNET
Document Code	GN2-05-017v7

**Authors:** Afrodite Sevasti (GRNET), Michal Przybylski (PSNC), M. Campanella (GARR), M. Marletta (GARR), F. Ferri (GARR), I. Pezelj (CARNet), M. Klobucar (CARNet), M. Klobucar (CARNet), T. Rodwell (DANTE), V. Olifer (UKERNA), F.X. Andreu (RENATER), M. A. Sotos (RedIRIS)

### Abstract

This document lays down the policy to be followed by those networks (including but not limited to the GN2 participants) which wish to take part in the end-to-end Premium IP (PIP) service that the GN2 project plans to introduce. In particular the document defines what the PIP service will offer, which networks are and are not eligible to take part, and what is expected of those that do. The document also describes the basic workings of the PIP service, and it highlights and contrasts the differences between those networks that comply with the SEQUIN recommendations for PIP implementation and those which do not. An important feature of a good PIP service is the ability to demonstrate that the PIP service is being delivered as promised, and so the requirements for monitoring PIP performance are examined. Related to this is the requirement for a Service Level Agreement (SLA) between the end-user and the service provider(s), and as such the format and production of SLAs are covered.

# Table of Contents

0	Executive Summary	v
1	Introduction	1
1.1	Premium IP in GN2	1
1.2	PIP service Compatibility	2
1.2.1	Definition of a PIP flow	2
1.2.2	Definition of a PIP compliant domain	3
1.2.3	Definition of a PIP Supportive Domain	4
2	Technical parameters of PIP service	5
2.1	PIP Network Parameters Set	5
2.1.1	Capacity	5
2.1.2	MTU	6
2.1.3	IPv4 and IPv6 Source and Destination Prefixes	6
2.1.4	DSCP	6
2.1.5	One-Way-Delay	7
2.1.6	IPDV (IP Delay Variation)	8
2.1.7	Packet Loss	8
2.2	PIP Service Parameters Set	9
2.2.1	User Acceptance Response Time	9
2.2.2	Domain Acceptance Response Time	9
2.2.3	Start Time	9
2.2.4	End Time	9
2.2.5	Reservation Request Lead Time	9
2.2.6	Periodicity	10
2.2.7	User Priority	10
2.2.8	Total cost and cost for each domain	10
2.3	Policing implementation	11
2.3.1	Location of policers	11
2.3.2	Domain border safety margin	12
2.3.3	Burstiness	13
2.4	Policy for monitoring the level of service of a PIP service instance	13

3	Operational issues of PIP service	15
3.1	Overview of a PIP service reservation request process	15
3.2	Service provisioning issues	17
3.2.1	PIP users, UserIDs and User-Groups	17
3.2.2	Policy for a PIP User-Group	18
3.3	Rules for Premium IP capacity reservation	19
3.4	Service maintenance issues	21
3.4.1	Service provisioning instance accounting	21
3.4.2	Accounting of PIP service provisioning within a domain	22
3.4.3	Policy for non-compliant domains	24
3.4.4	Disqualifying an SA3-compliant domain	24
3.5	Service Level Agreements (SLAs) for the PIP service	25
3.5.1	SLA parameters	26
3.5.2	End-to-end SLA	30
3.6	Integration of policies - recommendations for NOC operating procedures	31
3.6.1	Creation of User-Groups and UserIDs	31
3.6.2	Distribution of Quota	31
3.6.3	Pre-Quota Guidelines	32
3.6.4	Monitoring of QoS parameters	32
3.6.5	SLA handling	32
3.6.6	Accounting	32
3.6.7	Troubleshooting a PIP service failure	33
4	Conclusion and Next Steps	34
5	References	35
Appendix A	An example of a PIP compliant domain: GÉANT network	36
Appendix B	Suggested 'boundary conditions' for the PIP Allocation Policy	38
Appendix C	SA3 glossary	39

## Table of Figures

<b>Figure 3.1:</b> PIP flow	16
<b>Figure 3.2:</b> PIP request/PIP response sequence	16
<b>Figure 3.3:</b> Premium IP reservation procedure	20
<b>Figure 3.4:</b> The end-to-end SLA and the per-domain SLAs involved	26

## 0 Executive Summary

This document defines the policy, the technical parameters' definitions and the provisioning rules for the deployment of the Premium IP (PIP) service over the pan-European Research Networking Infrastructure.

The aim of the PIP service is to provide guarantees on bandwidth, one-way delay (OWD), IP packet delay variation (IPDV) and packet loss percentage for IP traffic across one or more interconnected domains. It is expected that the primary beneficiaries of this service will be the users of applications with real time constraints, such as video conferencing and remote control.

The SA3 activity of GN2 will extend PIP service provisioning to NRENs and other networks by developing the provisioning mechanisms and policy for the allocation of PIP. SA3 will develop a PIP Provisioning System which, when deployed in participating domains, will provide an automated procedure for requesting and delivering the PIP service. The PIP Provisioning System will be developed and deployed in phases, with new features being added in each phase.

In order to take part in end-to-end PIP a domain must be either PIP compliant or PIP supportive. A PIP compliant domain is one which precisely meets the requirements identified by the SEQUIN project, whilst a PIP supportive domain is one which does not implement all of the SEQUIN requirements but nevertheless provides an environment where in general PIP characteristics hold true. This document extends the SEQUIN definitions such that in addition to their other criteria PIP compliant and PIP supportive domains must deploy a Provisioning System which is inter-operable with the reference Provisioning System that GN2-SA3 will develop.

The PIP service technical parameters will comprise network parameters (which is to say the quality of service offered in terms of capacity, OWD, IPDV and packet loss) and service parameters, which stipulate the duration of the PIP reservation and also how quickly a user's request will be dealt with. To prevent one or more users from monopolising the PIP service a quota system will be put in place, limiting users to an amount of PIP sufficient for their needs. The quota available to a user is just one parameter in the policy that applies to that user's user-group, others being a limit on how long a reservation may be, and how much capacity it may use.

In choosing whether or not to accept a given PIP request, a domain's PIP Provisioning System will need to assess the additional load it will place on the network, taking into account existing PIP reservations. If a user's request is accepted then the PIP Provisioning System will automatically generate the configuration necessary to allow the approved PIP flow passage through the domain. To prevent a PIP flow from exceeding its agreed

Project:	GN2
Deliverable Number:	D.S.3.9.1
Date of Issue:	17/03/05
EC Contract No.:	511082
Document Code:	GN2-05-017v7

data rate and thus overloading one or more links it is necessary to police the flow as it enters the extended PIP domain.

If the user's request for PIP is accepted then technical parameters will be encapsulated in a Service Level Agreement (SLA), which itself will comprise two automatically created sub-sections, the Administrative Level Object (ALO) and the Service Level Object (SLO). The ALO provides non-technical information such as point of contact details whilst the SLO specifies technical performance guarantees

To ensure that SA3-compliant domains honour their SLAs they are expected to deploy appropriate monitoring systems. The monitoring systems will also provide data that can be used to assess the performance of the PIP service.

# 1 Introduction

## 1.1 Premium IP in GN2

This document provides the technical parameters' definitions, policy definitions and provisioning rules for the deployment of the Premium IP (PIP) service over the pan-European Research Networking Infrastructure.

The aim of the Premium IP service is to provide guarantees on bandwidth, one-way delay (OWD), IP packet delay variation (IPDV) and percentage packet loss for IP traffic across one or more interconnected domains. It is expected that the primary beneficiaries of this service will be the users of applications with real time constraints, such as video conferencing and remote control.

The Premium IP service is only available upon request. Due to the fact that network element re-configuration is required, a basic assumption of the PIP Provisioning Policy is that there is a finite lead time between a successful request and the start of a PIP service instance or session.

Currently the Premium IP service is supported in the GÉANT (GN1) backbone. One of the goals of the SA3 activity of GN2 is to implement the provisioning mechanisms and policy for allocation of PIP in order to:

- Extend PIP service provisioning to NRENs
- Establish a procedure, eventually automated, for requesting and delivering the PIP service

The GN2 PIP Provisioning System will be deployed in a series of phases, beginning with a basic system in Phase 1 which will support a minimal set of mandatory PIP features i.e. those features which make it beneficial for serving time-sensitive applications and those that make it conform to the IETF EF PHB definition ([RFC 3246] <http://www.ietf.org/rfc/rfc3246.txt>).

Subsequent Phases of the GN2 PIP Provisioning System will support extended feature sets. A notable difference between GN2 PIP Provisioning System Phase 1 and its evolution in subsequent phases is that in Phase 1 the System will provide service guarantees (one-way delay and packet loss) only in accordance with approximate domain-wide boundaries. In later phases such guarantees will be calculated on a per-request basis with much more precision. It is worth emphasizing that providing quantitative guaranteed boundaries only

on bandwidth for Phase 1 doesn't mean that the GN2 PIP service will be an Assured Forwarding (AF) style service as it will provide improved one-way delay and packet loss performance from the start, which the AF service does not.

The phased development of the GN2 PIP Provisioning System is in keeping with the phased deployment of performance monitoring equipment across the NRENs forming the extended PIP domain, which is being done in cooperation with JRA1. These monitoring systems are necessary for the measurement of the one-way-delay and loss parameters which together form the basis for Service Level Agreements between end-users and the PIP service providers.

## 1.2 PIP service Compatibility

In order for a network domain to take part in the extended PIP domain, it must **either** be PIP compliant **or** be PIP supportive (see sections 1.2.2, 1.2.3 below).

### 1.2.1 Definition of a PIP flow

The definition of a legitimate PIP flow within a single domain is provided here:

A **PIP flow** within a domain consists of all packets marked with DSCP 46<sup>1</sup>, which is agreed to be used for distinguishing PIP traffic from the rest of the traffic at the boundaries of a domain, and at the same time

- either contain the same source IP address prefix - destination IP address prefix pair (srcAd, destAd) in their IP headers
- or originate from the same upstream Autonomous System (AS) and destined for the same downstream AS

The source of an IP flow within a domain is the ingress router for the flow in the domain. Correspondingly, the destination of an IP flow within a domain is the egress router through which all packets of the flow exit the domain.

The PIP flow is by definition unidirectional from its source to the destination. Full duplex traffic between a source and a destination is therefore comprised of two different PIP flows. A service request may be unidirectional or bidirectional. In the latter case it includes a definition for two PIP flows which may have different parameters.

---

<sup>1</sup> Although the generic Diffserv RFCs allows varying DSCP values, Premium mandates for sake of simplicity the use of only the DSCP value of 46. This conforms to [RFC3246] which RECOMMENDS the use of codepoint 101110.

## 1.2.2 Definition of a PIP compliant domain

For the purposes of the GN2 project a network domain is considered to be PIP compliant when it adheres to the Premium IP specification of [SEQUIN D2.1 addendum 1] **and** it deploys a Provisioning System which complies with the requirements of [GN2-04-153]. The main criteria for compliance are:

- Classification of PIP packets. It must support the classification of PIP packets arriving from each upstream domain based on the DSCP value of 46 used to mark PIP traffic.
- Re-marking of traffic at the ingress. It must support checking each packet against a list of source/destination IP address pairs in its ingress routers if they border a non-compliant domain. Non-authorized flows must be re-marked to a DSCP of 0 (zero).
- Policing to control exceeding traffic. In the case of the origin domain for an authorized PIP flow, the domain supports policing with a dedicated token bucket profile the packets of the flow, dropping any out of profile packets. As with the re-marking described above, this policing is done on a per flow basis, using source/destination IP address pair filters. In the case of a transit domain, the domain may support policing of the total PIP traffic amount coming from a given upstream AS and destined for a given downstream AS. This is the recommended method since it means that a failure in a specific PIP flow will not have an impact on other compliant traffic. However, if a router is not capable of this kind of advanced AS-to-AS policing then it may either police on a per PIP flow basis (as is done in an origin domain) or it may police the ingress interface as a whole (for example, 30% of the interface line-rate). A transit domain should re-mark to DSCP 0 all out of profile packets.
- Admission control on PIP service requests. It supports the functionality of conducting PIP service requests' admission control (via a PIP Provisioning System) by providing a positive or negative answer to a request for serving a PIP flow from an ingress interface to an egress interface of the domain.
- Well-known QoS domain profile. It provides minimal intra-domain end-to-end OWD and IPDV as well as zero or negligible packet loss for all packets of each PIP flow that it has accepted to serve. In Phase 1, the minimum requirement of a PIP compliant domain is to provide a domain-wide worst case set of these three parameters that will always be an upper bound for each PIP packet served through the domain. Provisional values are OWD < 100 ms, Jitter < 10 ms and Packet loss < 0.5%. In subsequent Phases each domain should be able to define tighter bounds on a per-flow or on an edge-to-edge basis. In this case, the PIP compliant domain should provide statistical data on QoS metrics (OWD, IPDV and packet loss) for PIP traffic on an edge-to-edge basis.

It is assumed that a PIP flow will always take the route expected of it, which is to say it is assumed that the network's state is stable for the duration of the PIP flow, and known at the time of making the reservation. Note that, should a link fail before or during the PIP flow then although its path will no longer be that assumed by the Provisioning System the flow will nevertheless still survive because the network control plane will ensure it is properly re-routed, albeit possibly without the required QoS guarantees.

A PIP compliant domain can impose a set of rules and restrictions for serving PIP flows, accepting requests for PIP flows with a sum of average rates (denoted by the corresponding token bucket policers) not exceeding a small percentage (e.g. 10%) of the capacity of each one of its ingress interfaces. These rules have to be specified in a formal way in the policy for PIP service of the domain and be used in the request admission control for each domain.

### 1.2.3 Definition of a PIP Supportive Domain

For the purposes of the GN2 project a network is considered PIP supportive if, as per [SEQUIN D4.2], “[it] preserves Premium IP DSCP value and offers an environment where in general Premium IP characteristics hold (e.g. by over provisioning)” **and** it deploys a Provisioning System which complies with the requirements of [GN2-04-153]. Note therefore that a domain where there is congestion (shown by heavy packet loss or excessive delay) cannot be considered a supportive domain. In time, performance monitoring tools should be able to demonstrate whether or not a self-declared PIP supportive domain is truly providing a low congestion environment and therefore entitled to its status. It should be noted that a PIP supportive domain will not necessarily be able to mark to DSCP 46 or/and police incoming PIP flows and as such PIP traffic originating from them will need to be policed further downstream, by a PIP compliant domain.

## 2 Technical parameters of PIP service

The technical parameters of the PIP service can be divided in two classes:

- **Network Parameters Set.** This is a set of parameters related to the actual PIP packets and their treatment in the network. These parameters are used to determine if the requested PIP flow is technically possible to serve and also provide the information necessary to configure network equipment.
- **Service Parameters Set.** This is a set of parameters related to the users' service requests

When requesting a PIP service a user must specify the required Capacity, the Start and End Times of the reservation, and at least one IP Source and Destination address prefix Pair. For authorisation and accounting purposes the request must also include the unique identifier of the user on the PIP Provisioning System (UserID) as well as the organization/group for which the user is making the reservation (User-Group). All other parameters (e.g. required OWD) are optional and if they are not supplied will be set to default values.

An assumption for the implementation of Phase 1 of the PIP Provisioning System is that neighbouring domains are single-homed with one another, which is to say there is only one active link between them. This means that there cannot be any load-balancing between domains.

### 2.1 PIP Network Parameters Set

#### 2.1.1 Capacity

A PIP flow's Capacity is the amount of bandwidth (in bits per second) that a user wishes to reserve for the exclusive use of that PIP flow. The Capacity value includes the IP header. In general the specification of capacity requirements is specified by the following sub-parameters:

Project:	GN2
Deliverable Number:	D.S.3.9.1
Date of Issue:	17/03/05
EC Contract No.:	511082
Document Code:	GN2-05-017v7

- minimum assured Capacity or Committed Information Rate (CIR)
- Averaging Interval (AI) over which CIR is evaluated by the corresponding policer
- Burst Size

The Capacity value is a mandatory parameter and determines the resource allocation on network nodes along the path. The Burst Size is an optional parameter that the end user may specify. Burst Size is specified as a percentage of the maximum capacity, based on the individual policy and limitations imposed by the origin domain. For the internal implementation of the PIP provisioning system, it is recommended that its value should increase in the transit domains as a function of the distance from the source, and in any case as a growing function of aggregated PIP capacity. The Averaging Interval is an internal parameter of the system, and cannot be specified by the end user.

## 2.1.2 MTU

The Maximum Transmission Unit is the largest size that a layer 2 frame may reach on a single link. The MTU limits the amounts of user data in each frame. A larger MTU size is advantageous for large file transfers and may be requested for the end-to-end path. The end-to-end MTU is the smallest supported MTU along the PIP path. MTU is an optional parameter for end-users and it will not be supported in the early phases of the PIP Provisioning System.

## 2.1.3 IPv4 and IPv6 Source and Destination Prefixes

The user must declare one IP address prefix (host or network) as the source address, and one IP address prefix (host or network) as the destination address (host or network). Further addresses may be declared to be additional source and/or destination addresses, but such addresses will not be used as part of the path-finding calculation – it is the user's responsibility to ensure that the additional traffic follows the same route in and out of the extended PIP domain as the primary source/destination pair. Furthermore, the intra-domain PIP Provisioning System of each PIP compliant origin domain must ensure that when multiple source and/or destination addresses are requested in a single request, the corresponding policer is applied on the total of traffic produced from the multiple sources defined.

## 2.1.4 DSCP

The recommended DSCP value for PIP flows is 46 (EF PHB 101110). In case DSCP handling is not available in the used hardware or technology, the following mapping to Type of Service (ToS), MPLS QoS bit and VLAN QoS bits are recommended.

- DSCP to ToS mapping

Project:	GN2
Deliverable Number:	D.S.3.9.1
Date of Issue:	17/03/05
EC Contract No.:	511082
Document Code:	GN2-05-017v7

According to the definition of Differentiated Services, the DS field supersedes the existing definitions of the IPv4 Type of Service (ToS) octet and the IPv6 Traffic Class octet. Six bits of the DS field are used as the DSCP to select the Per Hop Behaviour (PHB) at each interface.

- DSCP to MPLS EXP mapping

MPLS packets have a 3 EXP bits field. DSCP is a 6-bits field so it is possible to have only eight possible EXP values and 64 possible DSCP ones.

- DSCP to IEEE 802.1p mapping

The IEEE 802.1p is an extension of the IEEE 802.1Q (VLANs tagging) standard and its specification enables Layer 2 switches to prioritise traffic. The prioritisation specification works at MAC framing layer. The 802.1p header includes a three-bit field for prioritisation, which allows packets to be grouped into various traffic classes.

Suggested values for these cases are:

Tagging field used			
ToS byte marking (3 bit IP precedence field)	1	0	1
MPLS EXP field marking	1	0	1
IEEE 802.1p Tagged Frame for Ethernet/TCI bytes/ User Priority field	1	0	1

**Table 1: PIP compliant marking values**

## 2.1.5 One-Way-Delay

Delay is the time between the transmission of a packet at its source and the moment it is fully received by the destination. The total delay can be divided into a few main components:

- Propagation delay: the time taken by the first bit to travel from the source to the destination, which is a function of the signal propagation on the medium.
- Forwarding delay, which is introduced by each network element that a packet crosses. Each network element introduces its own forwarding delay, and the total figure depends upon the number of the hops along the path. In a high quality network, such as an NREN backbone or the GEANT2 network, forwarding delay should be negligible compared to Propagation delay.
- Queuing delay, which is caused by instantaneous load and thus build up of packets in network element queue. The goal of PIP is to minimise overall delay by minimising queuing delay, and there should never be more than a small number of PIP packets queued.

- Transmission delay, which is the time needed to transmit all the bits of a packet on each link. Transmission delay is normally small compared to propagation time. For example, the transmission delay for a 1500 byte frame on a 1 Gigabit Ethernet interface is 12 microseconds, whilst the propagation delay on a 500km link can never be less than 160 microseconds, (and will normally be at least twice that).

A user does not request a specific value for the OWD parameter in a PIP reservation request, Instead, the user can request for an upper bound on the total OWD across the extended PIP domain experienced by the PIP packets. This upper bound is compared against monitored OWD statistics in each domain involved and the PIP request will only be accepted if the sum of the observed OWD values along the extended PIP domain does not exceed the user-requested upper bound.

### 2.1.6 IPDV (IP Delay Variation)

IPDV, sometimes called jitter, is the difference in the one-way-delay of two successive packets in the same IP flow. IPDV is mainly a function of any queuing a flow experiences as it traverses the end-to-end path, and is significantly affected by any packet reordering.

IPDV is an optional parameter in a PIP request. As with OWD, a user will not request a specific value for IPDV but rather an upper bound on the total IPDV encountered by the packets of the PIP flow along the extended PIP domain. This upper bound is compared against monitored IPDV statistics in each domain involved and the PIP request can only be accepted if the combined probability of the observed values along the extended PIP domain does not exceed the user-requested upper bound. A PIP Provisioning System-internal parameter for keeping monitoring statistics on IPDV per domain is the averaging interval during which IPDV measurements are processed and statistics are produced.

Forecasting IPDV is very complex and as such this parameter will not be supported in the early Phases of the PIP Provisioning System. However, each PIP compliant domain must ensure that no PIP flows' packet reordering occurs and thus minimise IPDV.

### 2.1.7 Packet Loss

PIP Packet loss is defined as the percentage of valid PIP packets discarded by the network. Packet loss may or may not be related to congestion in the network. Non-congestion related packet loss is mainly a function of the quality of the underlying network infrastructure, such as transmission lines and other equipment, and also routing stability. The PIP service has no real control over non congestion related packet loss, but it must ensure there is no congestion-related packet loss.

## 2.2 PIP Service Parameters Set

Service parameters are those non traffic related parameters which have to be specified in order to fully define the user request and its implementation in the network.

### 2.2.1 User Acceptance Response Time

User Acceptance Response Time is the maximum time allowed for the inter-domain Provisioning System to respond to a user's PIP reservation request. The User Acceptance Response Time is a function of the single-domain acceptance response times and the number of domains involved. This metric will be visible to the end user.

### 2.2.2 Domain Acceptance Response Time

Domain Acceptance Response Time is the maximum time allowed for an intra-domain Provisioning System to respond to a request for a PIP flow to transit its domain. The concatenation of Domain Acceptance Response Times will give the User Acceptance Response Time (above) but the individual metrics are transparent to the end user.

### 2.2.3 Start Time

The requested Start Time is the time by which the user requires the PIP service to be in place. In all likelihood the PIP service will actually be put in place before the requested Start Time but this is transparent to the end-user, except in as much as if they started their PIP flow before the requested Start Time their flow would be serviced as PIP rather than Best Effort.

### 2.2.4 End Time

The requested End Time is the time at which a user no longer requires the PIP service. It is probable that the service will not be de-configured until some time after the requested End Time. Once the PIP configuration has been removed the corresponding allocated network resources are released.

### 2.2.5 Reservation Request Lead Time

Reservation Request Lead Time is the minimum term period between the time of a user submitting their PIP reservation request and the requested Start Time. In the early Phases of PIP Provisioning System implementation, the Reservation Request Lead Time will be dependent upon two factors – the time it actually

takes to configure the network elements and the time(s) of day at which element configuration is allowed (which may be domain dependent). During Phase 1 of the Provisioning System all domains will be expected to observe a maximum lead time of 2 working days. This metric is an external one for the system and will be visible to the end user. It should be bounded so as to anticipate for the internal configuration lead time restrictions in the domains involved.

## 2.2.6 Periodicity

For some applications (like videoconference) it may be advantageous to schedule PIP flows periodically on a daily, weekly, or monthly basis. This may be a feature which is included in future versions of the Provisioning System.

## 2.2.7 User Priority

For Phase 1 of the PIP Provisioning System, no user priorities will be defined. Each request will be served on a first-come-first-served basis.

## 2.2.8 Total path-cost and path-cost for each domain

In order to ensure fair access to the PIP service, and prevent any one group monopolising it, each PIP request will have an associated path-cost ('cost' for short). This cost is not a financial one but will nevertheless be attributed to a certain User-Group, so that if necessary their use of the service can be moderated. As such the cost of a PIP service is a parameter to be considered when a user makes a request for allocation of PIP. This parameter is a compound value and may be dependent on one or more of the following parameters:

- Capacity requested
- Duration
- Current Priority assigned to the PIP flow (for future Phases of the PIP Provisioning System implementation)

In Phase 1 of the PIP service, as stated above, cost will not be accounted for. In Phase 2 the cost for each reservation in each domain will be directly proportional to the product of the requested PIP capacity multiplied by the duration of the reservation.

## 2.3 Policing implementation

At ingress to the PIP origin domain (which is to say at the ingress to the extended PIP domain) PIP packets are classified based on their source and destination IP address prefixes, and policed based on the contracted Capacity.

### 2.3.1 Location of policers

The purpose of policing, both in the origin domain and the transit domains, is to ensure that PIP flows do not exceed their pre-agreed flow rates, so that all PIP flows are forwarded without queuing and Best Effort traffic is not starved of network resources.

Each PIP flow must be policed as close as possible to the source of the traffic flow, which means policing must be done at the ingress to the first PIP compliant domain in the end-to-end path. This should ensure that any packets that need to be discarded (because they are out of profile) are discarded as early as possible and not use up time and space in the network before being discarded at a late stage in its journey. Policing need only be performed on ingress traffic.

Once the Premium IP traffic reaches the core of each domain, it is recommended that the principles of DiffServ are relied upon and no more policing is performed.

#### 2.3.1.1 Policing at ingress interface of origin domain

Policing is implemented at the ingress interface of the first router that a PIP flow enters within the origin domain. It should guarantee that the PIP flow does not exceed its allocated PIP capacity. If the origin domain is not PIP compliant then policing of the PIP flow must take place on the first PIP compliant domain that the packets of the flow will traverse along the end-to-end path. In combination with the PIP Provisioning System request processing module, the PIP flow policers should guarantee that the total amount of PIP traffic coming from several different users does not exceed the maximum amount of PIP traffic that can flow across each backbone link within the domain. Traffic exceeding the PIP capacity should be dropped.

#### 2.3.1.2 Policing at transit domain borders

For neighbouring PIP compliant domains policing should be implemented at the ingress of the downstream domain, which is to say the point at which the PIP flow enters the second domain, having left the first. If there is both a primary and a back up link between the domains then the filters and policers should be assigned to both interfaces, so that if the primary circuit fails PIP is properly handled by the back-up interface. Note that if the back up link is of a lower capacity than the primary link then a larger than usual amount of PIP traffic could use the link. Therefore it may be necessary to set the acceptable level of PIP below that which would normally apply to a circuit of the primary link's capacity, so as not overwhelm the back-up link in the event of traffic shifting to it.

The policing required at a transit domain border will depend on the nature of the two domains concerned.

Upstream domain	Downstream domain	Required policing
PIP compliant	PIP compliant	If possible policing should be done per upstream-AS/downstream-AS aggregate. However this type of policing is quite sophisticated and may not be supported by all equipment, in which case it is acceptable to police on a per PIP-flow basis, or even on a total incoming PIP aggregate basis.
PIP supportive	PIP supportive	PIP supportive domains are not required or expected to inspect or police packets based on DSCP values are involved.
PIP compliant	PIP supportive	Since policing should only be done at the ingress interface of the downstream domain, then this situation is effectively the same as above- no policing is expected from PIP supportive domains.
PIP supportive	PIP compliant	The PIP compliant domain should police <b>all</b> individual PIP flows coming from the PIP supportive domain, regardless of whether or not the traffic has been previously policed by another PIP compliant domain. This is necessary in order to ensure that no unauthorised PIP flows from the PIP supportive domain (which is not required to inspect or police DSCP values) are propagated further through the extended PIP domain.

**Table 2.1: Domain border policing requirements**

In a transit domain, traffic exceeding the PIP aggregate capacity must be remarked and treated as best-effort traffic, and not dropped.

### 2.3.2 Domain border safety margin

A policing safety margin of a small percentage (e.g. 20%) of the sum of all PIP incoming capacities should be kept when applying a rate limitation on transit domain borders.

### 2.3.3 Burstiness

It must be emphasised that minimum allowed burst size is a function of contracted capacity. It is an amount of data that can be transmitted in a given amount of time, given that there is room enough on the link to transmit it. Setting up a low ratio between Burst/Capacity can result in severe service degradation especially in transit domains, where contracted bandwidth increases, causing high packet loss and thus inability of the PIP flows in utilising all the contracted capacity. As long as PIP bandwidth grows, bigger bursts must be accommodated, using the following common rule:

$$\text{Burst size(bytes)} = \frac{\text{Contracted Capacity(bps)} * 1.2}{8}$$

Some vendor's implementations are able to accommodate also an extended burst size (to be considered as an optional parameter), that is intended as an instantaneous value that is added to normal burst, and is defined by the following formula:

$$\text{Extended burst size(bytes)} = \text{Burst Size(bytes)} * 2$$

Therefore the complete parameter set for policing is:

$$\text{Contracted PIP Capacity(bps)}, \text{ Burst Size(bytes)}, [\text{Extended Burst Size(bytes)}]$$

## 2.4 Policy for monitoring the level of service of a PIP service instance

To fully assess the PIP service provided to a PIP flow the following should be monitored:

- Bytes In/Out for the PIP flow
- Marked Packets In/Out for the PIP flow
- Remarked packets (46 to 0, faulty marking) for the flow
- Loss of packets for the PIP flow
- Packets of PIP flows remarked to BE packets due to them exceeding the corresponding policing profile in the ingress of a transit domain
- Traceroute
- OWD of PIP packets, ideally as a function of packet size

- IPDV of PIP packets
- The probability of a request blocking due to lack of available resources on the domain along the requested path.

Each of the above parameters must be monitored on an edge-to-edge basis by each PIP compliant domain. Statistical data must be available per ingress-egress interface pair and used in order to assess the service quality experienced by each PIP flow between these edges. PIP flow is unidirectional by definition; therefore for each PIP flow it is appropriate to consider only the inbound PIP monitoring parameters.

The above parameters can also be measured for the total of PIP traffic served through a specific router interface, in order to provide an assessment of the successful implementation of PIP as a whole.

The counters Byte In/Out are standard parameters that are commonly used for understanding the 'healthy status' of the network (PIP traffic+ Best effort traffic) at a specific point of the network. Monitoring the parameters allows verification that the real percentage of PIP flows over the total traffic does not exceed the predetermined threshold.

The 'Marked Packets In' parameter allows for evaluating the total incoming PIP traffic entering an interface. It is the most important monitoring parameter.

The 'Remarked Packets In' parameter, for the case of a transit domain, is used to monitor to what extent PIP flows violate their policers at the ingress interface of a transit PIP domain. This parameter is not relevant in the core of each PIP compliant domain.

The 'Dropped packets rate' is monitored at the ingress interface of the origin domain of a PIP flow.

As previously stated in the description of a PIP flow, it is assumed that a PIP flow will follow the route expected of it for the duration of the reservation i.e. it is assumed the network is stable. Nevertheless, in reality a link failure could cause a change in the network's topology and thus the path of the PIP flow. A mechanism to update the monitoring process in case of topology changes should be considered. Traceroute may be evaluated as diagnostic tool that allows determining the path of the packets of a flow, if executed in regular intervals during the reservation period of a PIP flow.

The role of a monitoring will evolve alongside the GN2 PIP Service Provisioning system.

In Phase 1, the number of measurement points for delay and jitter in NRENs will be minimal and monitoring will be limited to a process of inspection on a case-to case basis to verify or troubleshoot a configured service instance. In other words, during the Phase 1, the monitoring systems, where available, will be used on a reactive basis, manually, and not to provide to users quantitative guarantees on the OWD and IPDV boundaries. In future Phases it is expected an interface between the monitoring system and the PIP Provisioning System will be developed. Over time, the number of measurement points should be extended to include all ingress-egress pair domain interfaces and measurements will be conducted regularly e.g. on a day-to day basis.

## 3 Operational issues of PIP service

### 3.1 Overview of a PIP service reservation request process

The final PIP Provisioning System will provide a sophisticated interface for end-users applying for a PIP reservation. If the user's request can be met the user will be sent detailed confirmation, informing the user of the cost entailed (e.g. the reduction of his quota), the contact points for the PIP service provisioning instance etc. Similarly, if the user's request can not be met, they will be notified of the reasons for this and given sufficient information about the available resources so that if a viable alternative to their original request exists they may request that instead. In the first Phases of the PIP Provisioning System, only a limited subset of this functionality will be available.

In the initial Phases of the PIP Provisioning System implementation, each user will only be able to access the PIP intra-domain Provisioning System of the domain in which they are registered (i.e. their 'home' domain) with a UserID. Each user will be allowed to make a PIP request where the origin domain for the PIP flow is their home domain. In later Phases, roaming of user profiles and quotas will be supported, so that a user will be able to make a PIP request through the PIP intra-domain Provisioning System of a domain other than their home domain and/or where the origin domain for the PIP flow is other than their home domain.

A PIP reservation request will include:

- mandatory parameters (srcIPAd, destIPAd, Capacity, StartTime, EndTime, UserID, User-Group)
- optional parameters (Bi-Directional, IPDV, max OWD, User Priority Request, Periodicity)

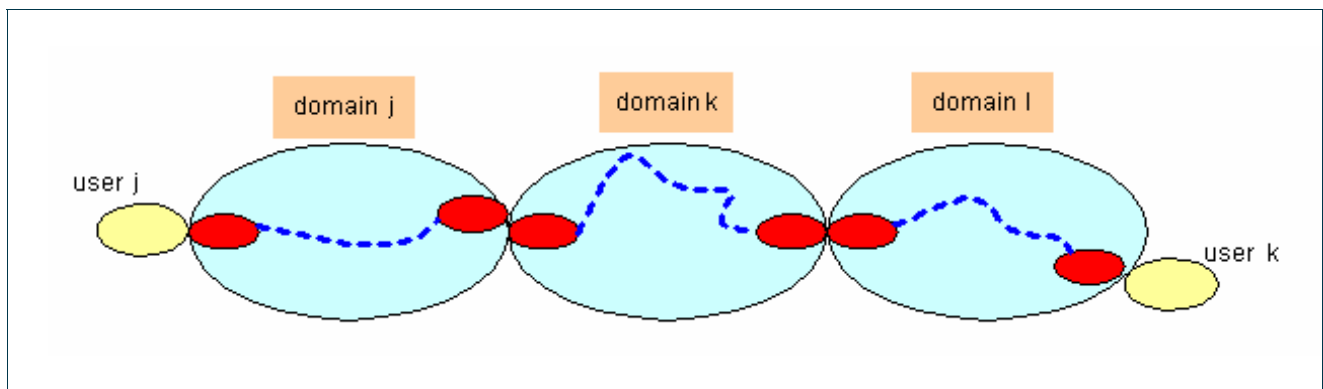
Each domain through the path from source to destination will evaluate the request against the resources available by checking:

- PIP traffic parameters along the path
- PIP service parameters along the path
- PIP flows already in place along the path

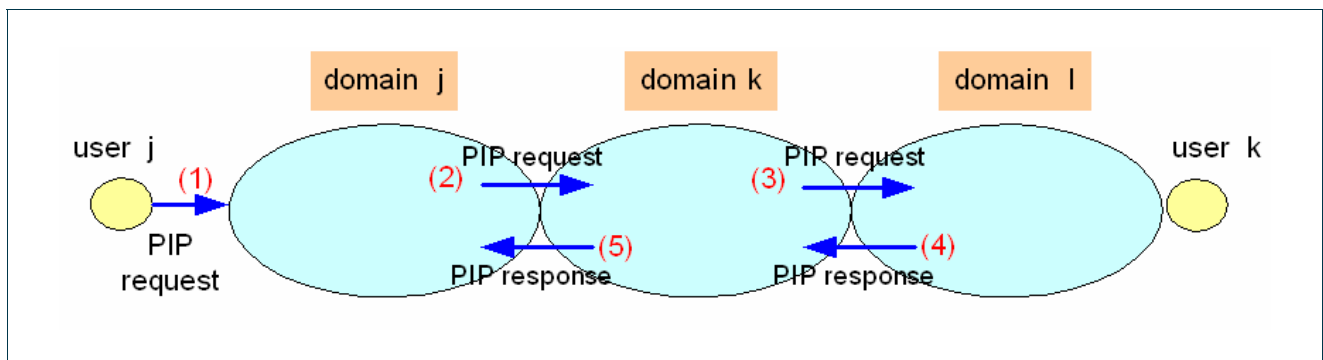
- Concurrent accepted PIP requests across the path
- Network information: internal and external routing protocols used by the domain, the current parameters of these protocols (e.g. link costs, BGP attribute values), and the current available PIP bandwidth of each link of the domain.

Based on this information the reservation system determines the path along which each PIP flow will be served. The PIP Provisioning System must be reliably and promptly updated with information on any changes to the network. Realistically this should be an automated process in order to minimise the possibility of error and such an advanced feature which will not be possible in the early Phases of the Provisioning System.

Each domain, in turn, either accepts or refuses the PIP request. The exchange of provisioning information between domains is executed in daisy chain fashion, as shown below. Figure 3.1 shows the actual path taken by the PIP flow whilst Figure 3.2 shows the sequence and direction of the requests and responses. Note that each domain communicates only with its neighbour – there is no system acting as a central manager, rather the domains operate as a federation of equal peers.



**Figure 3.1:** PIP flow in several domains



**Figure 3.2:** PIP request/PIP response sequence in two domains

Project:	GN2
Deliverable Number:	D.S.3.9.1
Date of Issue:	17/03/05
EC Contract No.:	511082
Document Code:	GN2-05-017v7

In the final stage of the request/response chain the origin domain is informed of network resource availability up to and including the destination domain. The user is then informed as to whether their request has been accepted or rejected, and in the case of a rejection the user is invited to submit a suitably modified request.

If the request is accepted then the Provisioning Systems in each domain will generate the appropriate configuration amendments that are required to support the request. These changes will apply to the ingress router of each domain and there may also need to be changes made to other devices e.g. firewalls.

As previously stated, in the early Phases of the PIP service all equipment configuration changes will be performed manually by NOC staff. In future Phases it is expected that the Provisioning System will be able to apply and remove the necessary equipment configuration changes itself.

## 3.2 Service provisioning issues

### 3.2.1 PIP users, UserIDs and User-Groups

A PIP user is an individual person or software application which is allowed to make PIP requests. A user authenticates themselves to the PIP Provisioning System (perhaps via a GN2 wide single sign-on system) based on their (individual) UserID, and then has their request evaluated based on the User-Group specified as part of the request. In other words policy and quota are applied to a User-Group, not a UserID, and as such, a user must be a member of at least one User-Group (note that in Phase 1 of the Provisioning System there will be no User-Group specific policy applied, and therefore all users will belong to some system wide default group). When a user submits a request their UserID is checked to see whether it is a legitimate member of the User-Group stated in the request. Depending on future GN2 developments, this check might be done by the Provisioning System or it might be done by the GN2-wide AA system.

User-Groups may be sub-grouped. Each User-Group's name is hierarchical in nature, very much like an Internet domain name. The root of a User-Group's name will be related either to the home domain (e.g. NREN or campus network) via which they access the extended PIP domain or, in the case of a large, quasi-independent international organisation such as EGEE, some other unique identifier. When a request is propagated through the extended PIP domain, each domain will choose at which level in the User-Group hierarchy it will apply policy. Typically, in the User-Group's home network (where their traffic enters the extended PIP domain) the most detailed accounting will be done – in subsequent networks the policy check (and quota reduction) may only be done at the top level.

The above statement is best illustrated by example. Consider User-Group Alpha which represents a department ('Da') in a university ('Ua') connected to NREN A. User-Group Alpha has a fully qualified group name which should be something like 'Na.Ua.Da'. NREN A (a PIP compliant domain) should have an entry in its intra-domain PIP Provisioning System which would show a detailed policy for User-Group Alpha, including max capacity per reservation allowed, max duration, etc, and of course quota. However, another NREN, NREN B, will probably not want to keep detailed accounts and policies on all departments in all universities in country A. Instead, it may have a policy which applies to all 'Na' subgroups, which would include one large quota out of

which all 'Na' subgroups' reservations are deducted. Alternatively NREN B may decide to apply a specific policy for, say, University A. In this case NREN B would maintain a separate policy record for 'Na.Ua.', and apply this to all User-Groups of the form 'Na.Ua.\*'. It would still have a policy for 'Na' subgroups in general but the 'longest match' approach taken to applying policy means this would not be applied to University A subgroups. An example of where this might happen is if NREN B observes that a specific User-Group is misusing PIP. In this case NREN B may apply a very limited policy just to that User-Group, or even block it altogether. Furthermore, because all PIP requests will also include UserID, in exceptional circumstances NRENS could even blacklist specific UserIDs if they so wished.

Top-level User-Groups (NREN-specific, or large international projects) are agreed amongst GN2 SA3 participants and their details manually added to each intra-domain PIP Provisioning System. Similarly, as part of configuring a new intra-domain PIP Provisioning System the top-level user groups will be manually configured. Each Top Level User-Group has a home domain PIP Provisioning System, which is the only System authorised to create sub-groups for that top-level User-Group. When a sub-group is created the home domain floods a message announcing the sub-group's existence. The other domains then have the choice of:

- Adding a new entry for that specific sub-group
- Amending the policy for the sub-group's parent (or grandparent) group e.g. increasing that group's quota parameters
- Do nothing

If subsequently any domain makes a change to its own policy concerning a top-level group or sub-group, then that message is flooded as well.

Typically an NREN will be the home domain of all User-Groups which connect to only one NREN (Universities, national projects etc) and the GEANT2 Domain Provisioning System will be the home domain for large international projects which have multiple ingress points to the extended PIP domain.

Suggested Top Level User-Groups to be implemented in all domains involved in the extended PIP domain of Phase 1 are:

- EGEE project
- NREN users

### 3.2.2 Policy for a PIP User-Group

The Policy for each User-Group defined in a PIP compliant domain must include the following parameters:

- Min allowed reservable capacity per request

- Max allowed reservable capacity per request
- Min allowed duration of PIP reservation per request
- Max allowed duration of PIP reservation per request
- Maximum amount of Quota (Q) that may be banked
- Replenishment rate (R) of the Q quantity, in terms of amount per unit of time. Typically, the unit of R will be in megabytes or gigabytes per day, but the replenishment frequency could equally be once a week, once a month etc.

The required Capacity for a PIP flow is related to the user application. A common low speed application is videoconferencing. One videoconferencing channel typically requires 128 kbps and therefore 100kbps is a suitable minimum value allowed for a PIP request.

A user can request any amount of PIP capacity above 128 Kbps, as long as the capacity-duration product of their request does not exceed their available quota for PIP service usage.

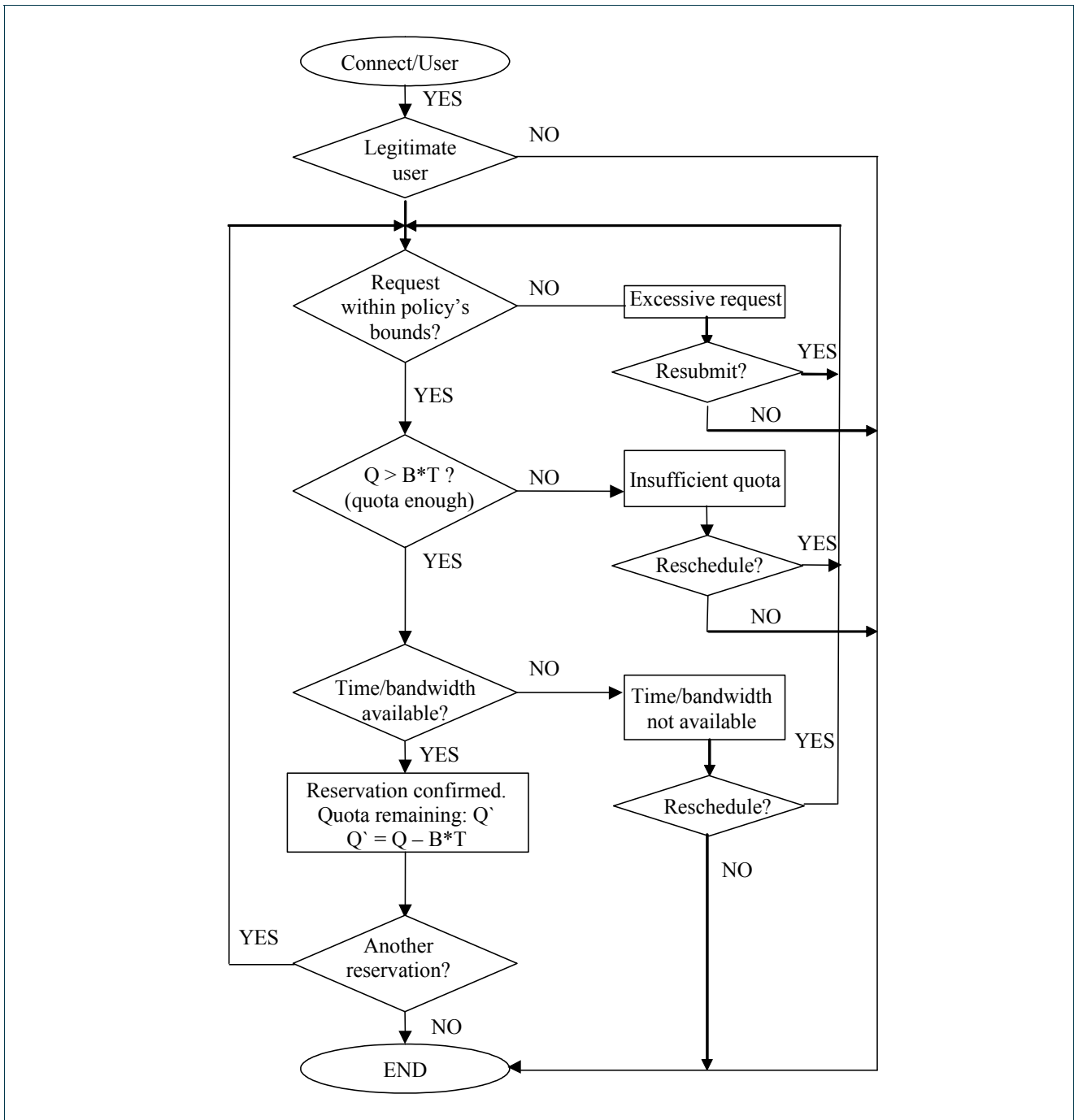
The maximum amount of PIP capacity that may be reserved will be a parameter defined on a per-domain basis. For the case of GEANT it is expected to be 100Mbps.

### 3.3 Rules for Premium IP capacity reservation

The PIP reservation agreement contains information about user rights, obligations, reservation conditions and consequences of improper usage. A user will have to sign this agreement when their account is first created.

All legitimate users are permitted to make a reservation for Premium IP capacity. If the User-Group they represent has available quota and if there are sufficient resources (PIP capacity), the reservation will be accepted and confirmed. In all other cases the request will be rejected.

The reservation procedure is shown in **Figure 3.3**. It starts with a user connecting to a reservation system to schedule a reservation.



**Figure 3.3:** Premium IP reservation procedure

The user only needs to be authenticated once, in their home domain (in later Phases of the PIP provisioning system roaming access to the service will be possible). As the request for PIP service is propagated, subsequent domains will check the eligibility of the request, by checking the policy (and quota) against the user's User-Group.

A user may query the Provisioning System to see if the required resources are available for the desired period and in the desired amount. In this way, the reservation process is made easier and faster since no rescheduling loop is necessary.

Note that in Phase 1 quota will not be used, which in effect means that everyone has unlimited quota. NRENs should monitor PIP requests to ensure that no one User-Group monopolises the PIP service at the expense of others.

Users are offered the following guidelines:

- When submitting a PIP reservation request there is a Minimum Reservation Request Lead Time. For Phase 1 of the PIP Provisioning System over GÉANT, the Minimum Reservation Request Lead Time will be two days.
- A user should reserve the resources only for the period required. User are discouraged from making speculative bookings, but it is acceptable to plan for an extended reservation so as to allow for unforeseen delays (e.g. video-conference meetings over-running).
- Premium IP capacity usage may be monitored. If users are found to be reserving capacity and not using it then that User-Group's policy may be made more restrictive and in extreme cases users and/or a user-group may be suspended.
- The minimum allowed duration of a reservation ( $T_{\min}$ ) will be determined by the shorter of that stated in the User-Group policy and the time it takes to configure and re-configure a network element. In Phase 1 (before specific User-Group policies are in effect)  $T_{\min}$  will be 1 day.
- The maximum duration of a reservation ( $T_{\max}$ ) will be determined by the User-Group policy, but in Phase 1 (before specific User-Group policies are in effect) the maximum time will be three months.
- Requested reservation Start and End times will be in the form of date-hour-minute. Note that since network elements are normally only configured during specific maintenance windows (typically once a day) the actual applied reservation will be slightly longer than that requested, running from the maintenance window before the requested PIP service, to the maintenance window after the requested PIP service (when the network elements will be de-configured).

## 3.4 Service maintenance issues

### 3.4.1 Service provisioning instance accounting

Associated with each User-Group is a quota, (Premium IP quota, or PIP quota), measured in bits (b). When a user makes a reservation then the Capacity-Duration product of the reservation is subtracted from the quota of

the User-Group that the user represents. So for example, if user U, representing User-Group G, makes a PIP reservation of Capacity 10Mbps lasting 1 week, then User-Group G's quota would be reduced by 756GB (1 week multiplied by 10Mbps). If there is insufficient quota to support a new reservation request then the request will be refused i.e. it is not possible for a user-group to have negative quota. In these circumstances the user may contact their NREN to discuss their PIP needs and apply for an increase in the user-group's quota.

Quota is periodically (period  $P$ ) replenished, at a rate  $R$  determined by the User-Group's policy (see section 3.2.2). Also specified in the policy is some value  $Q_{max}$  which represents the maximum value that that User-Group's quota can grow to.

Each domain, not just the origin domain, should calculate and deduct quota from User-Group's whose PIP flows transit their networks. In the case of transit domains it is recommended that rather than keeping separate User-Group policies for foreign User-Groups they instead have a single, aggregated policy for all User-Groups homed in a given domain i.e. there should be a single policy per Top Level User-Group (see section 3.2.1).

It is therefore up to each domain to determine its policy for setting quota and they may even choose to reduce a User-Group's allocated quota if they consistently overbook PIP (e.g. does not use reserved PIP capacity or uses significantly less than that requested). Quota amount can be re-assessed periodically and/or upon request according to the internal policy of each PIP compliant domain.

For GEANT2 implementation, suggested default values are:

- quota refresh period  $P$  is one day,
- replenishment rate  $R$  is 20 GB
- maximum quota is 6000 GB

The values would allow a user to make a permanent reservation of approximately 2Mbps (but note that if a user were to make such a reservation they would be expected make proper use of it)

### 3.4.2 Accounting of PIP service provisioning within a domain

For accounting, what needs to be measured are the resources reserved by each User-Group, and the resources they actually use.

In this particular case that resource is the Premium IP traffic. The accounting of the used IP Premium bandwidth should be done by the reservation system, based on the data it collects from the appropriate monitoring systems. The system will keep records on the usage of the resources and the requests made by users.

The data collected by the Accounting system can be used for a range of purposes, such as assessing the appropriateness of User-Group policies, determining service availability and promoting the PIP service. In

particular there will be the necessary data to study the behaviour, performance and operation of the service and so determine what adjustment, if any, needs to be made to the parameters of the service and the PIP Provisioning System.

In particular, the proposed records for general Accounting are:

- Total Number of requests
- Total Number of requests per period of PIP service operation
- Total Number of serviced requests
- Total Number of serviced requests per period of PIP service operation
- Total Number of denied requests
- Total Number of denied requests per period of PIP service operation
- Total amount of reserved PIP bandwidth
  - per period
  - per domain
- Total number of users

Another possible use for the accounting data could be the production of a Premium IP Network Accounting Record (PIP NAR). The fields of this record would be:

- Record Type
- Measurement Point Identification
- Flow Description
- Reserved Resources
- Used Resources
- Data Extension

The PIP NAR for a denied request would use only 2 fields:

- Record Type: Requests
- Data Extension: Explanation indicating why the requests are denied.

For approved and serviced requests all the fields of the PIP NAR would be completed.

The accounting scheme will provide for each user the percentage of utilization of their available quota in each service operation period. For each domain it will provide the utilization of the set-aside PIP capacity on core links in each service operation period.

### 3.4.3 Policy for non-compliant domains

Non-compliant domains can each be classified in to one of three categories as described in [SEQUIN D4.2]:

- incompatible: the domain resets or discards packets with Premium IP DSCP values
- indifferent: the domain preserves Premium IP DSCP value but applies only Best Effort treatment
- supportive: the domain preserves Premium IP DSCP value and offers an environment where in general Premium IP characteristics hold, (e.g. by over provisioning) (see also section 1.2.3 )

A PIP allocation through an incompatible domain is not possible: even if the network is of sufficient capacity and uncongested the DSCP value reset cancels all PIP mechanism established on other domains. The worst case for users is when the domain is at the beginning of the path through an extended PIP domain: DSCP values will be reset before the packets enter the extended PIP domain and the flow will not have any quality of service on the remainder of the way.

Although indifferent and supportive domains both treat DSCP in a transparent way, they are not grouped together. An indifferent domain, preserves DSCP value and therefore the existence of a PIP flow but it does not guarantee any QoS. In contrast supportive domains provide better service than Best Effort (normally via over-provisioning) and thus preserves the PIP flow existence and its characteristics. For this reason PIP supportive domains may, and are encouraged to, take part in the extended PIP domain, but indifferent domains may not.

Regardless of status, all domains on an end-to-end path which are not part of the extended PIP domain (e.g. those comprising “the last mile”) are encouraged to allocate resources to assist in troubleshooting and deploy a suitable monitoring infrastructure, such as that recommended in section 3.6.4.

A PIP supportive domain which joins the extended PIP domain must follow exactly the same rules as a compliant domain in term of SLAs. Note that in a PIP supportive network which relies on over-provisioning IPDV will most likely not be guaranteed to any high degree. This handling of IPDV is likely to be the main difference between the levels of service offered by PIP compliant and PIP supportive domains.

### 3.4.4 Disqualifying an SA3-compliant domain

An SA3-compliant domain may be disqualified if it does not operate a measurement system on its network in order to guarantee the SLAs, or if it habitually fails to respect the SLA parameters it offers.

Each parameter described in section 3.5.1.2 can disqualify a compliant domain if the measured value exceeds the threshold specified in the SLA. For example:

- When the OWD value for the 99% of packets served through the domain exceeds the value specified within the SLA, and then the domain is disqualified.

- If the available capacity for an accepted PIP flow within a domain is measured to be less than the requested capacity during the reservation process for a reason different than a failure in the normal routing path of the flow the PIP compliant domain is disqualified.
- When the IPDV exceeds the value specified within the SLA, then the domain is disqualified.
- If the MTU changes and becomes less than the value defined in the SLA, packets loss will appear, so this parameter is also disqualifying.
- When the packets loss is more significant as the defined value in the contract and persistent, the PIP flow throughput is no more guaranteed and then the compliant domain is disqualified.

When an SA3-compliant domain is disqualified, its neighbouring domains will disable the peering between their Provisioning Systems and its own. Where possible the GN2 PIP community will work with the disqualified domain so that they may improve their systems and/or network performance and so re-gain their previous status.

It should be noted that disqualification should only be used as a last resort and after efforts have been made to remedy the situation. For example, a domain might fail to meet its SLAs if it specifies impossibly high levels of service. In this case the situation could be resolved by decreasing the level of service offered to the true achievable level.

### 3.5 Service Level Agreements (SLAs) for the PIP service

For the successful deployment of the Premium IP service across Europe it is necessary that a series of SLAs are established among the participating domains. Each PIP compliant domain has to offer a set of SLA assurances to all its PIP compliant peers. The bilateral SLA that needs to be provided by a PIP compliant domain to each of its peers has been analytically presented in [SEQUIN D2.1 addendum 2]. This requirement extends the definition of a PIP compliant domain as it was presented in section 1.2.2 to also include the requirement for the domain to support monitoring by measurement tools the SLA parameters and supporting bilateral SLAs.

This section focuses on the establishment of end-to-end SLAs for each PIP service instance from one edge of the extended PIP domain to another, based on the intermediate per-domain SLAs.

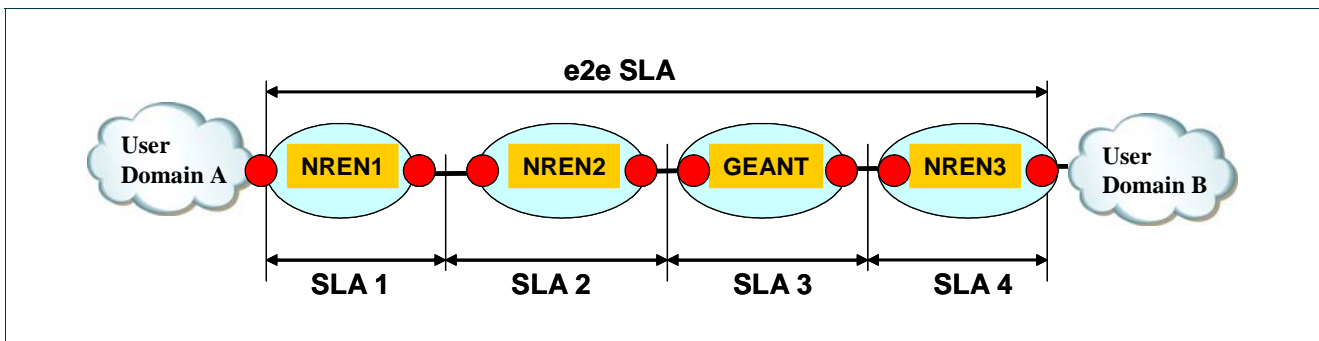
The goal of each such SLA is to provide end-to-end control on PIP service in a multi-domain context. In principle, the end-to-end SLA has to be established so as to contain all related and necessary parameters for the service instance specification between the two endpoints requiring PIP service, regardless of the underlying network and intermediate domains. Some considerations are pointed out here:

- Path. The provisioning of end-to-end services is based on investigation and monitoring of all the per-domain SLAs along the path.
- Unidirectional. Since each PIP flow is unidirectional, the related SLA is unidirectional too.
- Static or Dynamic negotiation. The establishment of the SLA can be negotiated between user and NRENs in static or dynamic mode. In Phase 1 of the PIP Provisioning System static negotiation will be supported

- Continuous Monitoring. The continuous monitoring of all SLA parameters subscribed to in the contract between user and the extended PIP domain is mandatory, for the duration of PIP session (see section 3.5.1.2 below). The extended PIP domain must provide to end-users evidence that the end-to-end SLA is respected. SLA monitoring must not impact on the network performance. A basic condition to implement an end-to-end SLA is that the corresponding PIP session spans only through SA3-compliant domains.
- Quick reaction to the violation of the SLA terms is very important. A contact point with good network knowledge should be available at each PIP compliant domain along the path in order to quickly address disruption events, in order to reduce unavailability and violation of the service provisioning. The contact points for each domain should be included in all its offered PIP SLAs and the end-to-end SLAs should include all the contact points along the end-to-end path.
- The QoS metrics of an end-to-end SLA across multi-domains have to be derived from the corresponding metric values at each traversed domain through the path. The main issue to build up a true end-to-end SLA is the multi-domain nature of networking between domains. Actually in a pan-European scale, involved domains (NRENs, GEANT etc.) are part of a confederation, each with their own independent administrative domain. The resulting end-to-end SLA has to be determined from the per-domain SLAs. For example assuming that availability inside a PIP compliant domain is independent from its peers, the availability ensured for each PIP session along the extended PIP domain is the product of the individual per-domain availabilities along the path. The domain chain described is depicted in **Figure 3.4**. From a practical point of view, the domain with the worst availability has the most significant effect on the resulting total availability in the chain.

$$e2eAvailability = \prod_i^N (DomainAvailability)_i$$

where the product is extended to the traversed domains along the path.



**Figure 3.4:** The end-to-end SLA and the per-domain SLAs involved

### 3.5.1 SLA parameters

An SLA is a set of technical and non-technical parameters agreed between customer and NRENs. As described in [SEQUIN D2.1 addendum 2], an SLA specification is composed of two parts:

- Administrative/legal part

Project:	GN2
Deliverable Number:	D.S.3.9.1
Date of Issue:	17/03/05
EC Contract No.:	511082
Document Code:	GN2-05-017v7

- SLS (Service Level Specification) part, as a set of parameters and their corresponding values that define the technical parameters of PIP service provisioning to a PIP flow

It is useful to deal with the SLA components as two objects, namely an Administrative Level Object and a Service Level Object.

### 3.5.1.1 Administrative Level Object (ALO)

The Administrative Level Object comprises of a number of fields that define the procedure and framework for the provision of the PIP service to each request. Each of the fields can include one or more parameters. The template structure of the ALO for a PIP compliant domain is shown in **Table 3**

For each PIP flow served through a domain a different copy of the ALO template must be produced and kept for reference purposes. This copy will be automatically produced and stored in electronic format within the PIP Provisioning System.

Parameters	Sub-field	Description
PIP Provisioning Administrative and Technical contacts	Full contact details (e-mail, telephone & fax nos., address) for an administrative and a technical contact person	This field should always contain updated information of a single point of contact within a PIP compliant domain for administrative (in case that a new user/ user-group has to be created) and technical (in case of some custom configuration or for troubleshooting) purposes. For each PIP session through the domain, the values of this field in the corresponding SLA instantiation will remain the same
SLA Duration		Period for which the SLA is valid. All requests for PIP flows being served through a particular PIP compliant domain can be served only if the requested Start Time and End Times fall within the SLA 'Duration' interval. For each PIP session, the value of this field in the corresponding SLA instantiation will have to be updated with the actual Start Time and End Time
Availability guarantees	Service Availability for the SLA Duration  Source of availability data  PIP service availability estimation	This field should contain all necessary information for the estimation of the availability of the PIP service as described in the SLO. A percentile of the time that the PIP service is actually available over the duration of the service provisioning should be included here, as well as a formula for deriving this percentage and the source of 'uptime' information should be provided. It is likely that the values of these fields will be the same for all PIP sessions within the

		domain.
Response times	Domain Acceptance Response Time Reservation Request Lead Time	This field should contain response times guaranteed by the PIP compliant domain to client requests for new PIP sessions or adjustment of existing ones and for carrying out a PIP provisioning instance, based on the necessary configuration of routers
Fault handling-trouble ticket procedures	Failure Indication Response Time Set of actions taken	This field should specify the maximum response time of the technical contact person to a user initiated trouble indication, as well as the steps followed to fix the problem (troubleshooting within the domain, contact upstream/downstream PIP compliant domains' contact persons). The contents of this field in each SLA instantiation for each PIP session within the domain will remain the same.

**Table 3: Administrative Level Object (ALO) template**

### 3.5.1.2 Service Level Object (SLO)

A Service Level Object (SLO) is an instance of the SLS template that contains the parameters and their values that qualify the transport properties of each PIP service provisioning instance.

To qualify a bi-directional service a combination of two SLOs is required, one for each direction. Each SLO is generated automatically when a service is negotiated. Again here, for each PIP flow served through a domain a different copy of the SLO template together with the corresponding values must be produced and kept for reference purposes. This copy will be automatically produced and kept in electronic format within the PIP Provisioning System. Since a number of fields in the SLO of a PIP flow refer to or depend upon ALO fields and vice versa, the two components of the SLA provided to each PIP session have to be kept as a unified SLA record.

The template structure of the SLO is shown in **Table 4**

Parameters	Description
PIP service instance scope	This field should contain the technical information of the ingress interface and the egress interface of the PIP compliant domain, between which a PIP flow is served. For each PIP session, this field in the corresponding SLA instantiation will have to be updated with the actual interfaces traversed by the packets of the PIP flow

<p>PIP flow description</p>	<p>This field should contain the DSCP value and the IP source-destination address pair (in case of a PIP origin domain) or the upstream-downstream AS number of the PIP compliant domains (for a PIP transit domain) that uniquely describe a PIP flow within the domain. This descriptor will have to match the policer definition data for the PIP flow. For each PIP session, this field in the corresponding SLA instantiation will have to be updated with the actual descriptors of the current PIP flow</p>
<p>Performance guarantees</p>	<p>This field should contain the values of the qualitative parameters guaranteed by the PIP compliant domain to the packets of the PIP flow. These values will be statistical data based on measurements across the domain. Also in the initial stages of the service deployment, the values will have a domain-wide validity for all PIP flows. In later stages, and as the monitoring infrastructure becomes more dense within a PIP compliant domain, it will be possible to extract specialised performance guarantees per PIP session. The individual parameters for this field are:</p> <p>OWD: An indicative value is the distance delay + 50ms. The distance delay can be roughly computed using a signal speed of about 7 us/km.</p> <p>IPDV: An indicative value is &lt;10ms.</p> <p>Packet loss: An indicative value is 0.5%.</p> <p>MTU: The suggested value for a WAN is 4470 bytes. MTU value is unlikely to change during the lifetime of each PIP flow. A variation on MTU could be due to rerouting on different path or mis-configurations along the path.</p> <p>Capacity: The PIP Capacity guaranteed to the PIP flow (It could change if the PIP path within the domain changes during the life of PIP flow)</p>
<p>Traffic Envelope and Traffic Conformance</p>	<p>This field should contain the details of the policer profile used for the PIP flow, thus the type of the Traffic Envelope used (e.g. Token bucket with a CIR and a Burst Size) and the actual conformance values used, together the corresponding Averaging Intervals, if available (see section 2.1.1)</p>
<p>Excess treatment</p>	<p>This field should specify how excess traffic (or out-of-profile traffic, according to the profile described by the Traffic Envelope and Traffic Conformance field) is treated by the PIP compliant domain. Unless there is a specific reason to do otherwise, the contents of this field in each SLA instantiation for each PIP session within the domain will remain the same.</p>
<p>Monitoring</p>	<p>This field should contain information on the PIP monitoring capabilities of the domain in terms of points where measurements are possible, the availability of</p>

Project:	GN2
Deliverable Number:	D.S.3.9.1
Date of Issue:	17/03/05
EC Contract No.:	511082
Document Code:	GN2-05-017v7

infrastructure	measurements (continuous, per-case, on-demand), which parameters are monitored at each point, granularity (time intervals) for the measurements of each of the parameters in the 'Performance guarantees' field and methodology for statistical processing of monitoring data
Reliability	<p>This field should include:</p> <ul style="list-style-type: none"> <li>• allowed maximum downtime for the duration of the PIP session (MDT)</li> <li>• maximum allowed time to repair (TTR) in case of downtime in the service provisioning</li> </ul> <p>For each PIP session through the domain, the values of this field in the corresponding SLA instantiation will remain the same</p>

**Table 4: Service Level Object (SLO) template**

### 3.5.2 End-to-end SLA

It would be ideal if the granularity for monitoring of SLO parameters were uniform along the whole path, across all domains. However, in the first stages of e2e PIP service Provisioning, the end-to-end SLA will simply consist of the set of per-domain PIP SLAs along the path within the PIP extended domain.

The minimum requirement for the end-to-end SLA is for its SLO parameters to be monitored as follows:

- Between cross-border links, for each pair of cross border links along the path within the PIP extended domain, and
- Between the ingress interface of the origin domain and at the egress interface of the extended PIP domain towards the destination user

This information will have to be included in the SLO of the end-to-end SLA and verified against the values provided in the constituent bilateral SLAs.

For Phase 1 of the PIP Provisioning System implementation SLA monitoring could be restricted on the aggregation of all PIP flows at the cross-border links. In later Phases it could be extended to individual PIP flows.

Project:	GN2
Deliverable Number:	D.S.3.9.1
Date of Issue:	17/03/05
EC Contract No.:	511082
Document Code:	GN2-05-017v7

## 3.6 Integration of policies - recommendations for NOC operating procedures

All NRENs which participate in PIP Provisioning should appoint at least two people to be POCs for all matters relating to PIP. These people (termed PIP POCs) should have detailed knowledge of PIP, in terms of both their own NREN's implementation of it and also the overall PIP service as offered by the participating GN2 partners. They should be fully conversant with all PIP documentation, including this Policy document, and should maintain close ties with the PIP POCs in neighbouring domains. The PIP POCs are responsible for producing and maintaining their own NREN's specific policy on the usage of PIP which, unless their NREN explicitly announces to the contrary, should adhere to what is contained in this Policy document. The PIP POCs should also provide instructions to the NOC as to what to do in the event of a problem with PIP (PIP packets being dropped, re-routing of PIP due to link failure etc). In particular, they should ensure that in the event of a PIP problem the NOC staff knows who in adjacent domains to contact if the PIP POCs themselves are not available. As far as possible PIP POCs should be chosen so as to minimise the likelihood of both/all PIP POCs being out of office at the same time.

### 3.6.1 Creation of User-Groups and UserIDs

With the exception of those working for specified pan-European projects, anyone wishing to use PIP will first need to apply to their domain's NOC for a UserID and (if they are not joining an existing User-Group) new User-Group. It is at the NREN's discretion as to whether such a user (termed 'domestic user' to differentiate them from users from special projects) should be allowed to use PIP. It is also the NOC's decision as to what boundaries should be set for a given User-Group (in terms of maximum possible duration of a reservation, minimum possible duration, maximum possible capacity of a reservation etc). Note, these boundaries form the main part of that User-Group's policy.

### 3.6.2 Distribution of Quota

PIP is free to the end-user at point of use and as such 'quota', as described in section 3.4.1, is the recommended means by which demand for the PIP service should be regulated. The quota for large-scale, pan-European projects will be discussed and agreed by participating NRENs (and DANTE) in advance. For all other User-Groups NOCs are free to allocate quota to whomever they wish and in whatever quantity they wish, but it is recommended that the quota for external User-Groups (i.e. those whose requests come from another domain's PIP Provisioning System) are aggregated into a single value associated with that domain. In allocating quota to User-Groups NOCs will be able to set a quota replenishment rate and a maximum quota level. Depending on the domain's Provisioning System, the NOC may also be able to set a frequency of replenishment (e.g. daily, weekly, etc).

Project:	GN2
Deliverable Number:	D.S.3.9.1
Date of Issue:	17/03/05
EC Contract No.:	511082
Document Code:	GN2-05-017v7

### 3.6.3 Pre-Quota Guidelines

Quota will not be used in Phase 1 of the PIP service, since the first version of the Provisioning System will not feature it. To say 'quota will not be used' is the equivalent of saying there will unlimited quota for all users, and therefore it is prudent for NRENs to take some other steps for managing demand during this phase of service. Whenever a request is accepted by the Provisioning System (which it will be if there is sufficient PIP capacity and the basic boundary limits are kept to), NOC staff should manually check the request to ensure it is in keeping with what was agreed with the user when their account was first created. A final check should be made by the NOC staff when they come to apply the configuration (which they must always do manually in the early phases of the Provisioning System, before automated configuration is possible).

### 3.6.4 Monitoring of QoS parameters

For each PIP flow which originates in their domain a PIP participating domain must be able to monitor the flow's data rate, averaged over not more than 10 minute intervals. PIP participating domains are strongly encouraged to deploy performance monitoring systems which either directly or indirectly measure the following parameters for a PIP flow:

- OWD
- IPDV
- packet loss

### 3.6.5 SLA handling

A Service-level Agreement (SLA), as would exist between two end-points, is defined in section 3.5.

NOCs should provide the means to monitor technical parameters defined in SLA agreement, such as capacity, MTU, IPDV, OWD and packet loss.

Specifics such as which parameters are important for end-to-end quality of service in Premium IP services should be agreed upon between domains and in accordance with the network policy of the domain (e.g. specific parameter monitoring may not be recommended on core routers in general, and might only be allowed only for a short period of time).

### 3.6.6 Accounting

Accounting will be conducted based on resources used. When accounting is introduced in Phase 2 the only two parameters of interest will be reserved bandwidth and duration of the reservation i.e. quota.

Reserved resources will be recorded in the PIP Provisioning System; used resources will be monitored by NOC through network monitoring system, as duration and amount of bandwidth used by the flow. Combined, this data will enable a NOC to generate a report on the PIP service in its network covering, in particular:

- capacity reserved
- capacity used
- number of successful reservation requests
- number of unsuccessful reservation requests (and reason for failure)

### **3.6.7 Troubleshooting a PIP service failure**

Each NOC is responsible for monitoring PIP parameters in its domain. In case of any violation of agreed parameters (bandwidth in time in the first phase of the implementation) on any part of the path within this domain, the cause of the violation should be investigated with urgency and the affected PIP user(s) informed, along with the neighbouring domains.

If the NOC observes that a link reaches a given threshold (as specified in the their procedures, but typically 50%), then steps should be taken to prevent any more PIP requests which would use that link from being accepted until the cause of the overload has been identified and fixed.

## 4 Conclusion and Next Steps

The PIP allocation policy was created as the first step towards a full production inter-domain PIP service. The PIP parameters were thus agreed amongst partners to be inclusive of a wide number of networks, including those which do not fully comply with SEQUIN principles. On the other hand, the precise details of the service, given here, will provide clear guidance to the newcomers as to what is expected of them in their role as end-to-end PIP participants. In conjunction with the reference Provisioning System being developed, it is expected that this will facilitate the wide adoption of inter-domain PIP service in the whole of the European research networking community.

The multi-domain PIP service will begin as a limited pilot, consisting of the GEANT2 backbone network, GRNET, GARR and Super JANET 4. Figure 4.2 below provides a rough sketch of the pilot. Each of these (and subsequent) networks will deploy an appropriate Provisioning System, which may include some or all the components of the SA3 Provisioning System reference model. Note that if a PIP service participant wishes to deploy their own Provisioning System it must nevertheless be fully compatible with the SA3 reference model, in that it must use the interfaces as defined by the SA3 PS development team.

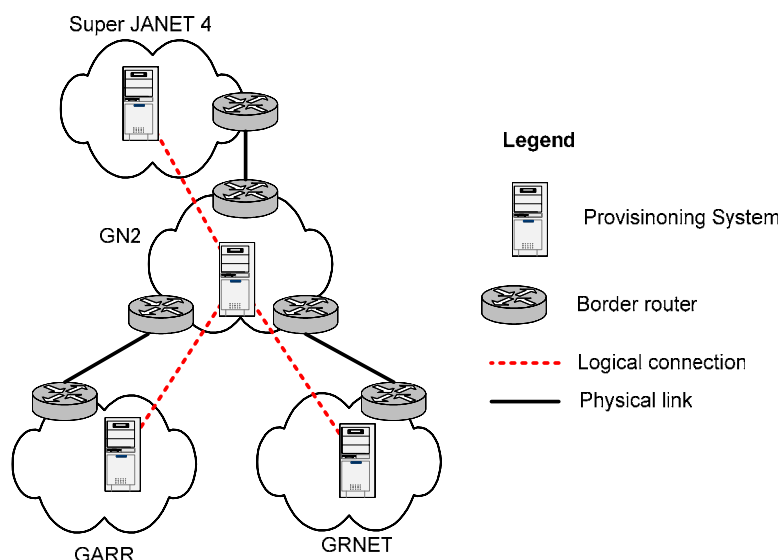


Figure 4.1: Pilot PIP Service

Project:	GN2
Deliverable Number:	D.S.3.9.1
Date of Issue:	17/03/05
EC Contract No.:	511082
Document Code:	GN2-05-017v7

## 5 References

[SEQUIN D2.1 addendum 1] M. Campanella, 'SEQUIN D2.1 - Addendum 1 Implementation architecture specification for the Premium IP service', 2002

[SEQUIN D2.1 addendum 2] A. Sevasti, M. Campanella, 'SEQUIN D2.1 - Addendum 2 SLA specification for a Premium IP service', 2002

[SEQUIN D4.2] M. Campanella, R. Sabatino, 'SEQUIN D4.2 QoS Implementation Plan', 31 May 2002

[RFC2474] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, 'An Architecture for Differentiated Services', December 1998

[RFC2598] V. Jacobson, K. Nichols, K. Poduri, 'An Expedited Forwarding PHB', June 1999

[RFC 3246] B. Davie, A. Charny, J.C.R. Bennett, K. Benson, J.Y. Le Boudec, W. Courtney, S. Davari, V. Firoiu, D. Stiliadis 'An Expedited Forwarding PHB (Per-Hop Behavior)', March 2002

[GN2-04-153] A. Patil, T. Rodwell, Premium IP Provisioning System - Design Specification, December 2004

## Appendix A An example of a PIP compliant domain: GÉANT network

GÉANT is a PIP compliant domain as follows:

The DSCP/ToS value used in GÉANT to identify PIP traffic is shown in the table below.

Service	DSCP value	ToS value	Juniper alias	ToS (hex)	DSCP	ToS byte (binary)
Premium IP	46	184	ef	B8	101110	10111000

**Table 5: Use of DSCP value**

A DSCP binary value of 101110 (46 decimal) (according to the [RFC 2598]) corresponds to a Precedence 101 (binary) and a ToS (110 binary) (low delay, high throughput, normal reliability).

GÉANT uses DSCP 46 to identify PIP traffic within its core

It identifies legitimate PIP packets by examining their source/destination IP address and distinguishes among different PIP flows at border routers. Traffic that is marked with the PIP DSCP that GÉANT accepts but doesn't match any source/destination address pair is remarked as best effort as shown in the table below

Service	Incoming DSCP value	New DSCP value
Un-authorized Premium IP traffic	46	0/5 <sup>2</sup>

**Table 6: Remarking**

It policies all packets of each PIP flow arriving at a border router with a token bucket policer imposing an average rate and a maximum burst size. Out-of profile packets are dropped, as shown in the table below

<sup>2</sup> The packets coming in on a Juniper M160 router are retagged with the DSCP value 0, while those coming in on a Juniper M40 router are retagged with the DSCP value 5 (PREC rewritten, three last bits of the DSCP field kept unchanged).

Service	Incoming DSCP value	New DSCP value
Identified Premium IP flow	46	46/drop <sup>3</sup>

**Table 7:** Dropping of out-of-profile packets

GÉANT PIP service provides a mechanism to examine whether a request for serving a PIP flow can be accepted or not and provides a response to such requests

GÉANT uses an absolute priority queue to serve PIP traffic in its core and a subscription level for PIP traffic of up to 10% on its links. This has been qualitatively shown to ensure bounded minimal end-to-end delay and reduced jitter as well as minimum packet loss for all packets of each PIP flow that has accepted to serve.

GÉANT implements its request admission policy for PIP based on the 10% subscription rule.

---

<sup>3</sup> Packets of an authorised PIP flow are evaluated against the flow policer. If the packets are in-profile, they are accepted and the tagging is kept unchanged. If they are out-of-profile, they are dropped.

## Appendix B Suggested 'boundary conditions' for the PIP Allocation Policy

Boundary Condition	User group(s)	EGEE	(All Others)
Min allowed reservable capacity per request		100kbps	100kbps
Max allowed reservable capacity per request		200Mbps	100Mbps
Min allowed duration of PIP reservation per request		1 day	1 day
Max allowed duration of PIP reservation per request		3 months	1 month
Max bandwidth-time product for a single reservation		16TB	12TB
Max number of concurrent active reservations in a domain		-	100 (total)
Max number of reservations (active/pending) per User-Group		(40)	3
Max number of starting/ending reservations		4	10
Max lead time allowed		3 months	2 months
Min lead time		2 days	2 days

## Appendix C SA3 glossary

<b>Premium IP service</b>	An added-value network service that provides guarantees on bandwidth and minimises one-way delay, IP packet delay variation and packet loss percentage for IP traffic across one or more interconnected domains
<b>One-Way-Delay (OWD)</b>	OWD is the time between the transmission of a packet at its source and the moment it is fully received by the destination
<b>IP packet delay variation</b>	IPDV for an IP flow of packets is the difference of the one-way-delay of one packet and the one-way-delay of the next packet of the same size in the same flow
<b>Origin domain</b>	The domain where a PIP-flow first enters the extended PIP domain.
<b>Transit domain</b>	All domains in the extended PIP domain other than the origin domain
<b>PIP flow</b>	A PIP flow within a domain consists of all packets marked with DSCP 46 and either contain the same source IP address prefix - destination IP address prefix pair in their IP headers or originate from the same upstream Autonomous System (AS) and are destined for the same downstream AS
<b>PIP compliant domain</b>	A domain that adheres to the Premium IP specification of [SEQUIN D2.1 addendum 1], thus implements the PIP service with an EF-PHB compliant manner and with the following features: Classification of PIP packets Remarking of traffic at the ingress Policing to control exceeding PIP traffic Admission control on PIP service requests A well-known QoS domain profile
<b>PIP session</b>	A PIP flow service instance across an extended PIP domain
<b>PIP supportive domain</b>	A domain that preserves Premium IP DSCP value, offers an environment where in general Premium IP characteristics hold (e.g. by over provisioning) and operates a Provisioning System.
<b>UserID</b>	A unique identifier for a user on the PIP Provisioning System
<b>User-Group</b>	The organization/group under which a user is allowed to use the PIP service.
<b>Extended PIP domain</b>	A contiguous chain of two or more PIP compliant and/or PIP supportive domains
<b>PIP Provisioning System</b>	A set of software modules used for the processing of PIP requests, the decision making process for accepting or rejecting them, for enforcing the configurations on network elements required, for maintaining PIP resources' allocation information and in general the automated –eventually- operation of the PIP service both at an intra-domain and at an inter-domain level.
<b>User Acceptance Response Time</b>	It is the maximum time spent by the inter-domain PIP Provisioning System to answer to a user request for serving a PIP flow

<b>Domain Acceptance Response Time</b>	It is the maximum time spent by an intra-domain Provisioning System to answer to a user request for serving a PIP flow either as an origin or as a transit domain
<b>Reservation Request Lead Time</b>	It is the maximum allowed interval between the moment when a user request for PIP service is submitted and the Start Time for service of the request
<b>PIP user</b>	He is an individual person or software application which is allowed to make PIP requests across an extended PIP domain
<b>PIP Quota</b>	A quantity that defines the maximum of PIP bandwidth B that can be reserved for the maximum allowed duration of a PIP reservation. The PIP Quota amount can be re-assessed periodically and/or upon request according to the internal policy of each PIP compliant domain.
<b>PIP incompatible domain</b>	A domain which resets or discards packets with Premium IP DSCP values
<b>PIP indifferent domain</b>	A domain which preserves Premium IP DSCP value but applies only Best Effort treatment
<b>SA3-compliant domain</b>	A PIP compliant or PIP supportive domain