26-05-2011

Deliverable DN4.3.2: Proof of Concept. Moving Data with Multicast, EUMETSAT/DANTE Collaboration

Deliverable DN4.3.2 v1.0

Contractual Date:	31-03-2011
Actual Date:	26-05-2011
Grant Agreement No.:	238875
Activity:	NA4
Task Item:	T3
Nature of Deliverable:	R (Report)
Dissemination Level:	PU (Public)
Lead Partner:	DANTE
Document Code:	GN3-10-247
Authors:	R. Hughes-Jones (DANTE), M. Sorrentino (DANTE), D. Vicinanza (DANTE), Shannon Milsom (DANTE)

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7 2007-2013) under Grant Agreement No. 238875 (GÉANT).

Abstract

This deliverable reports on the discussions with EUMETSAT to understand their network requirements for potential use of the academic network formed by GÉANT and the NRENs, and the joint evaluation of different connectivity options. It outlines the design and implementation of a collaborative Proof of Concept to demonstrate moving data over GÉANT using multiple simultaneous high bandwidth multicast flows. It briefly describes the results and analysis of performance measurements of the network infrastructure as well as that obtained when moving representative datasets from EUMETSAT using the TelliCast data moving software.

Table of Contents

Table of Figures

Table of Tables

Executive Summary

This Deliverable describes the collaborative work between EUMETSAT and GÉANT. Tests were designed jointly to investigate the IP multicast ability over the GÉANT network using the TelliCast software[1] licensed by EUMETSAT [RD5]. This document records the results from the tests performed in accordance with *Multicast over GÉANT Test Cases, EUM/OPS/TEN/09/4685* [RD2].

This document provides information on the experience and performance of operating high bandwidth multicast flows. Detailed descriptions of the tests and their configuration are in separate documentation [EUMPoC], which provides a technical information resource in conjunction with this Deliverable. It is expected that this documentation set will prove useful for both users with operational responsibility and for researchers interested in the behaviour of multicast flows.

EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) operates a number of satellites collecting earth observation data and images; they process the information and disseminate the resulting data products to a wide number of users and the research community. EUMETSAT expressed interest in using the academic network after meeting GÉANT at the GEO-V Plenary meeting held in Bucharest in November 2008. DANTE and DFN followed this up via meetings with EUMETSAT in Darmstadt and Cambridge, defined the user's requirements, explored operational aspects of the GÉANT service area, advised on GÉANT's connectivity services and discussed possible architectural solutions, including four different connection scenarios. From these discussions it was decided to embark on a technical collaborative Proof of Concept (PoC) to demonstrate moving data over GÉANT using multiple simultaneous high bandwidth multicast flows.

The objectives of the joint PoC were to:

- Extend the technical knowledge and operational experience in using the GÉANT network to reliably support high bandwidth multicast applications at the required data rates.
- Demonstrate, using EUMETSAT hardware and software located in GÉANT PoPs, that the multicast flows could successfully use the academic network to transmit data files over the TelliCast software to geographically dispersed locations.

[1] TelliCast is commercial software provided by Newtec for moving data using IP multicast. See [RD5] for more information.

Four sets of tests were planned and implemented:

- Initial tests using the GÉANT multicast beacons to explore the performance of multicast flows over GÉANT.
- Tests of the EUMETSAT hardware and software directly connected in a laboratory environment.
- Tests of unicast and multicast flows using the EUMETSAT hardware installed in the GÉANT PoPs.
- Tests using the TelliCast software to move EUMETSAT data files over multicast flows.

The PoC was successfully completed in February 2011. It proved to be an excellent collaboration on all levels including:

- Open exchange of technical information.
- DANTE supplying network resources and network testing tools.
- EUMETSAT supplying the hardware and the commercial TelliCast [RD5] software used to move their data.
- Performing tests jointly; GÉANT personnel were able to get information from the routers during the tests and investigate the TelliCast traffic patterns.
- Joint analysis of the data and discussion of the conclusions.

The initial tests focused on verifying that the environment to be used in the tests was correctly configured and the performance of the GÉANT network was understood. While these tests showed that care needed to be taken with the smaller switches in the point of presence (PoP) LAN (which did not always appear to operate as expected), the operation and performance of the GÉANT Juniper routers and the trunk links was excellent. Once the multicast distribution tree had been set up, no packet loss was experienced under normal conditions. In addition, the GÉANT network easily supported simultaneous multicast flows using different multicast groups or port numbers.

The back-to-back laboratory tests (laptops directly connected together) were vital in providing the optimum conditions to be used with the EUMETSAT hardware to ensure that it was the behaviour of the network which was being measured, <u>not</u> the operation of the end hosts. They demonstrated that the test setup with laptops running TelliCast had the required power and performance to provide lossless dissemination up to 120 Mb/s and, therefore, provided a base-line set of data of the network and data-moving performance. This data allowed comparison with the PoC tests made over the GÉANT network.

For security purposes, a rate-limiter was imposed for the EUMETSAT traffic. After the laptops were installed in the GÉANT PoPs in London, Amsterdam and Prague, the network performance was first measured without any rate-limiter using unicast flows (using the *udpmon* program [RD4]), then multicast flows (using the *mcastmon* program [RD10]). Without the rate-limiter, the network was found to operate correctly with both UDP (User Datagram Protocol) and multicast packets with a throughput of 1 Gb, no packet loss and 2µ to 4 µs of inter-packet jitter in full agreement with the laboratory tests. With the rate-limiter enabled, the flows were correctly limited to the 300 Mb/s that was set.

The multicast data flows generated by the TelliCast software were analysed using software tools developed for this PoC project (see *mcastmon* [RD10]). Analysis of the multicast data flows produced by the TelliCast software showed that every 20 ms the software generated bursts of packets at the 1 Gb connection speed of the host PC. The number of packets generated in each burst was correctly calculated to provide the average rates specified for all the configured flows. However, for data rates above 50 Mb/s, the bursty nature of the traffic triggered the rate-limiters agreed for the PoC (set at 300 Mb/s) when they were enabled. This contributed to loss of the data files. This behaviour of the TelliCast software could be important if the destination sites were connected with low bandwidth or rate-limited links. This demonstrated that it is important to consider the peak rates and not just the average flow rates when planning a network installation.

Summary of Using TelliCast Software to Move EUMETSAT Data Files

At TelliCast data rates of 8 Mb/s and 50 Mb/s, the average reception performance was very close to 100%. The only packet losses recorded by the TelliCast software client at these data rates were caused by a link outage on the Frankfurt-Prague link, which was due to emergency maintenance to replace an amplifier board on that link, and by one other event where there was exceptionally high network load on the London-Amsterdam link. GÉANT traffic monitoring confirmed the losses on the GÉANT trunks at these times.

At the data rate of 120 Mb/s, the average reception performance dropped below 100%. For the majority of the tests, the TelliCast application reported that there were no significant losses. However, when packet losses occurred, they did so in batches. Occasional *database full* errors were reported by the TelliCast client, which resulted in data file loss. These errors could occur if the space in the application database was not sufficient to hold details of the files currently in transit. Nearly all of the file losses (that could not be attributed to network-related events) coincided with database errors in the TelliCast client. This database error behaviour is not yet understood. However, in most cases no losses were actually recorded on GÉANT trunks at these times.

Two factors contributed to the losses observed on the London-Amsterdam trunk at the 120 Mb/s data rate: the planned upgrade of the London-Amsterdam trunk IP route from 10 Gb to 20 Gb, which (for other technical reasons) was delayed until after the PoC; and the LHC traffic patterns, which had changed in an unforeseen manner. These two factors together with the bursty nature of the TelliCast traffic caused packet loss on the London-Amsterdam trunk that was reported by the TelliCast software. With current data moving tools, such as gridftp and TelliCast, end users are becoming more aware of the performance of the underlying networks.

One of the other network related events that caused packet loss on the multicast flows was the operation of the perfSONAR BWCTL bandwidth throughput measurements between the monitoring points located on the LANs of the GÉANT PoPs. This suggests that great consideration should be given to the frequency of running intrusive throughput monitoring tests on live networks where users are moving bulk data or have time critical requirements. Use of light weight estimators of performance, such as observations of packet loss and round-trip delay from the ping tool or the Hades one-way delay tool (which are also part of perfSONAR), may be more appropriate.

EUMETSAT may consider the following for future work, with which GÉANT would assist if requested.

- Perform an in-depth analysis of the configuration of the GÉANT network equipment and the TelliCast software to optimise their interaction for this type of multicast traffic.
- Perform an in-depth analysis of the actions of the TelliCast server application to better understand the unexpected data bursting that was observed in the TelliCast server application on the network during the tests.
- Assess the impact of reconfiguring the test laptops for use on the GÉANT network.
- Assess the impact of the long uptimes on the test laptops.
- Perform additional laboratory tests to better understand the apparent discrepancy between the results of the original laboratory tests and those performed on the GÉANT network.

This PoC successfully demonstrated the use of IP multicast on the GÉANT backbone. It showed that GÉANT could be used by EUMETSAT to deliver data to a subset of their users.

Further PoC work may be required in performance testing IP multicast across NREN backbones and through campuses to the end users. GÉANT is willing to assist EUMETSAT in such future collaboration, if requested.

The collaboration with EUMETSAT also demonstrated the GÉANT connectivity services available to potential user projects from the academic network community and GÉANT's interest in working with new users to provide solutions that enable them to do their work efficiently.

Knowledge gained on the performance of high bandwidth IP multicast on the GÉANT network from this PoC with EUMETSAT could also benefit other user projects and communities, such as those involved in HDTV applications for remote learning and research.

1 Introduction

This Deliverable describes the collaborative work between EUMETSAT [EUMETSAT] and GÉANT. Tests were designed jointly to investigate the IP multicast ability to move data over the GÉANT network using TelliCast software. This document briefly describes the results from the tests performed in accordance with *Multicast over GÉANT Test Cases, EUM/OPS/TEN/09/4685* [RD 2]. The methodology, analysis and results are fully discussed in *Technical Report on Moving Data with Multicast in Collaboration with EUMETSAT* [EUMPoC].

First contact was made between GÉANT and EUMETSAT at the GEO-V Plenary meeting in Bucharest in November 2008, where EUMETSAT expressed interest in the connectivity services that GÉANT could offer, and in using the academic network. Future EUMETSAT missions (MTG, GMES, etc.) and imaging will generate significantly more data, and it is expected that more bandwidth will be needed. They explored operational aspects of the GÉANT service area and discussed different connection scenarios.

As a result, DANTE and DFN developed a technical collaborative PoC with EUMETSAT to demonstrate that the academic network can provide a viable and cost effective dissemination infrastructure to efficiently reach all user locations. The collaboration between GÉANT and EUMETSAT has proved positive at all levels.

1.1 Objectives

The objectives of the collaborative Proof of Concept (PoC) between GÉANT and EUMETSAT were to:

- Extend the technical knowledge and operational experience in using the GÉANT network to reliably support high bandwidth multicast applications at the required data rates.
- Demonstrate, using EUMETSAT hardware and software located in GÉANT PoPs, that the multicast option could successfully use the academic network to transmit data files using the TelliCast software to geographically dispersed locations.

1.2 Document Structure

Section 2 introduces the user requirements and possible connectivity solutions.

Section 3 describes the design of the PoC and tests.

Section 4 describes the results of the multicast tests and success of the GÉANT network to transport high-bandwidth multicast traffic.

Section 5 provides the conclusions based on the network performance and TelliCast test results.

The test configuration is described in more detail in [RD1] and in separate documentation [EUMPoC].

2 Networking Requirements from EUMETSAT

2.1 Introduction to EUMETSAT

The European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) operates a number of satellites collecting earth observation data and images. The data from the satellites is received by a number of ground stations and sent to EUMETSAT for processing; the resulting data is then disseminated as various data products to users and the research community using commercial telecommunication satellite links, in a similar way to satellite television. See Appendix A for a list of the EUMETSAT member states, which occupy a similar footprint to the GÉANT collaboration. Figure 2.1 illustrates the collection, processing and distribution of information. There is a wide range of applications for the data products, including numerical weather predictions, climate modelling, land surface analysis, ozone and atmospheric chemistry monitoring, hydrology and water management, and active fire monitoring. EUMETSAT aims to provide its service 24 hours a day, 365 days a year.

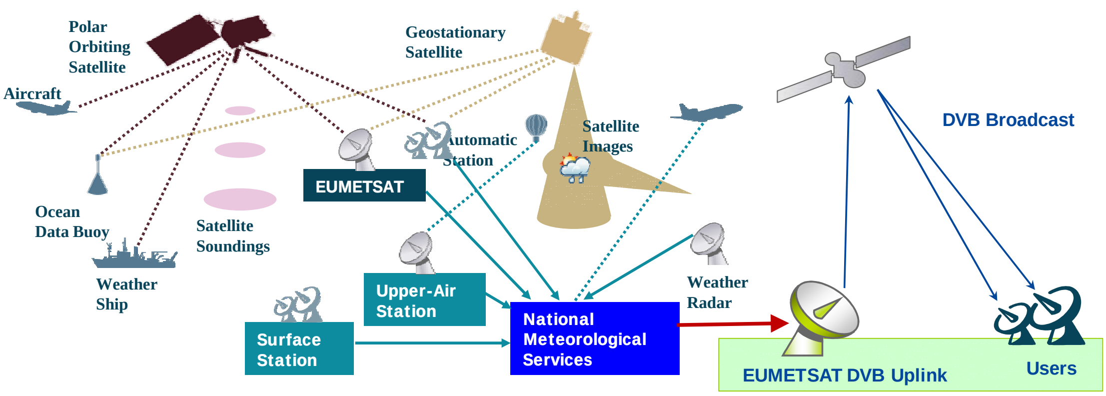

Figure 2.1: Diagram of EUMETSAT's data flow

2.2 Requirements for Data Movement

There is a continuous flow of data from acquisition, processing to dissemination. Timing is critical, as the information must be with the consumers within a predefined time of the images being taken.

Currently EUMETSAT's prime data dissemination mechanism is known as EUMETCast. EUMETCast is implemented using the TelliCast software supplied by Newtec [RD5] for transferring files via multicast over satellite systems using digital video broadcasting via satellite (DVB-S). The current bandwidth usage is 16.5 Mb/s, which is about half of the bandwidth of the current satellite (DVB-S) transponders. Future EUMETSAT missions (MTG, GMES, etc.) and imaging will generate significantly more data. It is expected that the bandwidth required for the new data products will increase by a factor of about 20 (to approximately 300 Mb/s), and might require terrestrial dissemination mechanisms for a subset of the data to complement the satellite broadcast system. See *Multicast over GÉANT Test Plan, EUM/OPS/TEN/09/4684* [RD1]. One such alternative dissemination mechanism could be to transfer files via multicast over a high-speed terrestrial network such as the GÉANT-NREN academic network. EUMETSAT expect that some users of the smaller data products will continue to use the satellite DVB-S system.

NA4 (*Liaison and Support* Activity) and SA1 (*Network Build and Operations* Activity) worked out the traffic patterns for data flows from Germany to the countries of interest to EUMETSAT, and demonstrated the possible reduction in the number of traffic flows if multicast was used to move their data.

2.3 Investigation of Possible Connectivity Options

The EUMETSAT data dissemination system is currently based on digital video broadcast (DVB) technology, with a bandwidth of 16.5 Mb/s. To support new image systems and advanced products, traffic rates potentially up to 300 Mb/s need to be correctly disseminated. Information is sent from the dissemination server at EUMETSAT to users of the data products located at the client sites.

Several unicast and multicast solutions, using both point-to-point circuits based on the GÉANT Plus [2] connectivity service or the routed IP academic network infrastructure to reach the user locations, were analysed. These possible solutions are presented and discussed in the following sections. For each potential architecture solution, there is a trade off in the complexity and capacity required in the network layers, as well as the manageability, hardware and computing power aspects of the server, client or application.

Multicast

Multicast is the delivery of messages (data) to a group of destination computers simultaneously in a single transmission from the source; copies are created automatically in the network elements (typically the routers) only when the topology of the network requires it.

---

2 GÉANT Plus is a point-to-point connectivity service across reserved, high-speed, guaranteed-bandwidth circuits on a pre-provisioned infrastructure.

Multicast is an efficient technique to deliver messages to several destinations in a network when several hosts are interested in the same message and/or when the receivers are unknown to the sender.

Messages are sent to a group which is identified by a group address. Hosts interested in the communication may join the group[3], and routers collaboratively create distribution paths for messages sent to multicast groups. The multicast distribution tree for a group is dynamically constructed when the receivers join the group. The tree construction is initiated by the routers that are close to the receiver. Just one copy of the message is sent by the source to the group. The links are used only once and copies of the message are generated just as the links to the multiple destinations split.

Unlike broadcast transmission (which is used on some local area networks), multicast clients receive a stream of packets only if they have previously elected to do so by joining the specific multicast group address.

Unicast

Unicast is communication between a single sender and a single receiver over a network. Separate unicast transmissions are required if different receivers need to be reached. As a consequence, the bandwidth requirement at the source increases as the number of users increase.

Possible solutions and their descriptions that would meet the requirements of EUMETSAT are as follows:

- Unicast over the IP Network.
- Multicasting over the IP network.
- Unicast over Point-to-Point.
- Multicast over Point-to-Point.

2.3.1 High Availability Services

To provide high availability (uninterrupted) services, all four connectivity options would require EUMETSAT and all client sites to be configured with suitable physically diverse backup links.

2.3.2 Unicast over the IP Network

Figure 2.2 shows a solution based on unicast. This approach is suitable for the exchange of data between EUMETSAT and their strategic partners.

To estimate the required bandwidth using, as an example, 30 client sites from the EUMETSAT member states, each requires a throughput of 300 Mb/s, giving a total of 9 Gb/s.

3 Internet Group Management Protocol (IGMP) is the network protocol that allows hosts to dynamically join a multicast group.

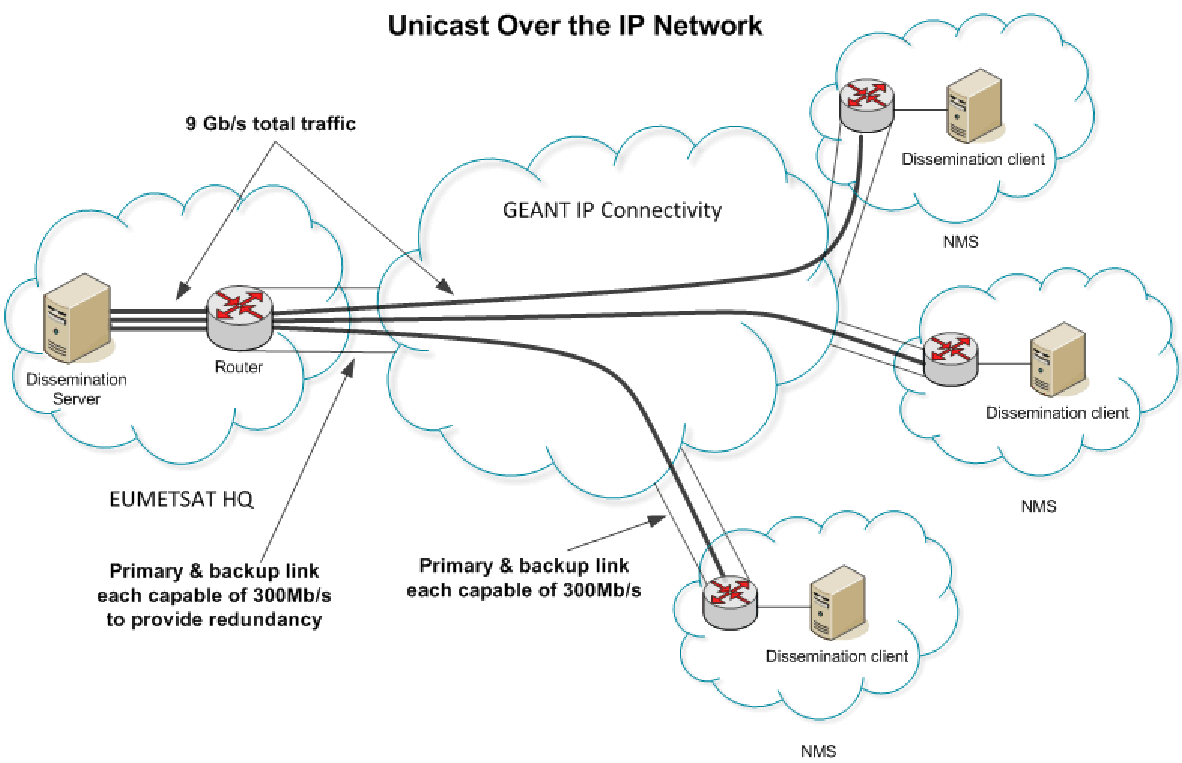

Figure 2.2: Unicast traffic flows over the routed IP network

Advantages of using the unicast dissemination strategy over the GÉANT/NREN IP network:

- The routed IP network offers inherent redundancy in the backbone.
- Low cost redundant connectivity at client sites.
- Simple configuration on client routers could be supplied by the user.
- Easier to debug problems than with the multicast solution.
- Routers at the client sites could be supplied/managed by the site.

Disadvantages of using the unicast dissemination strategy over the GÉANT/NREN IP network:

- No end-to-end bandwidth guarantee on the IP network.
- High volume connectivity required at EUMETSAT (9 Gb/s).
- EUMETSAT server must be capable of delivering 9 Gb/s.

2.3.3 Multicasting on the IP Network

Figure 2.3 shows a solution based on multicast. Note that the server, the clients and all the network equipment must be multicast aware.

Although there are 30 sites, only one multicast flow at 300 Mb/s is required from the dissemination server at EUMETSAT as the network replicates the data to the client sites as required.

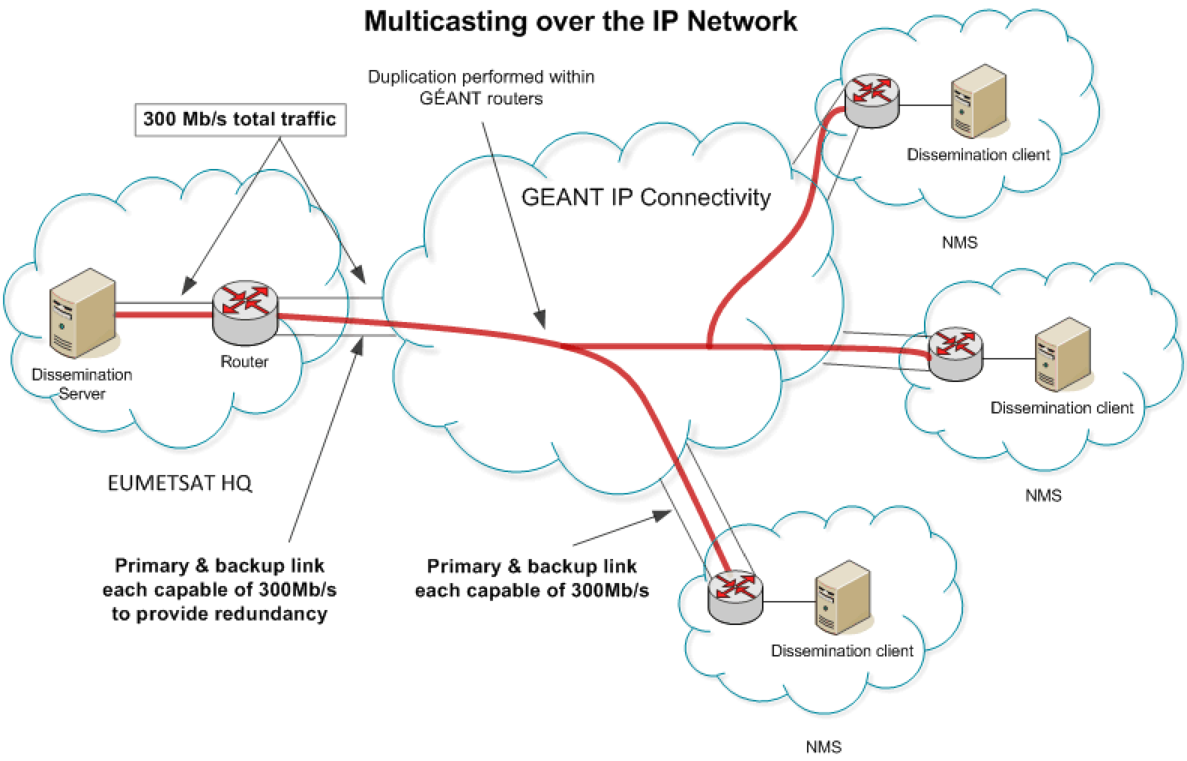

Figure 2.3: Multicast traffic flows over the routed IP network

Advantages of using the multicast dissemination strategy over the GÉANT/NREN IP network:

- The routed IP network offers inherent redundancy in the backbone.
- Low cost redundant connectivity at client sites.
- EUMETSAT server serves only a single set of datastreams at 300 Mb/s.
- Relies on IP multicast supported by the GÉANT network.

Disadvantages of using the multicast dissemination strategy over the GÉANT/NREN IP network:

- No end-to-end bandwidth guarantee on the IP network.
- Network problems are more difficult to debug in the multicast cloud.
- Complex configuration on client routers should be supplied by EUMETSAT.

2.3.4 Unicast over Point-to-Point

Figure 2.4 shows a solution based on unicast traffic flows, structured on point-to-point circuits provided using NREN and GÉANT Plus services.

Each of the 30 sites requires a point-to-point link operating at 300 Mb/s, but the server will need to operate at 9 Gb/s. High Availability requires 60 links.

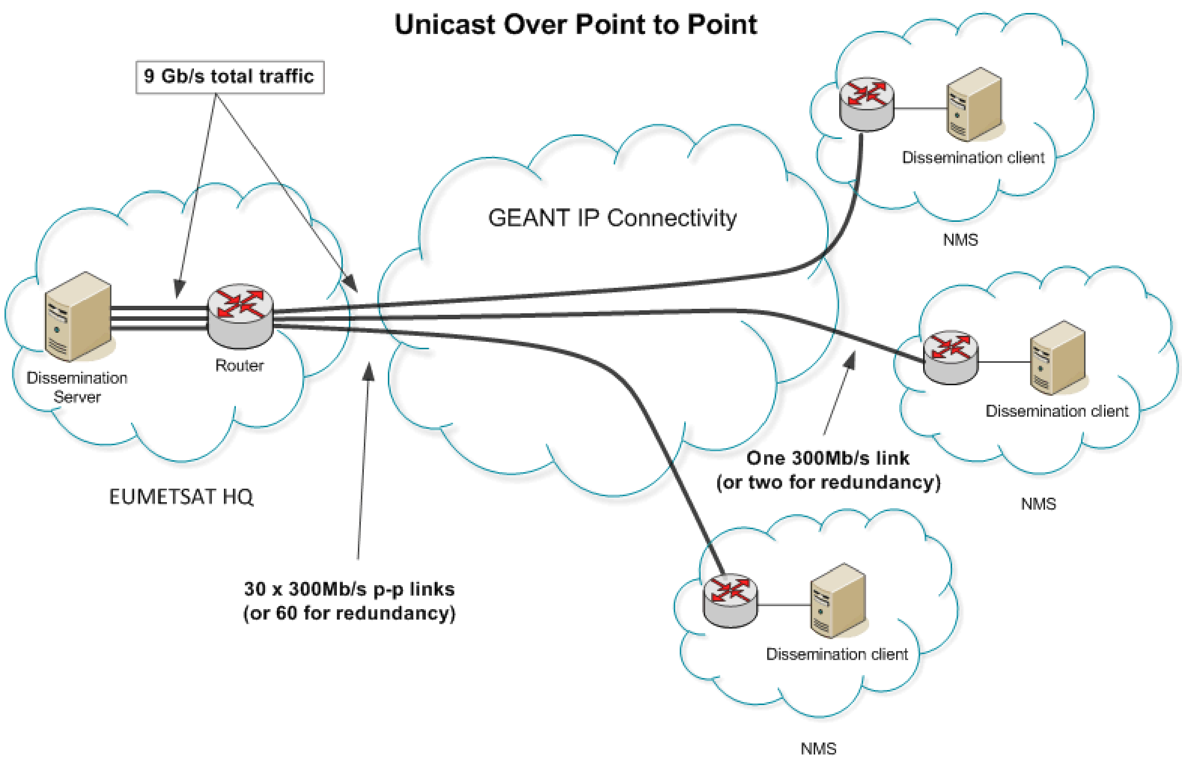

Figure 2.4: Unicast traffic flows over a series of point-to-point links

Advantages of using the unicast dissemination strategy over point-to-point connections:

- End-to-end bandwidth guarantee on the network.
- Network problems are easy to debug as all routers are visible, unicast and managed by EUMETSAT.
- Routers at the client sites could be supplied/managed by the site.

Disadvantages of using the unicast dissemination strategy over point-to-point connections:

- No inherent redundancy within the point to point lines; two lines would be needed.
- Cost of supplying full redundancy could be relatively high (60 point-to-point circuits).
- EUMETSAT server must be capable of delivering in total 9 Gb/s.

2.3.5 Multicast over Point-to-Point

Figure 2.5 shows a solution based on multicast. Note that the server, the clients and all the network equipment must be multicast aware.

30 links are required (one for each site) to carry the 300 Mb/s multicast traffic. The server will only need to operate at 300 Mb/s. High Availability requires 60 links.

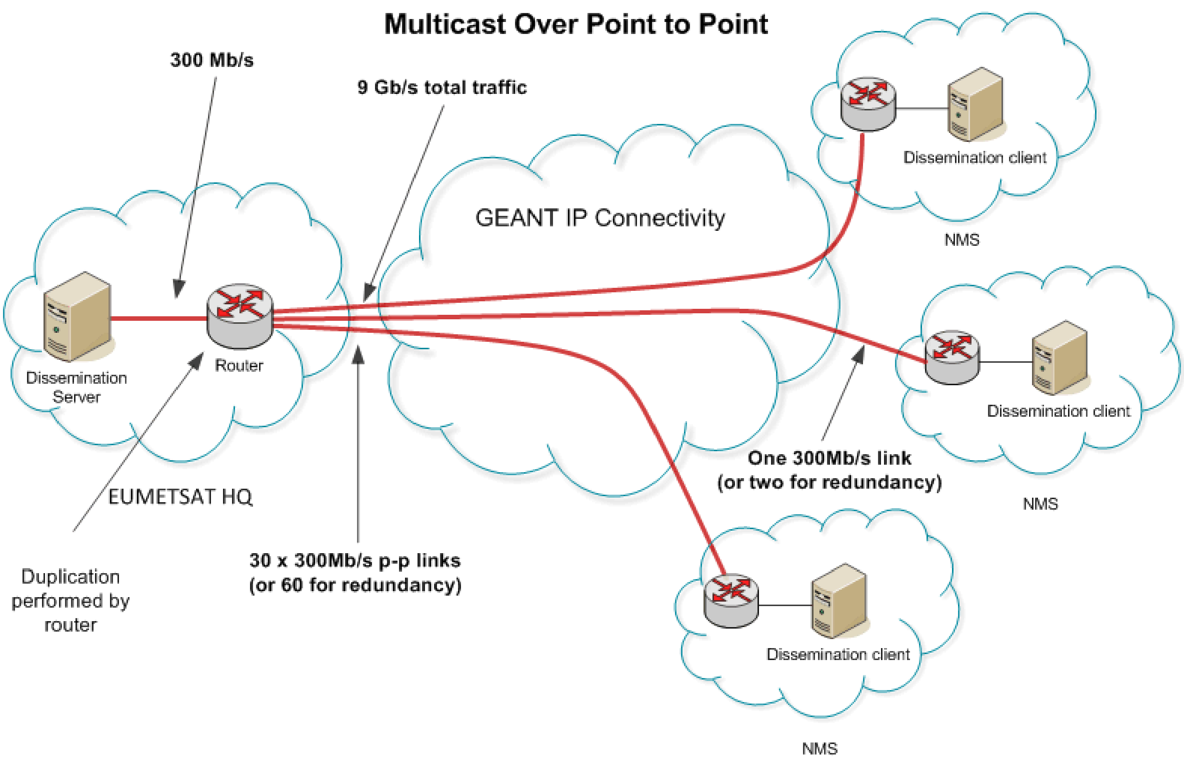

Figure 2.5: Multicast traffic flows over a series of point-to-point links

Advantages of using the multicast dissemination strategy over point-to-point connections:

- End-to-end bandwidth guarantee on network.
- EUMETSAT server serves only a single set of datastreams at 300 Mb/s.
- Network problems would be very easy to debug as all routers are visible and managed by EUMETSAT.

Disadvantages of using the multicast dissemination strategy over point-to-point connections:

- No inherent redundancy within the point-to-point lines; two lines would be needed.
- Cost of supplying full redundancy could be relatively high.
- Router must be capable of duplicating source data on to 30 interfaces (subnets).

2.3.6 Comparison of Moving Data Using Multicast and Unicast

Normal unicast flows using TCP/IP are a reliable way to move data, but in this application there would be many duplicated flows of the same data to different users at the same time. EUMETSAT would create a TCP link to push the data to a given user, which would require powerful, and probably multiple, servers to transmit the data. In addition, the access link from the network infrastructure to the EUMETSAT servers would require a bandwidth of at least 9 Gb/s. Figure 2.6 shows the routes over the GÉANT backbone which the unicast flows would take from EUMETSAT in Germany to each of the sites in each of the EUMETSAT member states (listed in Appendix A), assuming that each site required just one data product. As Figure 2.6 shows, often there are multiple flows of the same data over the backbone trunk links. In one case, there are seven duplicate flows.

Figure 2.6 demonstrates the advantage of using multicast from network and server points of view; the multiple unicast flows over the individual backbone links would be replaced by just one multicast flow. Users will only receive the data of interest as each type of data will be sent on a different multicast flow and the users will only join the multicast flows that interest them.

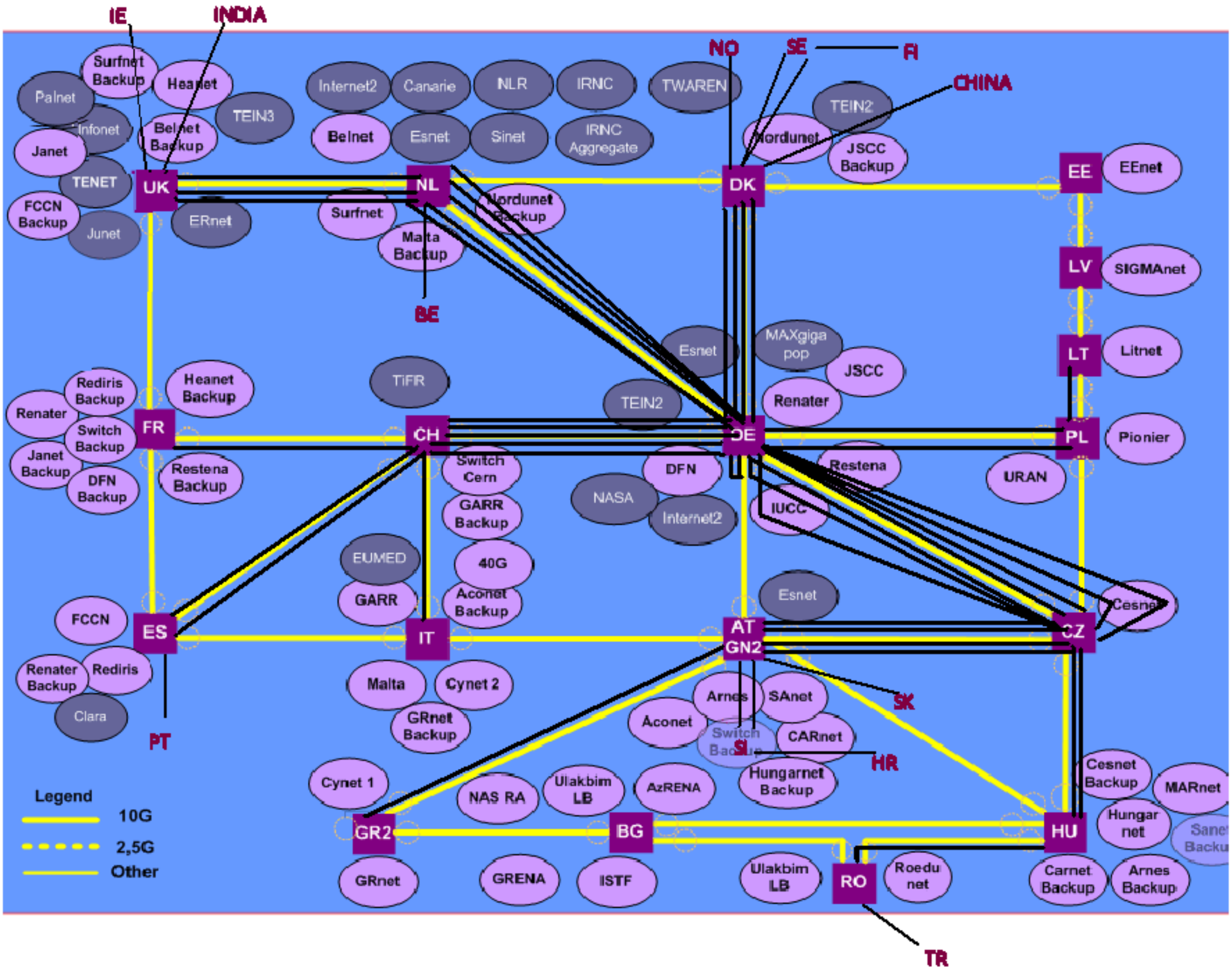

Figure 2.6: Unicast flows over GÉANT from EUMETSAT to major clients

3 Design of the Proof of Concept

This section provides high level descriptions of the jointly designed tests and preparations to run the PoC collaboration with EUMETSAT.

3.1 Preliminary Tests of the GÉANT Infrastructure

The preliminary tests used the existing multicast beacon servers to investigate the multicast capabilities of the GÉANT network. These servers act as multicast beacons [RD13] in the PoPs and are used to monitor the availability of the multicast service. These preliminary tests provided information on the performance of the LAN infrastructure in the PoPs, as well as the behaviour of the main GÉANT routers and trunk links. Security, operational and configuration issues were understood and solutions implemented. The servers involved were first tested for their UDP unicast performance, followed by performance tests using multicast flows.

3.2 Baseline Laboratory Tests of the EUMETSAT Equipment

Four high-end Dell laptops were provided by EUMETSAT with the TelliCast software installed. These were used in the PoC to transfer data files at the required data rates. Tests were performed in the DANTE and EUMETSAT laboratories to determine the baseline performance of the laptops and the TelliCast software without a wide area network. The tests performed included:

- UDP unicast performance: throughput, packet loss, packet jitter and one-way delay with the laptops directly connected with a 1 Gb link.
- The multicast performance: throughput, packet loss and packet jitter for single and multiple flows with the laptops connected on a Gigabit Ethernet LAN. (This used auto-generated Ethernet multicast addresses that match the required IP multicast group.)
- Moving data on a Gigabit Ethernet LAN using the TelliCast software.
- Determining the optimum conditions for performing the network tests to prevent the laptops from entering any power saving modes while performing a test.

3.3 EUMETSAT Tests over the GÉANT Network

Two general sets of tests were made when the EUMETSAT laptops were installed in the GÉANT PoPs. The first set of tests characterised the UDP unicast and multicast capabilities of the test hosts operating over the GÉANT network. The second set of tests used the TelliCast multicast software to transfer files via multicast between the test hosts by configuring various bandwidths for the data transfers. The following tests were performed:

- UDP unicast performance between the laptops following the network paths across GÉANT that will be used in the data flows: throughput, packet loss, packet jitter and one-way delay.
- The multicast performance of flows between the laptops in London and from London to Amsterdam and Prague: throughput, packet loss and packet jitter for single and multiple flows.
- Analysing the pattern of packet separation of the flows sent by the TelliCast software.
- Moving data over the GÉANT network, using the TelliCast software.

For more information about the tests, see [EUMPoC], which covers methodology and configuration in detail.

3.4 Network Planning

Figure 3.1 shows the high-level view of the network and installation of the EUMETSAT hosts used for the tests. The requirements were that two laptops were located in London (Lon1 and Lon2) to generate the multicast flows; one laptop to generate the data files and the other to serve these files over a set of multicast flows. Two further laptops (Pra and Ams) were located in different PoPs within the GÉANT network. These laptops required:

- Ability to send unicast and multicast traffic between the EUMETSAT end points.
- That the researchers had remote access to all laptops from the DANTE office and the EUMETSAT office.
- That all laptops were preconfigured and checked before delivery to the GÉANT PoPs.

Security Planning

A security risk analysis was performed by the DANTE operations group. The laptops were treated as accessible by members of DANTE and only EUMETSAT outside of GÉANT. Limitations were imposed by configuration of the firewalls in the GÉANT routers to only allow certain unicast and multicast traffic to and from the designated IP addresses. A rate-limiter was put in place for security purposes.

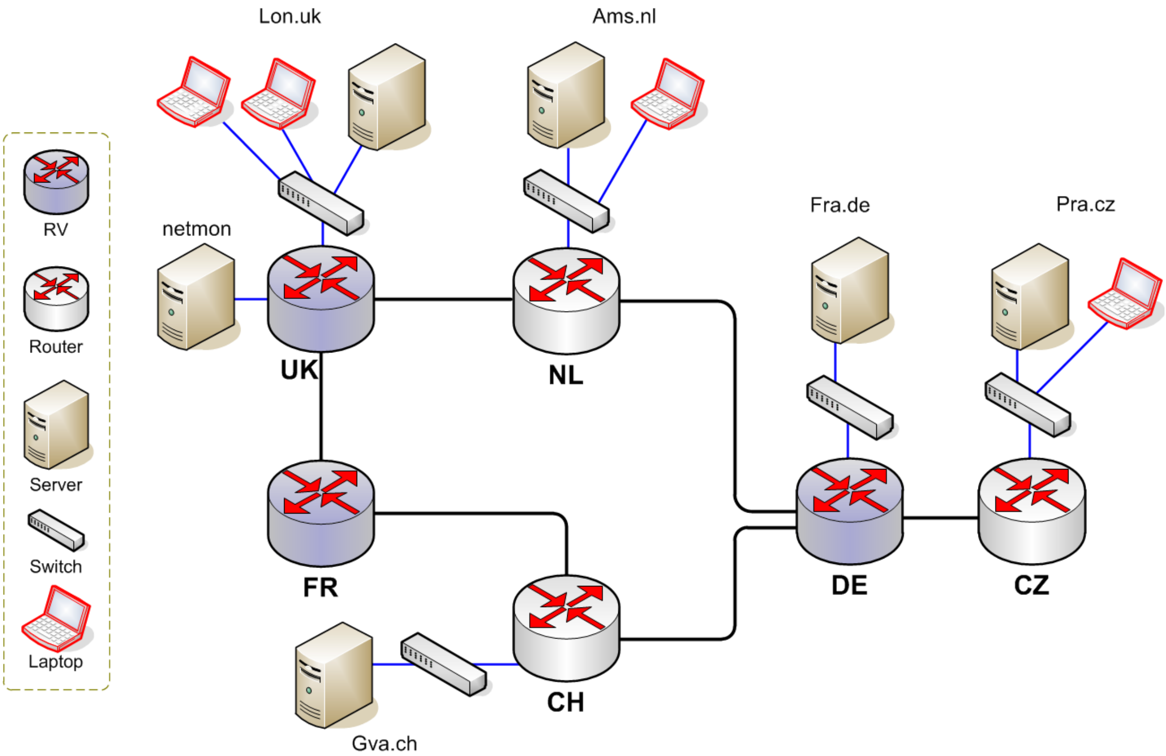

Figure 3.1: Topology of multicast implementation in GÉANT backbone and equipment used in PoC tests

4 Results from the EUMETSAT Multicast Proof of Concept

4.1 Preliminary Tests of the GÉANT Infrastructure

Measurements were made between the Sun servers in the GÉANT PoPs that were used as multicast beacons. The unicast tests were used to characterise the end hosts and to check the behaviour of the network interconnecting them. This includes the switches and the LAN in the PoPs as well as the GÉANT backbone. For detailed descriptions of the tests, see [EUMPoC].

The unicast tests explored the capabilities of the hosts and network to sustain high throughput data flows. The multicast tests explored the performance of high-rate multicast data flows. These tests were:

- Unicast UDP throughput, packet loss and packet jitter.
- Single flow multicast throughput and packet loss.
- Multicast inter-packet spacing.
- Long-run throughput.
- Multiple multicast flows.

4.1.1 Experience Gained

The multicast tests showed that the GÉANT network is capable of successfully transporting high-bandwidth multicast traffic. Once the multicast distribution tree was set up, no packet loss was experienced under normal conditions. The GÉANT network also supported simultaneous multicast flows using different groups. Long-run multicast tests showed that the network can sustain long-term multicast transmission. This indicated that the Juniper routers, the Rendez-Vous points and the backbone worked well in the provision of the multicast service.

During the multicast tests on the GÉANT network the switches that connect the end systems to their corresponding router in the PoPs proved to be the weak link of the chain. Experience was gained as follows:

- Switches should be multicast aware to correctly deal with multicast traffic.
- It is good practice to make the switches capable of recognising IGMP messages exchanged between hosts and their designated router (often referred to as enabling multicast snooping). IGMP enabled switches recognise the group-join requests and can isolate the ports that are interested in a particular multicast traffic. The switches avoid sending the multicast traffic to all ports on the switch.
- Some LAN switches have hardware limitations when IGMP was not enabled. For example, a slower interface connected to the switch can prevent the delivery of high-bandwidth multicast traffic to the interested hosts if flow-control is enabled. This caused a queue to build up in the switch, eventually causing losses on all flows through the switch.
- In addition to allowing the GÉANT GLOP multicast addresses, the multicast address range required by the TelliCast software had to be enabled through the security firewalls. Missing that, the TelliCast flows could not leave the LANs in the PoPs. Care was required to allow flows of all the required port numbers as well as the multicast groups.
- The need to access the Network Time Protocol (NTP) time service was overlooked in the requirements. This resulted in drifts in the times used by laptops in their logs, which made it more difficult to correlate the reports.

4.2 Baseline Laboratory Tests of the EUMETSAT Equipment

Detailed tests were performed on the EUMETSAT laptops in the DANTE laboratory to characterise their performance and network behaviour, and to determine the optimum conditions for the network tests that included preventing the CPUs from initiating any power saving measures. Unicast and multicast performance of the laptops was measured under these recommended test conditions. For detailed descriptions of the tests, see [EUMPoC].

EUMETSAT performed tests [RD3] of the TelliCast server and client software in their laboratory for different data transmission rates.

The following tests were performed in the laboratories:

- Laptop Unicast UDP throughput, packet loss and packet jitter.
- Laptop UDP one-way delays.
- Laptop Multicast throughput, packet loss and jitter.
- Moving data over a Gigabit Ethernet LAN using the TelliCast software.

4.3 Network Tests over GÉANT Using the EUMETSAT Laptops

With the laptops installed in the GÉANT PoPs, the network performance was first measured without the rate-limiter using unicast flows, then multicast flows. Without the rate-limiter, the network was found to operate correctly with both UDP and multicast packets with a throughput of 1 Gb, no packet loss and 2 µs to 4 µs of inter-packet jitter in full agreement with the laboratory tests. With the rate-limiter enabled, the flows were correctly limited to the 300 Mb/s that was set. For detailed descriptions of the tests, see [EUMPoC]. The following tests were performed:

- Unicast UDP throughput, packet loss and packet jitter.
- Unicast UDP one-way delay.
- Multicast throughput, packet loss and jitter.

4.4 Data Moving Using the TelliCast Software

4.4.1 Analysis of the Data Flows from the TelliCast Software

Tests were first made to verify the methodology proposed for analysing bursty traffic flows using the software tools developed for this PoC project; see *mcastmon* [RD10]. The methodology and the TelliCast multicast flows were then measured in the *netmon* server directly attached to the GÉANT router in London (see Figure 3.1). This proved that it is easy to distinguish the different types of traffic. Observations were made with the multicast data rates of 8 Mb/s, 50 Mb/s and 120 Mb/s configured in the TelliCast software. The *Technical Report on Moving Data with Multicast in Collaboration with EUMETSAT* [EUMPoC] provides detailed descriptions and analysis of the results.

Analysis of the multicast data flows produced by the TelliCast software at all the configured data rates showed that every 20 ms it generated bursts of packets at the 1 Gb connection speed of the host PC. The number of packets generated in each burst was correctly calculated to provide the average rates specified for all the configured flows. However, the bursty nature of the traffic triggered the rate-limiters agreed for the PoC (set at 300 Mb/s) when they were enabled. This contributed to the loss of packets, which led to loss of data files. See Table 4.1 and, in particular, items marked with “*”.

Analysis of 240 Mb/s Flow

It was not possible to collect a set of reliable results for the 240 Mb/s flow. The reason for this was probably related to the hardware performance and/or configuration of the test platform. As an example, the TelliCast software reported that the number of packets received was greater than the number of packets sent.

In the test configuration used, it appears that the TelliCast software was not able to transmit the files at the required rates. As a result, the input job queue builds up and, in order to maintain the configured bandwidth limitations, the server software deletes data files rather than transmit them in an attempt to catch up.

The disk performance of the laptops may also have contributed to performance issues, as the laptop disks were not the same specification as those used in the production servers. Disk performance was important, as the application required that data from the data generator laptop in London was written to the disk on the server laptop at the same time that data was read from the disk files by the TelliCast server for multicast transmission.

4.5 Overview of All TelliCast Tests

The data moving tests using the TelliCast software were performed as follows: Datasets, emulating the EUMETSAT traffic patterns, were generated by one of the laptops in the GÉANT PoP in London and sent to the second laptop that acted as the TelliCast multicast server. The data flows were multicast over the GÉANT network using several multicast groups. TelliCast multicast clients were run in the laptops located in the GÉANT PoPs in Amsterdam and Prague to receive the data flows. Several tests were performed at each of the following data rates, both with and without the rate-limiter enabled in London:

- 8 Mb/s.
- 50 Mb/s.
- 120 Mb/s.
- 240 Mb/s.

Table 4.1 shows a summary of all the results from the TelliCast tests and iterations. The percentages are based on the sent packet counter from the TelliCast server in London and the received packet counters from the TelliCast clients in Amsterdam and Prague. In general, the performance was considered very acceptable (98.6% is average value).

Test No/# Iteration/ Rate Limiter {On/Off}	Bandwidth	Date	Amsterdam	Prague
			Percentage of Transmitted Packets Received	
1 / #1 / Off	8 Mb/s	06.01.2011	100%	100%
1 / #2 / Off	8 Mb/s	13.01.2011	100%	100%
1 / #1 / On	8 Mb/s (*)	20.01.2011	99.9%	99.8%
2 / #1 / Off	50 Mb/s	10.01.2011	99.0%	98.7%
2 / #2 / Off	50 Mb/s	14.01.2011	100%	100%
2 / #1 / On	50 Mb/s (*)	21.01.2011	99.9%	99.9%
3 / #1 / Off(†)	120 Mb/s	11.01.2011	99.8%	81.5%
3 / #2 / Off	120 Mb/s	12.01.2011	99.0%	99.0%
3 / #3 / Off	120 Mb/s	17.01.2011	99.0%	99.8%
3 / #1 / On	120 Mb/s (*)	24.01.2011	98.9%	98.0%
4 / #1 / Off	240 Mb/s	06.01.2011	Counters not accurate due to server load conditions.	Counters not accurate due to server load conditions.

(*) Denotes that the rate limiter was enabled at 300 Mb/s on the LAN interface of the GÉANT router. (†) During this test there were many database full errors reported by the TelliCast software in Prague but not in Amsterdam. The reason is unclear, but no network related events were reported during this period.

Table 4.1: Summary of TelliCast test results

4.5.1 TelliCast Tests Summary

During the testing period, plots of the traffic statistics collected from the routers on the GÉANT backbone were examined to monitor and better understand the behaviour of the multicast flows as reported by the TelliCast software. All the tests were scheduled by EUMETSAT and needed to be performed during a given period of time. After investigating the packet losses reported for the conditions set for the 50 Mb/s tests with the rate-limiter set to off and 120 Mb/s with the rate-limiter set to off, three causes of the loss were identified as follows:

- The **congestion on the London-Amsterdam trunk** was caused by an exceptional network load due to significant changes in the traffic patterns from the particle physics community. As part of the policy to regularly review link capacity and usage, the planned upgrade of the GÉANT western ring was already in progress, but the upgrade of the London-Amsterdam link to 2 x 10 Gb had not taken place at the time of the multicast tests.
- Independent of this, while a scheduled test was in progress, there was a **link outage on the Frankfurt-Prague link due to the failure of an amplifier board**. This caused temporary loss of packets due to the automatic rerouting of the IP traffic.
- Network performance testing traffic was generated by the regularly scheduled perfSONAR tests that operated from approximately 00:00 until 06:00 on the LANs in the PoPs. The LANs have a throughput of 1 Gb/s, so this test traffic can cause congestion on the link to the router and subsequent **packet losses on the switch**. This affected the PoC tests, as there was a brief overlap of tests running at the same time around midnight.

4.5.2 TelliCast Related Issues

The *database full* errors that were observed in the TelliCast client log file occur if the space in the application database is not sufficient to hold the files currently in transit along with the incompletely received files for which the housekeeping has not yet been performed. Incompletely received files occur when packets are lost and cannot be recovered via forward error correction (FEC). Furthermore, once the application database is full, the housekeeping of the incompletely received files can create extra system or CPU load that itself can result in further packet losses. The situation only corrects itself when there is lossless reception (no more packets are lost) and no further database housekeeping is necessary.

The Dell laptops are high-end machines, however the TelliCast server software can exponentially consume computer resources as bandwidth usage is increased. The resources found to be impacted were CPU and disk IO throughput. As the backlog of files increase, disk IO activity increases and eventually the CPU usage will go over 90%, which forces a restart of one of the TelliCast processes. This results in files not being successfully transmitted.

Note: Newtec publish that the data rate supported by the TelliCast software is 80 Mb/s, so there could be issues when operating at higher data rates. See [RD5] and download the *TelliCast TL300 Product Leaflet*.

5 Conclusions

The preparatory multicast tests made with the multicast beacons and the mcastmon software showed that the GÉANT network is capable of successfully transporting high-bandwidth multicast traffic without any packet loss. Once the multicast distribution tree had been set up, no packet loss was experienced under normal conditions. Also the GÉANT network easily supports simultaneous multicast flows using different multicast groups or port numbers.

Measurements of the inter-packet arrival times, or packet jitter, with both the Sun multicast beacons and the EUMETSAT laptops showed a distribution with a width of about 4 µs which confirms that the GÉANT network transports the multicast packets in a stable manner and with minimal queuing.

As well as providing a base-line for comparison with the measurements made across the GÉANT network, the laboratory tests of the EUMETSAT laptops proved that they had the power required for the tests and provided insight into an optimal set of conditions to be used for making the tests to allow separation of the effects due to the end hosts from the real behaviour of the network.

Analysis of the multicast data flows produced by the TelliCast software showed that every 20 ms it generated bursts of packets at the 1 Gb line-speed of the host PC. The number of packets generated in each burst was calculated to provide the average rates specified for all the configured flows. However for the high rates used, this corresponded to several hundred packets in quite a short time interval (approximately 1 ms), followed by still periods. Even with a rate-limit configured for 300 Mb/s and a requested rate of only 120 Mb/s, there was a substantial number of packets rejected which resulted in some lost data files. It is anticipated that, due to the bursty nature, the traffic could be severely impacted by the policies and capabilities of the equipment at the end sites. This could be a useful point for discussion with the software suppliers.

The packet losses that were experienced on one of the trunks during one of the tests reinforces the importance of the GÉANT policy of regularly reviewing the link usage, as well as planning and implementing upgrades to engineer a service capable of transporting the peaks of the user traffic.

In the tests that were performed, the GÉANT network demonstrated its capability to transmit (between three hosts located throughout Europe) the multicast data stream that was generated by the TelliCast file transfer software.

At data rates of up to 50 Mb/s the average reception performance was very close to 100%. The only losses recorded at these data rates were caused by a link outage on the Frankfurt-Prague link, which was due to emergency maintenance to replace an amplifier board on that link, and exceptionally high network load on the London-Amsterdam link.

As part of the policy to regularly review link capacity and usage, DANTE has now completed the process of upgrading the London-Amsterdam link to 2 x 10 Gb/s. It is thought that this will be sufficient to prevent the network-load related losses for this level of multicast throughput. However, these tests demonstrated that there could always be the possible risk of packet losses due to exceptional network load. Use of QoS (Quality of Service) on the complete data path to the end site could be an advantage.

At the data rate of 120 Mb/s the average reception performance dropped below 100%. For the majority of the tests the TelliCast application reported no significant losses. However, when losses occurred they occurred in batches. Nearly all of the losses that could not be attributed to network-related events coincided with database errors in the TelliCast client and no packet losses were actually recorded on the GÉANT trunks at these times.

Note that in this context there were some discrepancies between the results from the TelliCast tests performed in the laboratory environment and those performed on the GÉANT network at 120 Mb/s. The TelliCast multicast tests that were performed in the laboratory environment demonstrated lossless reception up to and including a data rate of 240 Mb/s. In addition, DANTE multicast tests that were performed on the GÉANT network demonstrated lossless reception up to the imposed traffic rate limit of 300 Mb/s.

The PoC was successfully completed in February 2011. It proved to be an excellent collaboration on all levels including the exchange of information, sharing and supporting network resources and testing tools, performing jointly designed tests, and joint analysis of the data and conclusions.

EUMETSAT may consider the following for future work, with which GÉANT would assist if requested.

- Perform an in-depth analysis of the configuration of the GÉANT network equipment and the TelliCast software to optimise their interaction for this type of multicast traffic.
- Perform an in-depth analysis of the actions of the TelliCast server application to better understand the unexpected data bursting that was observed in the TelliCast server application on the network during the tests.
- Assess the impact of reconfiguring the test laptops for use on the GÉANT network.
- Assess the impact of the long uptimes on the test laptops.
- Perform additional laboratory tests to better understand the apparent discrepancy between the results of the original laboratory tests and those performed on the GÉANT network.

This PoC successfully demonstrated the use of IP multicast on the GÉANT backbone.

Further PoC work may be required in performance testing IP multicast across NREN backbones and through campuses to the end users. GÉANT is willing to assist EUMETSAT in such future collaboration, if requested.

This PoC has shown that GÉANT could be used by EUMETSAT to provide data to a subset of their users.

Appendix A EUMETSAT Member States

The list of EUMETSAT member states is defined in [EUMETSATMAP].

- Germany
- UK
- France
- Portugal
- Spain
- Croatia
- Turkey
- Latvia
- Denmark
- Norway
- Sweden
- Finland
- Ireland
- Poland
- Romania
- Netherlands
- Belgium
- Slovakia
- Austria
- Hungary
- Switzerland
- Greece
- Italy
- Slovenia
- Czech

References and Further Resources

[EUMETSAT]	http://www.eumetsat.int/Home/Main/AboutEUMETSAT/index.htm?l=en
[EUMETSATMAP]	http://www.eumetsat.int/Home/Main/AboutEUMETSAT/WhoWeAre/MemberandCooperatingStates/index.htm]
[EUMPoC]	*Technical Report on Moving Data with Multicast in Collaboration with EUMETSAT*, R. Hughes-Jones, GN3-11-127
[RD1]	*Multicast over GÉANT Test Plan*, EUM/OPS/TEN/09/4684
[RD2]	*Multicast over GÉANT Test Cases*, EUM/OPS/TEN/09/4685
[RD3]	*Multicast over GÉANT Test Report*, EUM/OPS/TEN/11/0187
[RD4]	*udpmon: a Tool for Investigating Network Performance* http://www.hep.man.ac.uk/u/rich/net
[RD5]	TelliCast, http://www.newtec.eu/products/ip-software/tellicast-tl-300/
[RD6]	*Host Extensions for IP Multicasting,* S.E Deering, RFC 1112, AUG1989.
[RD7]	*MULTICAST Tutorial,* RedIRIS/Red.es, OCT 2004 http://www.garr.it/emc_training/tutorials/mcast_tutorial.pdf
[RD8]	*Protocol Independent Multicast – Sparse Mode (PIM-SM): Protocol Specification (Revised)* B. Fenner, M. Handley, H. Holbrook, I. Kouvelas, RFC 4601, AUG2006.
[RD9]	*GLOP Addressing in 233/8,* D. Meyer, P. Lothberg, RFC 3180 (supersedes 2770), SEP 2001
[RD10]	*mcastmon*, http://www.hep.man.ac.uk/u/rich/Tools_Software/mcastmon.html
[RD11]	*IANA Guidelines for Ipv4 Multicast Address Assignments* M. Cotton, L. Vegoda, D. Meyer, RFC 5771, MAR 2010
[RD12]	*The Unicast and Multicast Performance of the EUMETSAT Laptops*, RHJ-10-20, 2-12-2010
[RD13]	*NLANR/DAST Multicast Beacon*, download from SourceForge http://sourceforge.net/projects/multicastbeacon
[RD14]	*EUMETSAT successfully distributes meteorological data using Newtec's TelliCast Software* http://www.newtec.eu/company/press-room/press-releases/detail-page/browse/2/detail/eumetsat-succesfully-distributes-meteorological-data-using-newtecs-tellicast-software-platform/?tx_ttnews%5BbackPid%5D=149&cHash=542431a557

Glossary

AS	Autonomous System
BWCTL	Bandwidth Test Controller
CPU	Central Processing Unit
DVB	Digital Video Broadcast
DVB-S	Digital Video Broadcasting-Satellite
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FEC	Forward Error Correction
ftp	File Transfer Protocol
GÉANT Plus	Point-to-point connectivity service across reserved, high-speed, guaranteed-bandwidth circuits on pre-provisioned infrastructure.
GLOP	(Addressing), a method to allocate multicast address space
GUI	Graphical User Interface
IGMP	Internet Group Management Protocol
IO	Input Output
IP	Internet Protocol
LAN	Local Area Network
mcastmon	Suite of programs used for investigating the multicast performance of a network
MTP/SO	Multiple Transport Protocol/Self Organising
multicast	Simultaneous communication in a single transmission to a group of destination computers
NACK	Negative Acknowledgement
NMS	National Meteorological Services
NREN	National Research and Education Network
NTP	Network Time Protocol
perfSONAR	PERFormance Service Oriented Network monitoring ARchitecture
PoC	Proof of Concept
point-to-point	Communication similar in meaning to unicast
PoP	Point of Presence
QoS	Quality of Service
RV	Rendez-Vous Point
TCP/IP	Transmission Control Protocol/Internet Protocol
TelliCast	Commercial data moving software licensed by EUMETSAT
UDP	User Datagram Protocol
udpmon	Software tool for investigating network performance
unicast	Communication between a single sender and a single receiver over a network