Project Number:	**IST-1999-20841**
Project Title:	**SEQUIN**

Deliverable D2.1 - Monitoring and Verifying Premium IP SLAs

Addendum 3

Deliverable Type:	PU-Public
Contractual Date:	30 April 2002
Actual Date:	30 April 2002
Work Package:	
Nature of Deliverable:	RE - Report

Author: **Athanassios Liakopoulos**

ABSTRACT:

SEQUIN investigates methods for providing QoS differentiated services over the IP-based networks such as the GÉANT network. Initially, the project defined a framework for supporting IP Premium service and described the procedures for establishing Service Level Agreements (SLAs) amongst networks such as GÉANT and the NRENs. In succession, SEQUIN investigates methods for monitoring the network behaviour and techniques for assessing the quality of delivered services to the end-users, i.e. NRENs.

This document reviews various IP measurements methods and describes monitoring architectures that may be deployed in large-scale backbone networks. Furthermore, it provides a general framework for monitoring QoS provisioning in GEANT.

Key Words: Monitoring, Network measurements, Quality of Service, IP Premium, GÉANT

D2.1 Addendum 3
SEQUIN: Monitoring and Verifying Premium SLAs

1. Introduction

Competition is forcing Service Providers (SPs) to differentiate themselves in the open telecommunications market by providing diverse service classes. Apart from the well-known Best Efforts service class, where no guarantees are provided, new service classes are specified with detailed SLAs agreements that strictly determine the network behaviour in terms of performance metrics. Signed SLAs agreements between SPs and their customers include explicit performance metrics at the network level, such as network delay, jitter or bandwidth.

By installing the necessary monitoring infrastructure, SPs will be able to check the conformance of the network behaviour to the agreed SLAs and verify its compliance to the agreements with their customers. In the case of a network problem, monitoring nodes would immediately inform the SPs Network Operation Centres and gather enough information to locate the problem (in a network). Another important reason for installing monitoring infrastructure is that customers would accept being charged more for enhanced services than they are being charged today for a Best Efforts service, under the condition they can verify the provided QoS service from the network. Consequently, SLAs compliance verification is an important and significant challenge for SPs.

This document provides a general framework for the monitoring of SLAs for different Classes of Services (CoSs) in a network. The document is organised as follows. In Section 2, we present issues regarding monitoring backbone networks, such as scope, synchronisation and presentation of monitored parameters. Section 3 presents the advantages and disadvantages for building a monitoring infrastructure based on freeware or commercial products. Section 4 briefly presents the GÉANT network and its supported CoSs. Finally, sections 5 and 6 present the proposed monitoring infrastructure based on RUDE/CRUDE tool and RIPE TTM test-boxes.

2. Monitoring Issues

As stated previously, SPs that offer diverse services classes to their customers should constantly monitor the performance of their backbone network and verify its compliance to the agreed SLAs. This is usually a complicated and extremely cumbersome task but it is a strict requirement imposed by the customers.

Today, SPs do not usually deploy a dedicated monitoring infrastructure as it requires the purchase and installation of expensive equipment, especially in large scale networks. Additionally, SPs would have to cope with the managerial burden of the extra equipment, which is also a dissuasive reason. Therefore, SPs usually partition their network into logical segments, e.g. small network areas or significant high-speed links, and they monitor each segment separately. This is generally a simple task, which is performed mainly by high performance routers without the use of extra equipment. Network statistics are collected in SNMP MIBs and are periodically sent to specific servers. The latter collect data from different network nodes and attempt to evaluate the performance of the network using post-processing analysis techniques. Although this method can be easily implemented in large-scale networks, it is inappropriate to determine the precise end-to-end behaviour of the network and is mainly used for troubleshooting purposes.

2.1 Monitoring Scope

SPs are only responsible for the part of the connection path between the end-users that belongs to their administrative domain. In some cases, results collected by the monitoring infrastructure may verify the compliance of the network behaviour to the customer SLA but, at the same time, the end-users may perceive significantly lower Quality of Service due to congestion problems on the customer internal network. Consequently, an SP can avoid disputes with its customers by carefully choosing measurement points for each customer separately and explicitly defining these points in the signed SLA contracts.

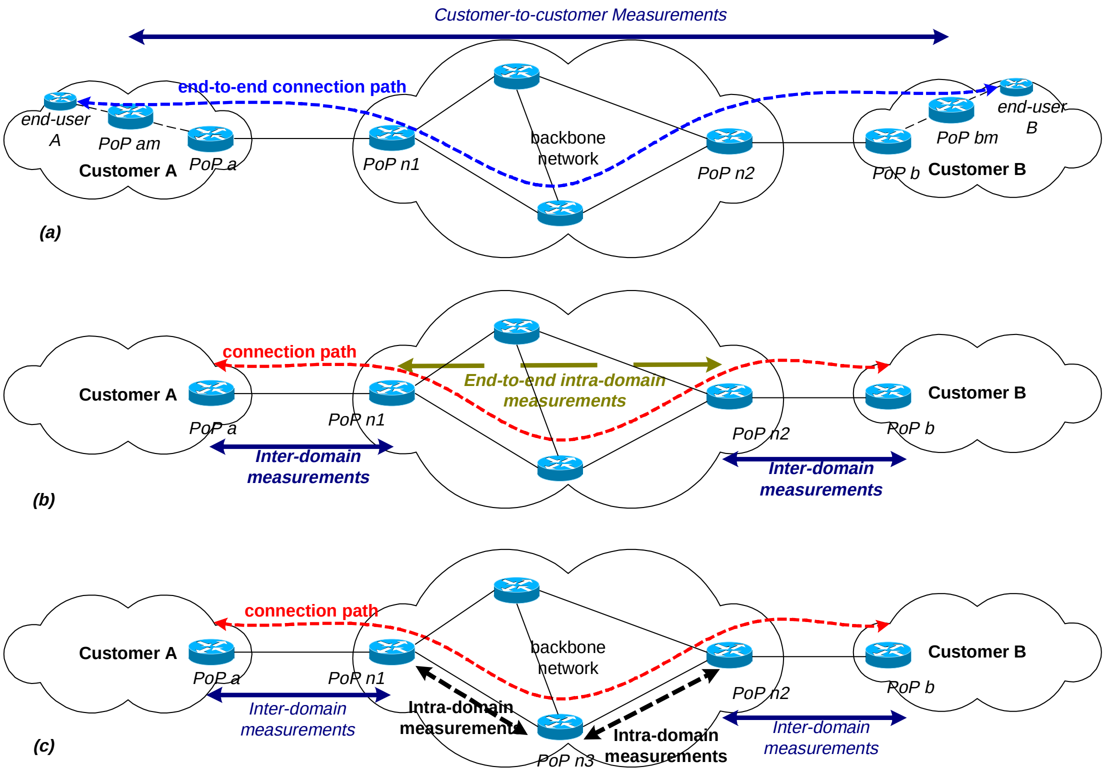
Customer-to-customer Measurements
end-to-end connection path
end-user A
PoP am
Customer A
PoP a
PoP n1
backbone network
PoP n2
end-user B
PoP bm
PoP b
Customer B
(a)
connection path
Customer A
PoP a
PoP n1
End-to-end intra-domain measurements
PoP n2
Customer B
PoP b
Inter-domain measurements
Inter-domain measurements
(b)
connection path
Customer A
PoP a
PoP n1
backbone network
PoP n2
Customer B
PoP b
Inter-domain measurements
Intra-domain measurements
PoP n3
Intra-domain measurements
Inter-domain measurements
(c)

Figure 1: Alternative topologies in deploying monitoring infrastructure

Figure 1 presents how packets originated from end-user A with destination to end-user B are forwarded in a typical network. At the beginning, packets will travel through the customer's internal network, passing through a router at major customer’s PoP, and the border router at PoP. Afterwards, packets will travel in the backbone network passing through the backbone routers at PoP$_{n1}$, PoP$_{n3}$, PoP$_{n2}$. Finally, packets will travel through the customer's internal network until they reach the final destination.

There are different alternative topologies for deploying monitoring infrastructure that audits the network behavior at the network in Figure 1.

1. Monitoring nodes are located at the major customers’ PoPs, e.g PoP$_{am}$ and PoP$_{bm}$, which differ from the customers’ access PoP towards the upstream network [Figure 1(a)]. Customer-to-customer measurements of the network performance, i.e. from PoP$_{am}$ to PoP$_{bm}$, are performed.

2. Monitoring nodes are located at the customers’ access PoPs and at the SPs’ aggregation PoPs, near to the border routers. Inter-domain measurements are

performed on the connection path between the customer and the SP, e.g. between the PoP_a and PoP_{n1}, and between the connection path inside the SP's network, e.g. between PoP_{n1} and PoP_{n2} [Figure 1(b)]. Alternatively, inside the SP's backbone network, multiple intra-domain measurements are performed along the connection path, e.g. between PoP_{n1} and PoP_{n3}, PoP_{n3} and PoP_{n2} [Figure 1(c)].

Deploying and operating monitoring nodes at the customers' PoPs [Figure 1(a)] may provide accurate measurements of the end-to-end network behaviour, as the monitoring nodes are closely placed to the end-users nodes. However, this scenario makes it difficult to locate where QoS service provisioning performance violation has occurred, as there is no way to distinguish which administrative domain misbehaves. Service degradation may occur due to congestion problems within the customers' domains or in the SP's backbone network. Consequently, monitoring nodes should be distributed at customers' and at SPs' PoPs [Figure 1(b)]. Different set of measurements are performed between the customers' access PoPs and the SP's aggregation PoP and between SP's aggregation PoPs. By deploying multiple monitoring nodes in all the aforementioned PoPs, it is easy to identify which part of the customer-to-customer connection path is responsible for the problem. For troubleshooting purposes, monitoring may be performed on per link basis (intra-domain measurements) [Figure 1(c)]. However, it is not efficient to calculate end-to-end intra-domain performance by combining measurements on each link along the path between two edge nodes.

2.2 Active / Passive Monitoring Methods

Monitoring methods are classified as either active or passive. In active monitoring methods, synthetic monitoring traffic is generated and injected into the network by monitoring nodes. Monitoring traffic shares the same network resources with the real traffic and therefore it encounters the same queue delays and packet losses in congested networks. As monitoring traffic mainly consists of counters and accurate time data (timestamps), it is used for post-processing analysis of the network behaviour. Although active monitoring methods consume network resources and put additional load to the network, they are widely used because they provide accurate measurements of the network behaviour.

Passive monitoring methods are based on information collected by network nodes, usually backbone routers. While real traffic is passing through the network, routers collect information concerning the network parameters, such as the number of packets successfully delivered to a specific destination network or discarded during queue congestion. Passive monitoring methods cannot easily be used for evaluating the network behaviour due to the lack of accurate network synchronisation between the network nodes. However, they are easily deployed and they do not burden the network with extra traffic.

2.3 Synchronisation

One-way packet delay is one of the most significant measured parameters that describe the network behaviour. However, performing accurate time measurements between two nodes of the network is a difficult task as fine time synchronisation of the source and destination nodes is required. Obviously, the accuracy of the clock synchronisation of the monitoring nodes limits the precision of the one-way delay measurements.

There are three different approaches in network synchronisation that differ significantly on the accuracy of their results. Actually, there is a trade-off between accuracy, complexity and cost. In order of ascending accuracy precision, the synchronisation methods are the following:

a. *Internal Time Source:* Each node has its own clock for measuring time. Synchronisation with “global” network time is performed only at the start-up phase of the node, where the right initial time value is set and the internal clock is tuned to the right speed.
b. *Phase Locked Loop:* Each node continuously adjusts its time based on the deference between its internal-clock time value and the external-clock time value.
c. *External Time Source:* Each node uses a precise external time source, such as a precise atomic clock or more commonly a GPS receiver.

In today's networks, the last two methods are commonly used. Generally, network equipment is synchronised according to the phase locked loop method using NTP protocol. Few accurate NTP servers, however, are synchronised with external time sources, usually GPS receivers.

The next paragraphs briefly present the NPT protocol and the GPS:-

2.3.1 NTP protocol

The Network Time Protocol [1] is designed to timely synchronise nodes over the network. It is based on UDP protocol for exchanging time related information between an NTP server and its clients.

An NTP server distributes time information through the network to other (less accurate) servers and clients. The latter adjust their clocks based on the information they receive from the server. In some cases, few packets per minute are necessary to synchronise the two nodes to within a millisecond to each other.

Primary NTP servers are perfectly synchronised to authoritative external time sources, usually not very expensive GPS receivers. Secondary servers, i.e. servers that do not get their time from accurate external time resources, receive time information over the network from the primary servers or other secondary servers. This method of distribution time information imposes inaccuracies in the synchronisation of the secondary servers. In general, the further away in terms of “ntp hops” a client is from the primary server, the lower the accuracy of synchronisation.

The synchronisation procedure between the primary and secondary NTP server is based on time stamp information that is exchanged between them. The primary server sends time information to the secondary server using UTP packets. The latter has to estimate the time offset between its time and the primary server’s time. However, time offset estimation procedure has to take in to account the time elapsing while the UTP packet is transferred over the network, i.e. the one-way delay of the packet. However, the secondary server can only measure the round-trip-delay to the primary server and afterwards an estimation of the one-way-delay is made. Consequently, estimation of one-way-delay is not accurate.

2.3.2 Global Position System

The Global Position System (GPS) [2] consists of a constellation of 24 satellites orbiting the earth. It is a position and navigation system, which is funded and operated by the US military.

The GPS system, i.e. the satellite constellation and the control earth stations, has its own time clock, which derives from the composite clock of the UTC (USNO) [2] and all satellite internal clocks. The master control earth station gathers time data from the satellites in service, calculates the timing errors for each satellite clock, and uploads the appropriate clock corrections to each satellite.

Standard Positioning System (SPS) provided from GPS system is a service available to all the users, free of charge, on a continuous, worldwide basis. Each satellite periodically broadcasts navigation messages, which contain information accurate time information apart of position data. The achieved accuracy of the time information is 167ns. However, the timing information received by end-users may be worse as low cost GPS receivers may impose time accuracies.

2.4 Presenting and Analysing Monitoring Data

As mentioned previously, SPs should publish results collected from their monitoring infrastructure so that the customers can verify the compliance of the network behaviour to the agreed SLAs. On one hand, the ability of the SP to publish monitoring information may influence the SLAs that can be supported. For example, if the SP is not able to provide monitoring information with less than 10ms time precision then SLAs that guarantee maximum jitter of less than 10ms may not be supported. On the other hand, SLA agreements define the requirements when presenting the monitoring data. For example, if the SLA agreement refers a delay percentile evaluated every five minute period then the monitoring plots should be created accordingly.

In the following paragraphs some suggestions are provided on how to analyse and present monitoring data in a coherent way for the SPs and its customers:

- Measurement information should be available for each monitoring node in an administrative domain. Otherwise, it may be impossible for the customers to evaluate the overall performance of the network and the QoS provisioning compliance to the SLAs for each pair of destinations.
- Real-time and historical data from the measurements should be available to the customers via Web or other interface. In addition, data plots of the measured parameters are required as plots provide an easy way of disseminating available information in a compact and comprehensive way for customers Furthermore, processing of collected data in almost real time is needed according to the SLAs agreements. For example, calculations of median and 95% percentiles of delay during recent period for a specific window frame are usually embedded in the SLAs. Data plots should be created according to these requirements. Finally, raw data files and a set of scripts for processing collected data should be provided by the SP to the customers in order to perform more extended analysis of the network behaviour.
- Tools for generating reports on demand should also be developed. The SP may provide to the end-users Web interfaces where they are allowed to apply various processing functions to the collected monitoring data. For example, an end-used may identify two monitoring nodes of the network; specify the measurement period (e.g. last two weeks); and ask the mean jitter value for each one-hour-window period.
- SPs may give to the end-users a limited access to their monitoring infrastructure. In this case, end-users are allowed to perform their own network measurement tests according to their needs. An example of such service is “Iperf Servers” [3] [4]. By

enabling such services, it will be more easily for the end-users to understand the QoS service provisioning of the network and apply for a supported service.

3 QoS monitoring Infrastructure

There are two different approaches for building a monitoring infrastructure in a large-scale network. The first one is based on common open-source or freeware software management tools (SMTs) [5] [6] that can evaluate the performance of a network. The second approach is based on a commercial product [7] [8] [9] [10] [11], [23] which is specially designed and developed for auditing the performance of IP networks.

SMTs or commercial products usually consist of two entities: a sender and a receiver. The sender injects traffic into the network according to a predefined traffic contract (bit rate, packet size, DSCP/IP ToS code, inter-packet delay etc.) The receiver collects monitoring traffic from one or more senders and stores it in text or binary files. Collected data are processed either locally or sent to a specific purpose central server.

Each of the two approaches has its advantages and disadvantages, which significantly differ in terms of complexity, scalability, installation and operation cost, availability, etc. In the following paragraphs, we present some of these characteristics in more detail.

3.1 QoS monitoring based on SMTs

Advantages

- Open architecture: SMTs usually consist of open-source or freeware software that is freely distributed at the Internet. Different software groups continuously develop measurement tools with diverse features and various APIs that allow easy manipulation of monitoring data. Consequently, SPs may define various measurement scenarios according to their network topology and characteristics. Also, SPs may utilise more than one performance tools to build a QoS measurement infrastructure. In addition, the SMTs software code may be changed or extended in order to be adapted to the specific requirements of a SPs network. For example, in a network that supports multiple CoSs, the monitoring infrastructure is adjusted to perform different test for each CoS.
- Distributed system: SMTs follow the client – server model, where the sender injects traffic to the network, which is afterwards collected by the receiver. Therefore, SMTs are easily deployed in distributed monitoring infrastructures where the collected data is stored and analysed at the receiving monitoring node. If a centralised model is selected by the SP, additional software is required for integrating the monitoring entities.
- Easy manipulation of data: SMTs provide common APIs for collecting measurement data in text-based or binary format. Therefore, real-time or post-possessed analysis of collected data is feasible with diverse software programs. Note that when data are analysed in real time, SMTs may be used continuously evaluating the network performance. However, as SMTs are not commercial products, SPs usually have to develop tools for processing and presenting results in a consistent fashion.
- Low implementation cost: SMTs may be used for building inexpensive monitoring infrastructure. The software is freely distributed and supported in various operating platforms. Also, the sending or receiving entities of an SMT require very few resources, in terms of CPU and memory, from a common network server. Finally, the NTP protocol may provide a cheap solution for distribution of time information amongst the monitoring nodes. In that case accurate time synchronisation is not required among the end nodes.

- End-to-end monitoring architecture: Customers may easily expand the monitoring infrastructure of their upstream provider inside their network by using similar SMTs tools, as the fundamental principles remain the same. Consequently, SMTs may be used for building an end-to-end (non-mesh) monitoring infrastructure in a large-scale environment composed by different integrated administrative domains. Attention should be given, however, to the requirements of the monitoring infrastructure laced closer to the end-users.

Disadvantages

- Cumbersome deployment: As SMTs are simple network tools, SPs should define the appropriate measurement scenarios, which is a complex and challenging process. Obviously, performed measurements should be designed according to the standards provided by the standard bodies, such as IETF IP Performance Metrics Working Group [12]. In succession, the monitoring infrastructure should be deployed and the software be tested. Furthermore, initial measurement results should be crosschecked in order to detect possible software bugs and measurement errors (calibration phase). Significant effort may also be required for analysing collected data from multiple receiver entities. This requires software developers to have very good understanding of the measure QoS performance parameter definitions and good knowledge of the network topology. Also presenting data in a unified environment using automatically generated graphs may be difficult task. Mechanisms for storing collected data in a central server are also required. All the above subsystems have to be developed and deployed by the SP.
- Security vulnerabilities: Most of the SMTs do not provide security features. Consequently, collected measurement data may be tampered by a malicious user that sends fake traffic to the STM receiver. Consequently, changes in the SMTs software may be required in order to be supported mainly authentication and authorisation features. Furthermore, misbehaving monitoring nodes may become potentially harmful to the network.

3.2 QoS Monitoring based on Commercial Products

Advantages

- Ready for service product: Usually, by the time a product is released, it incorporates functions that can satisfy most of the customer needs. Therefore, a SP that deploys a monitoring infrastructure based on a commercial product needs minimum effort before being able to provide the new service. For example, TTM is developed and supported by RIPE as commercial product. The RIPE TTM development team deals with the definition of measurement scenarios, the definition of monitoring procedures, the collection and analysis of measurement data, the presentation of data in graphical way, the storage of data in databases. The software team is also responsible for fixing software bugs and upgrading the TTM software.
- Accurate measurements: Commercial products, such as RIPE TTMs, include a GPS receiver for providing time reference information. Consequently, the accuracy of the measurements is extremely precise and definitely adequate for QoS monitoring in IP networks. Commercial products without accurate external clock reference are usually not suited for SP networks.

Disadvantages

- Closed architecture: The customer that uses commercial products for network auditing has limited ability to change the implemented monitoring model and parameters. For example, in most of the cases the measurement scenarios are predefined, monitoring traffic schematics and parameters can not easily be changed by the customer, analysis procedures are predefined, etc. Generally, the customer has to choose to deploy one of the available commercial products that more closely satisfy its needs. If a feature is not supported by a product, the customer has to wait until it is available. This imposes significant delay in the deployment of new services and requires additional cost for the customer.

- Scaling - Centralised architecture: Usually commercial products forward all of their collected data to central servers, dedicated for analysing and storing data. Scaling issues becomes an issue if too many monitoring nodes are installed in the network, as a significant amount of traffic should be injected into the network and forwarded to the central servers.
- High installation cost: Commercial products usually require a GPS receiver to be purchased and installed for providing time synchronization between the monitoring nodes. However, the installation cost of antennas for a large-scale infrastructure is not negligible. Furthermore, many installation problems have to be overcome especially if the equipment is not installed in customers premises.

Table 1 presents the advantages and disadvantages of the two different approaches for building a monitoring infrastructure in a large-scale network. SMTs vs. commercial product.

Table 1: Comparing monitoring infrastructures based on STMs vs commercial products

	Advantages	*Disadvantages*
SMTs	- Open architecture - Distributed system - Ease in manipulation of data - Low implementation cost - Easily expanded to end-users	- Cumbersome deployment - Security vulnerabilities
Commercial Products	- Ready for service product - Accurate measurements	- Close architecture - Scaling - centralised architecture - High installation cost

4. GEANT Service Specifications

Networks such as the pan-European Gigabit Research and Education Network (GÉANT) [13] is designed to support the following Class of Services (CoSs) [14]:

- Best Effort (BE)
- IP Premium

In brief, BE is a service that does not guarantee any QoS to the network traffic and is suited to non real time applications such as Web browsers or FTP clients. IP Premium, on the contrary, is a service that provides bounded delay and jitter guarantees and negligible packet loss to the conforming traffic. Therefore, it is a service suited for QoS sensitive applications, such as audio/video streaming, IP telephony etc.

In the future, backbone networks may support another CoS, IP+ [14], which is a less strict service than IP Premium, as packet losses may occur during network congestion periods.

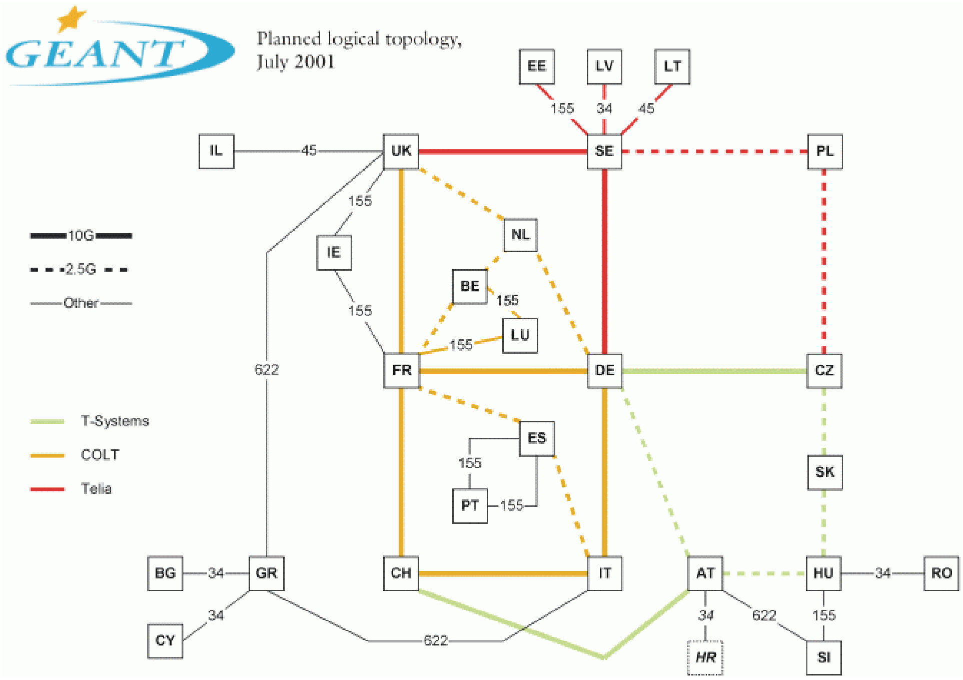

Figure 2: GÉANT logical topology map (06/2001) as an example of backbone network supporting several classes of service.

4.1 SLA Parameters for CoS

In order to validate the conformance of the network behaviour to the agreed SLAs for IP Premium and BE traffic, different parameters should be continuously monitored. According to SEQUIN document [14], IP Premium service performance can be assessed by monitoring the following four network parameters:

- One-way packet delay
- One way IP packet delay variation (IPDV) – "jitter"
- Packet losses
- Bandwidth

Extended definitions for the above metrics could be found in the document [14].

BE service, as mentioned previously, does not provide any QoS guarantees to traffic. However, for operational and management purposes, the following parameters are suggested to be monitored:

- Delay
- Packet losses

It should be mentioned that monitoring network behaviour for IP Premium traffic is more complex and cumbersome task than for BE traffic. As BE CoS does not provide any guarantees to traffic, there is no need for accurate measurements of the network parameters. On the contrary, IP Premium service requires accurate clock synchronisation of the monitoring infrastructure so that correct delay measurement statistics can be collected.

4.2 Monitoring Scope in GEANT

As noted in section 1, an SP is only responsible for the part of the connection between its customers that belongs to the SP's administrative domain. Consequently, monitoring infrastructure in a network only audits the network performance within its own boundaries while the customers are the NRENs and not the "real" end-users.

A backbone network monitoring infrastructure will comply with the 2nd topology presented in section 1. Monitoring nodes will be installed at each aggregation PoP and the customer access PoPs. An example can be found in the logical topology map of GEANT in Figure 2.

4.3 Synchronisation

As mentioned in section 1, monitoring nodes should be synchronised in order to perform accurate measurement of time-related parameters. Synchronisation can be achieved either by using NTP protocol or by high precision external clocks such as GPS receivers.

One-way delay is the only network parameter that needs network synchronisation among the monitoring nodes. One-way delay consists of propagation delay, transmission delay, queuing delay and switching delay. In the first operational period of a new generation backbone network, it is expected that switching[1] and queue delays in the backbone routers will be negligible. Backbone links such as STM-64/PoS or STM-16/PoS, the transmission delays are also negligible[2]. Therefore, propagation delay is the most significant factor in the one-way delay that packet encounters in the network. Propagation delay of 1msec per 150 Km is used in the following calculations. Jitter is another significant time-sensitive network parameter. However, jitter does not require accurate synchronisation between the monitoring nodes as the jitter measurements evaluate the relative delay between successive packets. However, only high clock resolution in the receiving node is essential.

Synchronisation problems will be minimised if the network operators choose to deploy monitoring a infrastructure based on commercial products, as GPS receivers are used by default for synchronisation. Otherwise, if they choose to use SMTs, then synchronisation of the monitoring nodes will be an issue. In the latter case, synchronisation between the monitoring nodes should be based on the NTP protocol. It is difficult to estimate the time accuracy between two nodes that can be achieved with NTP protocol[3]. NTP accuracy is affected by the network topology and the network congestion in both directions of the path between the client and the NTP server. Improvements in synchronisation could be achieved by giving high priority to NTP traffic, such as by marking NTP traffic as IP Premium traffic. This method will minimise the asymmetric delay that NTP traffic encounters traversing from the client to the NTP server and backwards. Furthermore, software patches for NTP software and operating system [15], will also improve the achieved synchronisation of monitoring nodes. However, in any case, a small number of primary servers with GPS are required for synchronising all the monitoring nodes with acceptable level of accuracy. The number and the exact location of the GPS receivers depend on the network topology and the SLAs that will be provided to the customers.

It is envisaged that the end-to-end delay that packets will encounter in the network on a European scale will be a multiple 10s of msecs. This was confirmed during the SEQUIN

[1] Switching delay in routers is estimated to range from 60usec to 200usec.

[2] Transmission delay of 512bytes packet over 2.5Gbps link is approximately 1,6 usec.

[3] At the moment of writing, a TF-NGN [20] activity is in progress and tries to evaluate the accuracy of synchronisation with NTP protocol.

H.323 tests where the maximum measured round trip delay was around 50msec[4]. These values are justified because the distance between two end-users from different countries is usually hundreds of Km and the transmission speed is continuously decreasing as we move closer to the end-users. Intuitively, it can be expected that it is not mandatory to install a GPS receiver for each monitoring node in the network. Of course, conclusions that are more accurate will be drawn in the future, after the initial implementation of the services.

5 Monitoring in GÉANT – SMTs solution

The software tool that is suggested to be chosen for building the monitoring infrastructure is the **Real-time UDP Data Emitter** (RUDE) and **Collector for RUDE** (CRUDE) [6]. RUDE is a simple but flexible program that generates UDP traffic, while CRUDE is used to receive and log monitoring traffic.

RUDE/CRUDE software may be used for monitoring delay, jitter and packet losses. It can generate two different traffic patterns; constant bit rate traffic or trace-based traffic. In the first traffic pattern, it is possible to define the packet size and the frequency according to which the packets are injected into the network. In the second traffic pattern, a text-file provides a simulated trace for transmission of different packets with time resolution of one microsecond. For each packet injected into the network, RUDE sets “tx timestamp” and “flow id” values. In the receiving end, CRUDE adds “rx timestamp” in each packet. Also, for each packet arrived CRUDE exports a tuple containing the packets flow id, packet sequence number, source and destination IP address, source and destination UDP port, time of transmission and time of arrival, and its size. Exported information may be stored into text or binary files.

5.1 Monitoring Scenario

- *Monitoring Matrix:* RUDE/CRUDE servers (group A) should be installed in all the backbone PoPs. These servers should be connected in a mesh topology, i.e. each of them will exchange monitoring traffic with all the others. Group A of servers will be used for evaluating the backbone network performance and QoS service provisioning. Furthermore, another group of RUDE/CRUDE servers (Group B) may also be installed in each customer's access PoP. These servers will exchange traffic only with the corresponding servers of Group A and they will be used for evaluating the performance at the customer's access links.
- *Traffic characteristics:* RUDE/CRUDE servers are suggested to generate monitoring traffic with the following characteristics:
  - Bit rate: VBR
  - Packet size: 64 bytes and 1500 bytes
  - Pattern: Every four 64-byte packets a 1500-byte packet follows
  - DSCP value: Set to 46

  Monitoring traffic that consists of single sized packets is enough for measuring delay, jitter and packet loss in the network. However, for troubleshooting purposes, a more complex pattern is suggested. As packet reordering occurs in the networks, two different packet sizes are used[5]. Also, comparing delay and jitter measurements per packet size could also reveal network misbehaviour. Finally, there are no specific guidelines on how often monitoring packets should be ingested in the network in order to accurately audit

[4] During the H.323 [21] tests, RTT of 219 msec was measured between GRNET and RUS. These results where obtained during a severe DOS attack to GRNET network.

[5] SEQUIN tests revealed that packet reordering occurs among packets of different sizes because backbone routers “treat” small packets differently than big ones.

the network behaviour while using minimum network resources. For the time being, one packet per second is suggested to be send for each Mbps of IP Premium service delivered between two backbone PoPs. For example, if the SLAs refer to 5 Mbps of IP Premium traffic between PoP A and PoP B, then 5 packets per second should be sent from the monitoring servers placed in the two aforementioned PoPs.

- *Overhead:* The monitoring matrix described previously defines multiple flows between RUDE/CRUDE servers of A group. Therefore, monitoring traffic may potentially impose significant load to some backbone links and thus affecting the provided services. Therefore, it is suggested that the upper limit of monitoring traffic over a link is less than 1% of the link capacity. For example, in STM-4 links monitoring traffic should never exceed 6 Mbps. Note that 30Mbps of IP Premium traffic is monitored by using around 84kbps traffic, according to the aforementioned traffic pattern.

5.2 Alarm system

Collected monitoring data should be analysed on a per flow basis. Median and percentile values for each monitored parameter, e.g. delay or jitter, should be evaluated continuously for the most recent period. A five-minute time window is suggested for evaluating the network performance in terms of mean packet delay or jitter. Obviously, the time window should change according to the signed SLAs/SLSs. Furthermore, the monitoring nodes should compare the measured parameters with the SLSs values agreed in the SLA agreement. If the SLA is violated then the monitoring node should trigger an alarm. Appropriate notifications should be sent to the provider NOC and to the related customers. Also, the monitoring flows may automatically be modified according to a predefined scenario. These means that in case of a SLA violation, the monitoring flows may be altered in order to provide more information to the network administrators. This procedure should be carefully designed as increasing the monitoring flows bit rate may additionally affect the network performance. For example, increasing the bandwidth of a monitoring flow when there is network congestion will further congest the network. In addition, for troubleshooting purposes, the recent network measured values should also be compared with the long-term values. The suggested period for the long term values should be 2 hours. If the evaluated results are outside the expected range, a notification should be sent to the corresponding NOC.

5.3 Other Aspects

- *Archiving:* All data collected by the monitoring nodes could either be stored locally in the node or be sent over the network to specific archiving servers. In the distributed scenario analysis, presentation and storage of data is easily implemented. On the contrary, if monitored data is sent over the network and stored at central servers scalability problems may arise. A distributed archiving method is usually preferable.
- *Security:* Performing QoS measurements can be easily tampered by malicious servers that inject artificial traffic to the network, which may be accepted as legitimate by official servers. Therefore, authentication techniques should be used where appropriate to guard against malicious traffic. Unfortunately, current version of RUDE/CRUDE does implement any authentication/authorization mechanisms. However, a malicious user has to know the flow id, sender/ receiver UDP port in order to tamper monitoring data.
- *Presenting Data:* There are various tools for presenting measurements results at each monitoring node. Common tools such as [16] [17] [18] could easily create data plots by accessing output CRUDE text files. Also, simple scripts, such as scripts developed during the SEQUIN H.323 tests [22], may easily used for analysing raw data files in order to calculate median and percentile values for packet delay and jitter.

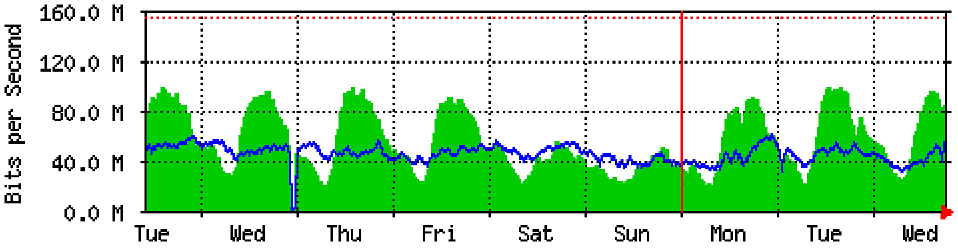

Figure 3: Data plot created with MRTG tool

- *Additional tests:* Before starting a monitoring test it is significant to register the network path that traffic is going to follow. Therefore, "traceroute" must be initially executed in order to verify that the traffic is routed as intended. Also, it is suggested a modified version of "traceroute" to be used, which can validate that the packet DSCP value is not altered in the network, probably due to mis-configured routers. Monitoring tests will be continuously performed, "traceroute" may be performed once a day.

6 Monitoring – RIPE TTM test-boxes

RIPE is providing a service called *Test Traffic Measurements (TTM)* [7], which aims to do independent measurements of connectivity parameters, such as delay, packet losses and routing-vectors.

The TTM service utilise a dedicated measurement infrastructure, composed by *test-boxes*, which performs active measurements. The test-boxes are rack mounted PCs[6] that run a modified operating system developed by RIPE software team. For providing accurate time-related measurements, the test-boxes use a GPS clock as an external time source. Thus, the accuracy of the measurements is of the order of tens of microseconds.

Data is collected daily from all the test-boxes and sent to a central server, which is dedicated to analyzing and presenting the results of the monitoring tests. The server combines data from each pair of test-boxes and converts it into an easily processed data format. Finally, the server processes the collected data and generates different sets of plots for all the measured parameters. Data is afterwards archived for future analysis.

The end-users do not need to have any access to the hardware or software of the test-boxes. They are centrally managed from the (central) server, which forwards to them the appropriate configuration and controls the monitoring procedures. Therefore, the TTM service requires minimal maintenance effort from the RIPE support team.

6.1 Monitoring Scenario

- *Monitoring matrix:* TTM test-boxes should be installed at the backbone network PoPs (Group A) and at each customer access PoP (Group B). Due to the TTM's service implementation constrain, all test-boxes (in Groups A and B) will start exchanging monitoring traffic between each other and with all the other test-boxes already in service[7]. Obviously, as test-boxes are fully connected in mesh topology, scaling problems may arise. The amount of data that should be sent over the network and centrally processed is increased quadratically to the number of test-boxes. Nevertheless, without any change to the current RIPE central server's set up, up to 200 test-boxes could currently be supported. In addition, in the next software releases, peering groups of TTM test-boxes, i.e. test-boxes that exchange traffic only with test-boxes that belong to the same peering

[6] Some specifications of the hardware are Celeron 600MHz, 10/100 Base-T port, GPS interface
[7] In April of 2001, more that 60 TTM were in service.

group, will be supported. This feature will decrease the amount of traffic injected into the network and the collected information sent to the central servers. Obviously, this new capability of the test-boxes will solve most of the scalability problems of RIPE service[8].

- *Traffic pattern:* Each TTM test-box injects two 100-byte packets per minute per each destination box. The current software release of the test-boxes allows the modification of monitoring traffic rate, while in the next releases the packet size will also be configurable. Unfortunately, all parameters are common for all the test-boxes in service and, consequently, the RIPE monitoring infrastructure is not easily adjusted to the needs. Furthermore, TTM test-boxes are currently DSCP/IP Precedence unaware. In other words, the monitoring traffic injected into the network is marked as best-effort traffic. In the future releases, the TTM test-boxes will be able to perform measurements for different CoSs by setting the appropriate DSCP value in the IP headers. Therefore, additional flows would have to be established between the test-boxes per each CoS supported. As a conclusion, the current traffic pattern of test-boxes may not be appropriate for monitoring IP Premium traffic in networks such as GÉANT. However, future software releases of test-boxes will overcome much of the aforementioned problems.
- *Overhead:* The monitoring matrix described previously defines multiple flows between test-boxes in service. However, as only two packets per second per destination are injected into the network, it is not expected that test-boxes can impose significant load to the backbone links. Also, collected data from the test-boxes are sent to central servers during off-peak hours once per day without possibly affecting the network performance.

6.2 Alarm system

Currently, RIPE test-boxes have very limited alarm functionality. Each test-box maintains a long-term median and percentile delay values for all the test-boxes in service. It constantly compares the values for the most recent measurements against the results taken over a long period. If a recent measurement is outside an expected value range, the test-box triggers an alarm by sending a message to the corresponding NOC of the test-box. In the following software releases, alarms may be generated according to predefined fixed thresholds. Therefore, in the future the alarm thresholds will be easily set according to the SLAs parameters.

6.3 Other Aspects

- *Archiving:* Collected monitoring data is sent over the network to a central server for storage. With 60 test-boxes currently in service, 20 Mbytes per test-box are sent over the network during off-peak hours. As a result, 1,2 Gbytes of monitoring data are stored in the central server per day. These numbers are increased by a factor of $O(N^2)$, where N is the number of test-boxes.

[8] Note that intra-domain measurements are outside the scope of the work done by RIPE. Therefore, peering groups in the same network may not be supported by RIPE service. This makes difficult to extend the monitoring infrastructure inside the end-users domains.

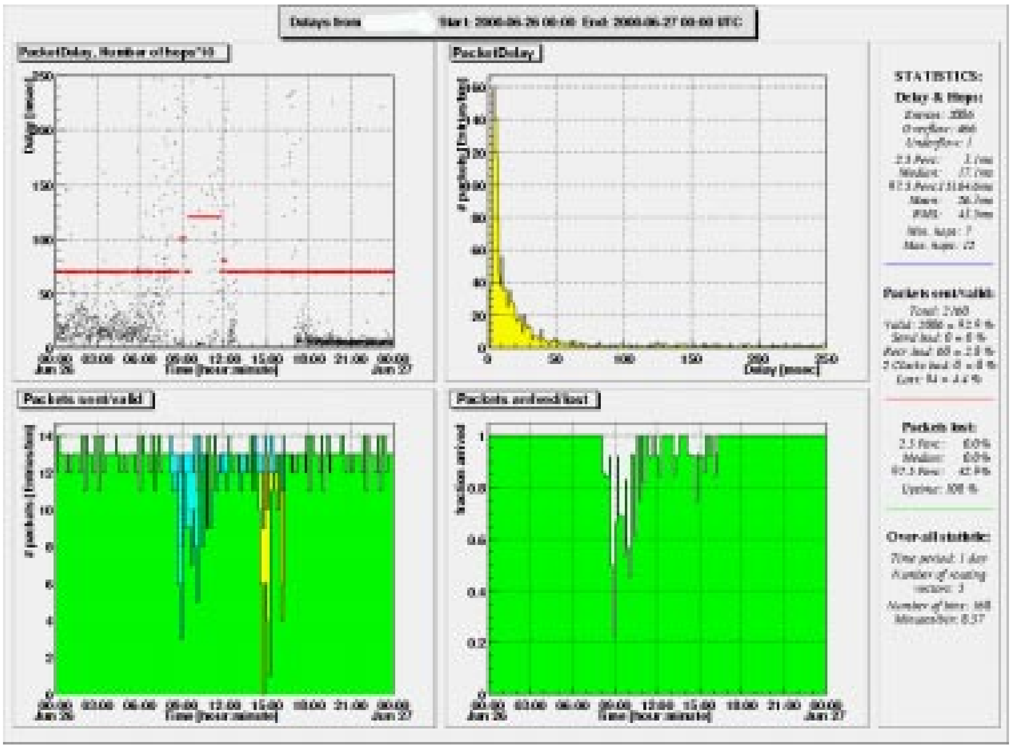

Figure 4: Delay plot from RIPE TTM service

- *Security:* Security is not an issue for TTM service as the operating system of the test-boxes implements all the necessary security mechanisms.
- *Presenting Data:* A well-designed web GUI provides a visual presentation of the monitored network parameters for each pair of test-boxes. Each measured parameter is presented with three sets of four plots. For example, the delay plots in Figure 4 provide the delay measurements as a function of time, the distribution of delays during a specific period, the experimental conditions and the packet losses. Furthermore, on the right side of the plots, various statistical information related to the measured parameters are given, such as the median or the percentile values of delay. In general, the TTM service plots make available information for the one-way delay, the one-way delay variation and the jitter that monitoring traffic encounters in the network. It should be noted that each set of plots derives from the collected data in 24-hour, 7-day or 30-day periods. Also, a test-box user may generate plots for arbitrary time intervals. A detailed description of the provided plots could be found in [19].

Finally, the following Table presents in a tabular format the most significant advantages and disadvantages of deploying a monitoring infrastructure with RUDE / CRUDE software or with RIPE TTM test-boxes.

Table 2: Comparison of RUDE/CRUDE and RIPE TTM test-boxes monitoring infrastructures

	RUDE/CRUDE	*RIPE TTM Test Boxes*
Open architecture	Yes	No
Distributed system	Yes	No (Centralised architecture)
Monitoring matrix	Flexible (Partial or full mesh)	Full mesh (Partial mesh in the future)
Implementation cost	Low	Medium
Ready for service product	No (Software should be developed by scratch)	No (Required functionality is still missing)
Cumbersome deployment	Yes (Subsystems have to be developed)	No
Ease manipulation of data	No	Yes
Expanded to end-users	Yes	No
Accurate measurements	Yes, if required	Yes
Scaling problems	Minor	Major
Alarm support	No, but easily implemented.	Yes, but minimum functionality
Security vulnerabilities	Yes	No

7 Conclusions

This document specifies the requirements for implementing a monitoring infrastructure in a large-scale network, such as GÉANT. Building a monitoring infrastructure is a complex and cumbersome task, but QoS service provisioning and SLAs verification necessitate the deployment of such infrastructure SPs.

In the previous sections, issues regarding monitoring scope, measurements methods and synchronisation were presented and analysed. Also, a comparison was made between SMTs and commercial products that are appropriate for monitoring purposes. In addition, two suggested solutions for building the monitoring infrastructure were presented based on RUDE/CRUDE software and RIPE TTM test-boxes. Each of the two solutions exhibits different advantages or disadvantages. On one hand, the embedded functionality of RIPE TTM test-boxes will fulfill most of the network monitoring requirements, especially with the following software releases. On the other hand, RUDE/CRUDE software could be used for building an open-architecture monitoring infrastructure, which is easily extended to the end-uses.

It is suggested that a backbone network should start building its monitoring infrastructure based on RUDE/CRUDE software. For synchronizing monitoring nodes, it is proposed the NTP protocol to be used. In some PoPs, however, GPS receivers will be needed for providing accurate time data to the NTP primary servers. After the initial implementation of the IP Premium service in GÉANT, more accurate conclusions will be drawn. If it turns out that the achieved accuracy based on the NTP protocol is insufficient for monitoring one-way delay, the backbone networks may choose to synchronize the monitoring nodes with GPS receivers or it may reconsider using RIPE TTM service.

8 References

[1] D. Mills, “Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI”, RFC2030, October 1996.
[2] SEQUIN internal report, “Presentation and Comparison Study for NTP and GPS”, March 2001.
[3]: GPN Iperf Server, http://noc.greatplains.net/measurement/iperf.php.
[4]: Iperf Server, http://nic-mn.northernlights.gigapop.net/noc/measurement/iperf.php.
[5] Corporative Association for Internet Data Analysis (CAIDA) http://www.caida.org/.
[6] Real-time UDP Data Emitter and Collector for RUDE, http://www.atm.tut.fi/rude/.
[7] RIPE Test Traffic Measurements, http://www.ripe.net/test-traffic.
[8] Chariot, http://www.netiq.com/
[9]: Brix Networksr, http://www.brixnet.com/
[10]: Matrix, http://www.matrixnetsystems.com/
[11]: Service Assurance Agent, http://www.cisco.com/
[12]: IP Performance Metrics (ippm), http://www.ietf.org/html.charters/ippm-charter.html
[13] GÉANT, http://www.dante.net/geant/index.html
[14]: M. Campanella et. al., “D2.1 SEQUIN Quality of Service Definition”, April 2001
[15] D.Mills, P.Kamp, “The nanokernel”, Proc. Precision Time and Time Interval Applications and Planning Meeting, Reston VA, November 2000
[16] Cricket, http://cricket.sourceforge.net/
[17] Multi Router Traffic Grapher, http://www.mrtg.org/
[18] CESNET projects, http://www.cesnet.cz/english/project/
[19] Test Traffic Project Results, http://www.ripe.net/test-traffic/General/tt_matrix.html
[20]: Task Engineering Next Generation Networking, http://www.terena.nl/task-forces/tf-ng
[21] SEQUIN internal report, “H323 Test Scenario”, November 2001.
[22] Simon Leinen, SEQUIN IPDV script, http://results.sequin.switch.ch/raw/ipdv.tar.gz
[23]: Ipanema, http://www.ipanematech.com/