14.06.05

Deliverable DJ.3.2.1: GÉANT2 Bandwidth on Demand (BoD) User and Application Survey

Deliverable DJ.3.2.1

Contractual Date:	31/03/05
Actual Date:	14/06/05
Contract Number:	511082
Instrument type:	Integrated Infrastructure Initiative (I3)
Activity:	JRA3
Work Item:	2 (Two)
Nature of Deliverable:	R (Report)
Dissemination Level	RE (Restricted)
Lead Partner	GRNET
Document Code	GN2-05-086v11

Authors: Vladimir Rabljenovic (CARNet/SRCE), Gabor Ivanszky (Hungarnet), Simon Muyal, E. Camisard (Renater), Maarten Büchli (DANTE), Radek Krzywania (PSNC), Afrodite Sevasti (GRNET), Esther Robles (RedIRIS)

Abstract

The objective of this deliverable is to identify key Bandwidth on Demand (BoD) requirements of advanced applications that exist at the beginning of GN2, in order to provide a comprehensive overview of user perceptions of what a BoD service should offer them. This information will be a valuable input for the GN2 BoD service specification process within JRA3 Work Item 3 (WI-03).

Table of Contents

0 Executive Summary iv

1 Identification of BoD service users 1
1.1 Projects 2
1.1.1 e-VLBI (http://www.evlbi.org/) 2
1.1.2 DEISA (http://www.deisa.org/) 2
1.1.3 GRID projects 2
1.1.4 MUPBED (http://www.ist-mupbed.org/) 3
1.1.5 GN2 Testbed 3
1.1.6 LOFAR (http://www.lofar.org/) 3
1.1.7 Evergrow (http://www.evergrow.org) 4
1.2 Connectivity to research infrastructures 4
1.2.1 HEC Facilities 4
1.2.2 Earth Observation 5
1.3 Connectivity with special networks/sites 5
1.3.1 CERN 5
1.3.2 USA & Starlight 5

2 BoD Requirements of Advanced Applications 6
2.1 General issues 7
2.1.1 EGEE 7
2.1.2 Earthquake simulation project 8
2.1.3 SEE-GRID 8
2.1.4 CrossGrid 8
2.1.5 GRIDCC 9
2.1.6 CANARIE 11
2.2 Control Issues 11
2.2.1 Overview 12
2.2.2 Signalling interface 12
2.2.3 Interaction between application and service reservation interface 13
2.2.4 Duration and frequency of BoD service usage 14
2.3 Network Issues 15
2.3.1 Overview 15

2.3.2 Bandwidth requirements 16
2.3.3 Traffic patterns 17
2.3.4 Description of network infrastructure resources 18
2.3.5 Description of future network infrastructure resources 19
2.3.6 Bandwidth usage monitoring 20
2.4 Reliability Issues 22
2.4.1 Overview 22
2.4.2 Results of the BoD questionnaire 23
2.4.3 Reliability requirements 25
2.5 Other Issues 26
2.5.1 Overview 26
2.5.2 Extended requirements and future plans 27
2.5.3 BoD usage overview 27
2.5.4 BoD service test participation 28

3 Conclusions 30

Appendix A BoD requirements' questionnaire 32

0 Executive Summary

The Joint Research Activity 3 (JRA3) is focused on specifying, developing and piloting a Bandwidth Allocation and Reservation or Bandwidth on Demand (BoD) service over interconnected domains that may implement heterogeneous network technologies. In particular, the activity will design and implement the necessary bandwidth reservation systems and signalling interfaces that will allow end-users to make advance reservations with automated provisioning. The aim of JRA3 is to have a pilot service running by the end of the GN2 project.

A BoD service provides an end-to-end connection with a guaranteed capacity possibly spanning different administrative domains using preferably connection-oriented technologies. This type of service is sometimes also referred to as a 'lightpath' although it is envisaged that the BoD service developed in JRA3 will not be exclusively based on all-optical transmission. It is also envisaged that the service will be automated to a large extent, the intention here being to give the end-user, middleware or application a certain degree of control over the network.

As a first step in this direction, information has been collected about GRID activities and other advanced applications, user groups and communities that have potential requirements for high capacity, connection-oriented BoD services. Although a large number of potential uses of BoD service were initially identified, it was only possible to collect contact details and establish communication channels for no more than a dozen of them. To gather adequate BoD requirements information, a BoD questionnaire has been defined, and categorised with five main areas to analyse. Emphasis has been placed on identifying the broadest spectrum of functional specifications that might be required from applications, in terms of control issues (signalling interfaces, interaction between user applications and the service reservation interface as well as duration and frequency usage parameters), network issues (bandwidth size and granularity, possible traffic patterns, network infrastructure resources already used and their future development, and bandwidth usage monitoring), reliability issues (latency, frame loss and reordering, confidentiality, path restoration in a case of failure), and other issues that users perceived as important to mention (like BoD usage information for user statistics and accounting, willingness to participate in BoD service testing process). Also, potential user groups have been given an introduction into JRA3 objectives, in order to better understand what is expected from them. Since most of the potential user projects are also in the initial process of deployment and there is the possibility of new user groups and requirements emerging then there is an obvious need to plan a revision process of gathering BoD requirements. This will be done after the first 18 months of the project.

After the BoD questionnaires were distributed to the potential user groups, some of them responded with the detailed information requested, but there were some that did not respond at all, or have postponed it for some time in future. According to the available information, BoD service requirements analysis is performed and presented in this document.

1 Identification of BoD service users

In order to identify a set of potential users for a BoD service, a questionnaire provided by DANTE to the NRENs in advance of the GN2 project launch was used as a starting point. The aim of this DANTE questionnaire was to get from NRENs the projected GN2 Network utilisation in order to adequately plan the network capacity requirements over the next four years. The information filled by NRENs included also the projects and special users' bandwidth requirements.

To complete the collected information from DANTE, the GÈANT Access Port Managers (APMs) and the TF-NGN user community was also contacted.

Clearly, three groups of potential BoD users have been identified:

- Projects: In this group, we have identified a number of currently running projects (or those starting in the near future) that require high levels of bandwidth between different sites across Europe to be provided by an added value service such as BoD.
- Connectivity to research infrastructures: A number of infrastructures have been identified to which access is required for collaboration and working purposes of research users, resulting in anticipated huge requirements in terms of bandwidth and advanced connectivity service characteristics.
- Connectivity with special networks/sites: Certain user communities have expressed their high bandwidth requirements for connectivity with some special networks/sites. Hence, it is possible that some of those requirements are served using the BoD service.

During the investigation of users' and applications' BoD requirements in the European research and educational community, only a subset of the identified entities was able to provide a detailed specification of their requirements. These detailed specifications are thoroughly presented and analysed in Chapter 2 of this deliverable. The rest of the identified entities (projects, user groups, applications) with potential need of BoD in the near future are briefly presented in the next sections. It is envisaged that more detailed requirements from them will be available and included in future revisions of this deliverable.

1.1 Projects

1.1.1 e-VLBI (http://www.evlbi.org/)

The European VLBI[1] Network (EVN) is an interferometric array of telescopes spread throughout Europe and beyond, which conducts unique, high resolution, radio astronomical observations of cosmic radio sources. It is the most sensitive VLBI array in the world, thanks to the collection of extremely large telescopes that contribute to the network. Observation data is collected at the remote telescope sites and correlated at a central location (currently the Joint Institute for VLBI in Europe – JIVE – located at the ASTRON facility in the Netherlands). Up until now, these collections have been done using magnetic tapes but recently experiments have been undertaken in which the data collected is transported over the European R&E network infrastructures either in real-time (live data streams) or through the use of off-line, high speed file transfers. When these "eVLBI" techniques are proven then future data collection would be done through real-time transfers at 1Gbps in the near future (2004-2007) and up to 8-30 Gbps in the longer term (2007-2010).

1.1.2 DEISA (http://www.deisa.org/)

DEISA (Distributed European Infrastructure for Supercomputing Applications) is a consortium of leading national supercomputing centres in Europe that intends to jointly build and operate a distributed tera-scale supercomputing facility. This objective will be attained by a deep integration – using modern Grid technologies – of existing high end national High Performance Computing (HPC) platforms.

The tightly coupled distributed supercomputing infrastructure proposed by DEISA, is a single system image extension of the national supercomputing clusters. This infrastructure is a tightly coupled Grid system that operates below the traditional middleware toolkits. This is the most efficient way of providing reliable high performance services to a large user community, and requires the availability of a high performance network infrastructure (probably including reserved bandwidth) across the DEISA computing platforms.

1.1.3 GRID projects

Within the individual European NRENs there are initiatives to coordinate the research groups which are interested in the GRID technology, both at a development and applications level.

Examples of these initiatives are IRIS-GRID in RedIRIS (http://irisgrid.rediris.es), GridIreland in HEANET (http://www.grid-ireland.org/), NorduGrid in Nordunet http://www.nordugrid.org/, SEE-GRID in the South-East Europe countries (http://www.SEE-GRID.org/) (detailed information about the BoD requirements of this initiative are provided in chapter 2) or the Hungarian ClusterGrid Infrastructure.

[1] VLBI means Very Long Baseline Interferometry

JRA3 will keep a close eye on these initiatives and try to cater for their BoD demands as much as possible in the near future.

1.1.4 MUPBED (http://www.ist-mupbed.org/)

The main goal of the this project is to integrate and validate, in the context of user-driven large-scale testbeds, ASON/GMPLS (Automatically Switched Optical Network/Generalised Multi Protocol Label Switching) technology and network solutions as enablers for future upgrades to European research infrastructures.

This goal will be pursued by creating a large experimental environment to assess the proposed network solutions, and that will be offered as an open test platform to other European research projects and users. The testbed will represent a multi-layer network based on IP/MPLS and ASON/GMPLS technologies, equipped with a unified control plane and designed to support the highly demanding applications of the European research community.

1.1.5 GN2 Testbed

The high profile of networking research and development activities within the GN2 project requires a dedicated testbed facility. The Technology and Service Testing research activity (JRA4) will provide a distributed testbed to accommodate the methodical testing of the technologies and techniques developed by the GÉANT2 research programme.

The test facility will incorporate gigabit IP routers and switches, as well as appropriate DWDM transmission equipment, in order to provide an environment for testing the technologies that will be used as building blocks for the implementation of the next-generation network (NGN) architecture.

The test facility is primarily intended for the use of research activities and network development within GÉANT2. However, it is possible that the facility may also be made available to external users (dependent on availability). In particular, the test facility may be able to offer support to other projects that fall within the scope of the EU's Sixth Framework Programme. The most flexible use of this infrastructure could probably be achieved by the use of the BoD service implemented within JRA3. However, a production BoD service is not anticipated to be in place (even on GÉANT2 alone) until the end of the GN2 project. This means that a JRA3-type BoD service may well be of use in implementing future incarnations of a testbed. For the period of operation of the JRA4 testbed under the GN2 project, JRA3 will most likely be one of the most active users of the testbed itself for the purposes of validating its BoD service designs and middleware component implementations.

1.1.6 LOFAR (http://www.lofar.org/)

LOFAR started as a new and innovative effort to force a breakthrough in sensitivity for astronomical observations at radio-frequencies below 250 MHz. LOFAR is the first telescope of this new sort, using an array of simple omni-directional antennas instead of mechanical signal processing with a dish antenna. The electronic signals from the antennas are digitised, transported to a central digital processor, and combined in software to emulate a conventional antenna. The cost is dominated by the cost of electronics and will follow Moore's law, becoming cheaper with time and allowing increasingly large telescopes to be built. So LOFAR is an IT-telescope. The antennas are simple enough but there are a lot of them - 25000 in the full LOFAR design. To make radio pictures of the sky with adequate sharpness, these antennas are to be arranged in clusters that are spread out over an area of ultimately 350 km in diameter. (In phase 1 that is currently funded 15000 antennas and maximum baselines of 100 km will be built). Data transport requirements are in the range of many Tbps and the processing power needed is tens of Tera-FLOPS.

1.1.7 Evergrow (http://www.evergrow.org)

The goal of the project is to build the science-based foundations for the growing global information networks of the future.

Networks will soon not only give access to all the world's knowledge, but the entire society will be network-based, from private life and business to industry and the processes of government. Above all, a number of today's highly manual processes must be automated, such as network management, network provisioning and network repair on all levels.

The project objectives are to integrate advances in their knowledge of data networks and the content they carry, through new techniques of measurement, advances in the theory of the emergent phenomena which are encountered in these large, ever-growing dynamic networks, with a focus on mechanisms to regulate their growth and channel user behaviour in the most effective directions, and advances in the architecture and functioning of overlay networks, which are the critical tool for exploiting this growing wealth of computing and communications power.

The project brings together the best research groups in three main areas: complex systems, peer-to-peer systems and experimental networking.

To do this it will provide a common platform and laboratory, the "virtual observatory", of computers coupled through the latest GRID technology. In the observatory it will provide archived measurements, tools to extrapolate them to 2025, test environments in which new services can be prototyped and evaluated against network mock-ups, and the computing power to support the extensive computer experimentation that is needed to allow surprises to surface in the theoretical study of such complex systems as these.

1.2 Connectivity to research infrastructures

1.2.1 HEC Facilities

High End Computing (HEC) facilities are already being connected to the research networks in order to be used by the research community. Even more, some national governments are giving an incentive to the creation of such facilities as it is the case in Ireland or in Spain with the MARENOSTRUM supercomputer. It is expected that these HEC facilities will attract communications/bandwidth requirements in the future.

1.2.2 Earth Observation

Earth Observation images show the world through a wide-enough frame so that complete large-scale phenomena can be observed to an accuracy and entirety it would take an army of ground-level observers to match. While satellite acquisitions are most often presented in the form of pictures, they are actually digital data. So the same raw data can be processed with computer software in many different ways to extract whatever information the particular end user requires.

To stress the importance of this work, an agreement has been reached by several countries in order to establish a Global Earth Observation System of Systems (GEOSS). GEOSS is envisioned as a large national and international cooperative effort to bring together existing and new hardware and software, making it all compatible in order to supply data and information at no cost.

1.3 Connectivity with special networks/sites

1.3.1 CERN

All European NRENs connect research centres that collaborate within projects where the access to the CERN facilities is a key point. As a result from the DANTE questionnaire, NRENs ask for huge amounts of bandwidth to and from CERN for several projects as CMS, tests with H-TCP, Delphi etc.

1.3.2 USA & Starlight

Similarly to the case of CERN, some of the DANTE questionnaire answers show special requirements in terms of bandwidth between sites (such as FERMILAB, a partner in projects with CERN and other sites) connected through USA research networks and Starlight

2 BoD Requirements of Advanced Applications

Based on the identification of potential BoD user applications, it was necessary to abstract generic and specific network requirements. Emphasis was placed on identifying the broadest spectrum of functional specifications that might be required from applications. In order to gather these specifications, a set of detailed questions was grouped into the BoD questionnaire form. According to their logical context, they were divided into following issues:

- General Issues
- Control Issues
- Network Issues
- Reliability Issues
- Other Issues

The content of the BoD requirements questionnaire, is provided in addition to this document, in Appendix A. In this section we present and analyse the complete set of answers provided by different projects to this questionnaire with the exception of the Earthquake simulation project, which provided just a short overview of their projected requirements, and did not complete answers in the questionnaire form.

Classification of the set of requirements from a potential BoD service was done using the categorization mentioned above. Analysis of gathered BoD requirements is given in the next sections of this document. Since the content of this deliverable depends a lot on the information provided in the responses to the BoD requirements questionnaire, which was in many cases presented in a rather sparse manner, the extent of information regarding particular user groups can often vary in quantity and quality. In general, stated BoD requirements could be categorized as *required* or *desirable*, depending whether each particular user group provided this classification or not.

2.1 General issues

General issues are ones that were actually provided by user groups, in order to have their overview while analysing given BoD requirements. There are 6 projects that responded to the given questionnaire, four GRID projects (EGEE, SEE-GRID, GRIDCC and CrossGrid), the Earthquake simulation project, and CANARIE network (the Research and Education Network of Canada). CANARIE is not quite a user group, but rather represents here a BoD-like service already implemented over a network infrastructure with which the GN2 BoD service is highly probable to interoperate in the future.

2.1.1 EGEE

EGEE is proposing a large-scale Grid for eScience applications covering the European Research Area and other world regions. This Grid will be a distributed infrastructure and will therefore be dependent on a communications network that links its resources. The European Research Area is currently served by the National Research and Education Networks (NRENs) linked via the high-speed pan-European backbone, the GEANT network. The EGEE Grid will use the European research networks to connect the providers of computing, storage, instrumentation and applications resources with user virtual organizations.

Distinct from the resources supplied to the Grid, EGEE will construct an integrated and scalable system to manage:

- Access to the resources of the Grid
- Relationships between parties interacting via the Grid
- Operation of the Grid as a reliable service.

All of these dimensions must have a network perspective for successful provision of Grid services.

- EGEE needs access to bandwidth allocation and reservation services which will be offered by the underlying network providers, i.e. the pan-European research network - GÉANT and the participating National Research and Education Networks (NRENs).
- The clients of these network services will be the higher layer EGEE Grid middleware, such as the Workload Management System, Data Replication Service etc. They will be referred to as the Higher Layer Middleware (HLM).
- EGEE develops a component of middleware referred to as the 'Bandwidth Allocation and Reservation Service' (BAR). BAR sits in between the HLM and the network, and provides an abstracted view of the network services in a similar way as the Computing Element (CE) and the Storage Element (SE) do for computing and storage resources.

2.1.2 Earthquake simulation project

Information about this user group is limited, since they did not fully reply to the BoD questionnaire provided by JRA3. The objective of this project is to provision for coupling the major EU earthquake engineering research labs together over the Grid. They will be basing the connection around the US Network for Earthquake Engineering Simulation (NEES) (see http://www.nees.org/), but will also be developing it to suit the European needs.

2.1.3 SEE-GRID

The SEE-GRID project will implement a Grid infrastructure in South East Europe, in order to assist the SEE countries to integrate progressively to the Pan-European Grid infrastructure (EGEE). The SEE-GRID infrastructure is based on EGEE middleware, and will deploy on it a selection of both ported EGEE applications, as well as applications that will be developed from the SEE-GRID partners.

The additional, to those of EGEE, SEE-GRID applications include at the moment the following two:

- A Search Engine for South-Eastern Europe (SE4SEE) application, which is an on-demand crawling and categorization engine for the SEE-GRID infrastructure. This application is meant to establish a basis for a grid-enabled information dissemination and sharing infrastructure.
- A Volumetric Image Visualization Environment (VIVE) application which is an interactive application for Web-based grid-enabled visualization of 3D medical data. VIVE aims to provide advanced visualization capabilities for users equipped with a standard Web browser by employing grid resources to store, access and process volumetric data.

SEE-GRID needs access to bandwidth allocation and reservation services which will be offered by the underlying network providers, i.e. the SEEREN network, the pan-European research network – GÉANT, and the participating National Research and Education Networks (NRENs).

As the SEE-GRID middleware will be the one developed in EGEE, the clients of these network services will be the higher layer EGEE Grid middleware, such as the Workload Management System, Data Replication Service etc. They will be referred to as the Higher Layer Middleware (HLM).

Additionally, EGEE develops the component of middleware referred to as BAR. The deployed applications will utilize the above middleware components. As an example, the VIVE application will primarily use services provided by EGEE Grid HLM to allocate and reserve bandwidth required for submission of 3D medical images to CE and for interactive visualization on user workstation of data produced by CE.

2.1.4 CrossGrid

The main feature of the applications of the project is that they are going to be compute-intensive and data-intensive applications that are characterized by the interaction with a person in a processing loop. Such applications require a response from the Grid to an action by a human agent in different time scales. Most of them are will require real-time behaviour and responses.

Services related to GRIDs which need special network requirement are:

- All "urgent request" communication services between websites.
- Exchange messages between processes, specially MPI
- Data transfer -> programmed or not programmed
- Consolidation data transfer
- Detectors data transfer and experiments
- Visualization

2.1.5 GRIDCC

Recent developments in Grid technologies have concentrated on providing batch access to distributed computational and storage resources. GRIDCC will extend this to include access to and control of distributed instrumentation.

Instruments work in real-time and their successful operation often requires rapid interaction with either conventional computing resources or other instruments. The control of instrumentation is often an interactive process. Both the real-time and interactive nature of instrument control mean that the extension of Grid technology into this area requires the definition of acceptable quality of service for the interactions between the different entities on the Grid.

GRIDCC will develop these definitions, where appropriate, building on work that is being carried out in the various standards organisations. Access to the instruments will be via an interface to a Virtual Instrument Grid Service (VIGS). VIGS are a new concept and their design and implementation will be a key part of the GRIDCC programme.

GRIDCC will, wherever possible, build on the work of other Grid projects whilst at the same time seeking to influence the developments within these projects. In this way GRIDCC will access existing Grid resources.

Control and monitoring of experiments. This application involves the use of the Grid in a real-time environment to control and monitor remote large-scale detectors. This application will make use of a High-Energy Physics (HEP) experiment, the CMS detector which is currently under construction at the future LHC collider at CERN. CMS consists of 20,000,000 electronics channels that will be read out by a complex distributed data acquisition (DAQ) system feeding a large processor farm charged with filtering an input rate of up to 100 kHz down to only ~100 Hz of physics events. The DAQ system involves a very large number (a few thousand) of intelligent modules and computers, data throughputs of ~100 Gbps. These characteristics, along with the selectivity of one event in 1,000, are unprecedented in the field, and introduce requirements on the control and monitor of the experiment's data-taking.

The so-called "Run Control" and "Detector Control Systems" of the experiment are charged with supervising the full configuration of the detector, but also with monitoring the data read out, their analysis and on-line interpretation. Both systems are of paramount importance for the correct operation of the experiment. Monitoring this detector, and potentially changing settings as a result of analysis on the monitoring data, is a complex task shared by a few hundred people distributed in geographically distributed laboratories. This task requires continuous analysis and display of large amounts of data generated by the detector and in the past was done in a counting room near to the detector. In the context of GRIDCC, this application will be made to run in a completely distributed fashion, over the Grid.

(Far) Remote Operation of Accelerator Facility. Far remote operation of an accelerator facility (i.e. the Elettra Control Room in Italy) involves the planning of accelerator operations, the maintenance of the accelerator and its troubleshooting, the repair of delicate equipment, understanding and pushing performance limitations, performing studies, performing commissioning and set ups and routine operations. All these activities are based on large amounts of information, which are at present accessible only at the accelerator site.

Remote control of an accelerator facility has the potential of revolutionising the mode of operation and the degree of exploitation of large experimental physics facilities. The involvement of one of the partners with the International Committee for Future Accelerators give also the possibility to present this application of GRIDCC to a potentially world wide community. Far remote operation combines elements of immerse (i.e. providing the feeling to be present at the remote location) communication and cooperation technology. This includes video and audio presence, allowing the simultaneous operation of the same instruments, having access to the same accelerator controls and the relevant data, meeting easily and spontaneously and providing full awareness of the presence of the collaborators.

The security and networking issues include adequate user management, session and floor control, and secure, synchronous and reliable data transmission and distribution.

Power Grid. In electrical utility networks (or power grids), the introduction of very large numbers of 'embedded' power generators often using renewable energy sources, creates a severe challenge for utility companies. Existing computer software technology for monitoring and control is not scalable and cannot provide a solution for the many thousands of generators that are anticipated. GRIDCC technology would allow the generators to participate in a Virtual Organization, and consequently to be monitored and scheduled in a cost-effective manner.

A specific test bed application will be built and demonstrated within the GRIDCC project by means of computer simulation and emulation. Existing software at Brunel University will allow the real-time simulation of a representative power network and the associated generators. New software will be created to interface the generator simulations to the GRIDCC environment. Distributed generator scheduling algorithms will be modified to utilize GRIDCC technology. The test bed will demonstrate the performance of the emulated system under various conditions, ranging from light power system loading (where energy economics is most important) to power system emergency conditions (where overloaded power circuits necessitate co-ordinated generator control to avoid power black-outs).

2.1.6 CANARIE

CANARIE's mission is to accelerate Canada's advanced Internet development and use by facilitating the widespread adoption of faster, more efficient networks and by enabling the next generation of advanced products, applications and services to run on them.

Across Canada and around the world there are many large remote sensor and instrumentation projects whose data output will need to be delivered to computing facilities many thousands of kilometres away. As well, researchers will want to remotely control these instruments and, from time to time, re-direct the flow of data to different computing destinations. Examples of such facilities include the European eVLBI project- JIVE, the US-Canada undersea network NEPTUNE, the Canadian Light Source Synchrotron project (CLS), the Ottawa University-NRC Nuclear Magnetic Resonance Instrument Facility (NRC-NMR) , and of course, the predecessor of them all, the CERN Large Hydron Collider (LHC).

Up to now the interconnection of these instruments to networks and remote computers was static and pre-configured by network or software engineers. Increasingly researchers need to dynamically control how these instruments are connected to the network and redirect or share the data flow amongst different geographically distributed researchers. As well they may want to control how the data is delivered as for example through an IP data flow with guaranteed parameters, or through an IPsec tunnel for secure transmission, or an independent managed lightpath for high volume transfers.

In some cases the expected data flow demands of these projects, in terms of throughput and duration will easily exceed by order of magnitude, that of conventional Internet traffic flows. In anticipation of these requirements CANARIE has been leading the development of User Controlled LightPaths in order to permit researchers to create “discipline” or “application” specific Layer 1 VPN or IP networks which would create separate high speed networks for these specific applications. As opposed to traditional VPNs the UCLP allows end users to do their own traffic engineering, change the topology of their VPN, and cross connect their VPN over multiple domains to VPNs operated by other users. Hence the moniker of “articulated” Private Networks.

2.2 Control Issues

This section discusses the control issues related to BoD service and its requirements. Three questions, raised to the user groups needing a BoD service, made it possible to determine these requirements in terms of control issues:

- Which are the forms of signalling interfaces required by the project application to interact with BoD service?
- What kind of responses are needed (if any) from the service reservation interface?
- What is the typical usage duration, frequency and day-time of the BoD service required by the application?

2.2.1 Overview

The main goal in JRA3 is to provide at first a streamlined manual but later on a dynamic provisioning system for end-to-end BoD reservations over multiple domains. In this way, GRID or advanced applications with guaranteed bandwidth requirements will be able to communicate directly with the BoD system and thus have a certain degree of control over the network.

In order to ensure communication between applications and the BoD system, two main factors must be taken into account: compliance between the signalling interfaces and the type of messages exchanged between applications and BoD system.

Generally, interfaces of applications rest mainly on technologies such as Web-page interfaces for the manual configuration and Web Services (WS) in the case of dynamic configuration. It will be important to know the set of signalling interfaces used by the potential applications in order to support them in the BoD system.

The other main point to consider is the type of responses awaited by the application from the BoD system. Two categories of messages exist: reservation messages and event notifications. Reservation messages permit to confirm or reject requests of user applications and sometimes, exchange other configuration parameters. From the application point of view, reservation messages can be very diverse, ranging from sophisticated UNI-like signalling interfaces to use of plain mailing services. In addition, event notifications allow to inform the concerned applications of unexpected changes in the BoD system and existing reservations. Event notifications are closely tied with the criticality of the application. For example, certain applications will have to be informed when there is a performance degradation while others will be informed only when there is a failure in the BoD reservation system.

Finally, it's important to know the usage duration of BoD applications as well as its frequency. Thus, it will be possible to characterize the profile of one or a set of applications and take into account these patterns in the design of the BoD system.

In order to obtain useful input for the design of the BoD system that JRA3 will develop in later stages, the expectations in terms of control from the users/applications of section 2.1 have been collected and are presented here.

2.2.2 Signalling interface

The form of signalling interface is very important, because via the signalling interface, the application needing BoD service will be able to communicate with the BoD Service. In order to anticipate possible implementations, it is crucial to consider what particular user groups can use. Overview of user group requirements is given in the **Table 2.1**.

User group	Signalling interface required by the application toward BoD service
EGEE	EGEE BAR Service needs to communicate with the BoD service via API;

User group	Signalling interface required by the application toward BoD service
	Use of Web Services if the latter can support asynchronous notifications messaging is also appropriate
Earthquake simulation project	N/A
SEE-GRID	Use of EGEE BAR API toward BoD service. Manual web interface to be discussed
CrossGrid	N/A
GRIDCC	Both automatic (Web Service) and manual through web page
CANARIE	Through a WSDL Web Service for automatic signalling; and through a Web portal for manual signalling

Table 2.1: Overview of required signalling interfaces toward BoD service

There are many possible forms of signalling interfaces for user applications: API, manual web interface, Web Services, etc. However, some GRID applications have or envisage to have two kinds of signalling interfaces (automatic and manual). Consequently, we can assume that it will not be necessary for the BoD service to implement all possible signalling interfaces.

Considering Web Services (WS), an interface mostly used by the user applications, we can perceive an interaction between the BoD service and applications based on WS for the automatic requests. A Web-page interface can also be useful for the manual configuration.

Other kinds of signalling interface would be implemented if WS and Web-page interfaces did not cover all the cases (it depends on the choice of some projects like EGEE or SEE-GRID that will make a decision with regard to the development of a new signalling interface in the next months).

2.2.3 Interaction between application and service reservation interface

In this section, it is examined whether a user application needs some kind of response from the service/user reservation interface. Overview of given answers is provided in the **Table 2.2**

User group	Interaction between user application and service reservation interface
EGEE	Permanent status messaging facility during BoD service usage (reroute, restoration, failure, duration, symmetric or asymmetric real amount of bandwidth…)
Earthquake simulation project	N/A
SEE-GRID	Permanent status messaging and event notifications
CrossGrid	Required: Event notifications

User group	Interaction between user application and service reservation interface
GRIDCC	Required: Confirmation of successful BoD reservation operation (WS-based for automatic requests or through the web interface for manual operations). Notifications on warning and error situations
CANARIE	All events should be sent by email.

Table 2.2: Interaction between user application and service reservation interface

In the BoD reservation phase, the BoD system will have to respond with a message of acceptance or refusal (and possibly more parameters) to a reservation request. It should also support permanent asynchronous events and notification messages (e.g. send failure messages to the applications). These events and notifications will be sent according to the signalling interface used by the application (WS, Web-page interface, email…). Additionally, the BoD system should be able to send event notifications by email even when other signalling interfaces such as WS are being used. To conclude, the BoD service is required to provide some form of a permanent status notification and management facility during its usage to each a particular application.

2.2.4 Duration and frequency of BoD service usage

This section analyses BoD requirements regarding typical usage durations, frequency and day-time availability of the BoD service required by the user application. An overview of given answers is provided in the **Table 2.3**.

User group	Time and frequency parameters of BoD service usage
EGEE	Hours, days, weeks or months are equally possible; repetition pattern; exceeding working hours (e.g. nightly, scheduled)
Earthquake simulation project	N/A
SEE-GRID	Session may last from hours to days and also weeks and months. Repetitive pattern is expected. Usage sometimes may exceed working hours. BoD provisioning response time in terms of seconds.
CrossGrid	In case of visualization: the session will last around 10-15 minutes.
GRIDCC	Control and monitor of experiments: application running approximately 8 months/year, continuously. (Far) Remote Operation of Accelerator Facility: in the range of 6-7 weeks (continuous) per experiment. Power Grid: 24x7 continuous traffic flow.
CANARIE	Minimum weeks - maximum year

Table 2.3: Duration and frequency of BoD service usage overview

If we analyse the answers, we can note that the usage duration is very variable depending on the application: Certain applications need occasionally 15 minutes of guaranteed bandwidth while others need almost permanently guaranteed bandwidth. It's very difficult to define a pattern for the whole set of applications; it is not even possible to define a certain profile for each of the applications since their needs in terms of BoD service vary a lot: for a given application, the usage duration ranges from a few hours to months. It is generally concluded that there are no daytime limitations for the usage of the BoD service, so it is required that the service is available on a 24x7 basis.

The only parameter we can more easily characterize is the frequency of requirement for BoD connectivity. The GRID applications repeat the same operation regularly, e.g. synchronization of data between distributed servers (the time of the operation cannot be estimated because it depends on the volume of data, but it is known that this operation will be repeated each month for example).

2.3 Network Issues

BoD requirements regarding the network infrastructure used, are discussed in this section. These are: bandwidth requirements, description of traffic patterns, description of available and future network infrastructure resources used and possible monitoring of service usage.

2.3.1 Overview

The Bandwidth on Demand service should empower potential user groups to take some level of control of available network resources, in order to get end-to-end guaranteed bandwidth, logically separated from the rest of the traffic sharing the network. For JRA3, to be able to design the BoD service, it is crucial to anticipate bandwidth requirements, in terms of size and granularity. From an initial view, it is perceived that the BoD service should deal only with bandwidth sizes from 10 Mbps and above (the upper limit, of course, is not set and it will be defined and continuously extended as new network technologies are implemented). The main reason for this is that bandwidth less then 10 Mbps falls into the scope of the new Premium IP service, which GN2-SA3 will develop. Also, the BoD service should define an environment for those user applications with such large bandwidth, latency sensitive requirements that they cannot be satisfied with Premium IP service.

Traffic patterns that user applications will create on the underlying network infrastructure would probably be of different types, since potential user groups have different application needs. It is important to be aware of the different traffic patterns expected, in order to take them into account when designing solutions which should be optimised as much as possible. For example, data that should be transferred could be in the form of a stream using maximal bandwidth for all the time, or burst transmissions with some pauses of some duration. Also, the size of data units is important to know when analysing possible network technology solutions.

User groups already use some network infrastructure resources for their needs. It is anticipated that this is built on different network technologies, both at the physical layer and at higher layers. Different physical interfaces may be implemented (gigabit Ethernet, SDH, optical lambdas, etc) on different parts of an end-to-end data transfer path (customer side, campus network, NREN, etc). In some cases, MPLS may be used in different layers (e.g. MPLS L3 or L2 VPNs). Also, the logical network topology depends on the number of end-points that are used in particular user applications (e.g. point-to-point, point-to-multipoint). All these network aspects are of particular relevance to the BoD service specification, since BoD service will be implemented on top of the network infrastructure available. Also, perceived future needs of BoD users (in terms of bandwidth, technologies supported etc) is valuable information to the design process of the BoD service within JRA3.

BoD service availability monitoring in distinct points of time, reliability characteristics during transmissions, communication load between two or more endpoints where the service is provisioned, and bandwidth which is actually achievable, may represent important issues for user applications signalling and monitoring process. In this section these issues and their importance for the BoD service deployment are examined and analysed.

2.3.2 Bandwidth requirements

Bandwidth and bandwidth granularity are probably the most important parameters that user applications need to specify as part of a BoD service request and the results from the survey are analysed here. An overview of the answers given is presented in **Table 2.4**.

	EGEE	Earthquake simulation project	SEE-GRID	CrossGrid	GRIDCC	CANARIE
Minimum	10Mbps	1 Gbps	1 Mbps	less than 1 Mbps	2 Mbps	1 Mbps
Maximum	Order of 1 or 10 Gbps	N/A	10 Mbps	Order of Gbps	800 Mbps	10 Gbps
Granularity	N/A	N/A	Order of 1 -10 Mbps	N/A	80–800 Mbps 2–50 Mbps 4–80 Mbps	N/A
Future need	N/A	N/A	1 Gbps	N/A	N/A	N/A

Table 2.4: Bandwidth requirements overview. Note: Earthquake simulation project have not yet managed to set up a suitable link between their labs in order to evaluate bandwidth requirements, so they rely on information from NEES (which requires a minimum of 1Gbps between labs).

Requirements presented should be categorised as required. The range of required bandwidth is large: ranging from 1 Mbps (or even less than that) to the order of 1 Gbps or even 10 Gbps. In some cases (e.g. GRIDCC), the application will need to reserve a maximum declared bandwidth because it exhibits a specific traffic pattern (a stream of a few Mbps with occasional bursts of unknown frequency that could be of the order of the declared maximum bandwidth size).

Note: from the answers given, it is not clear whether a minimum size of the order of 1 Mbps should be examined as a valid BoD requirement. But according to the described traffic patterns from some of the applications, it is likely that such values do not take into consideration occasional bursts that exceed the order of at least 10 Mbps.

In order to be as precise as possible, analysis conclusions for this requirement are done literally from the answers given. So, bandwidth size are in the range of 1 Mbps minimum to 10 Gbps maximum, with granularity in the order of 1, 10 , 100, 1000, 10000 Mbps.

2.3.3 Traffic patterns

In this section, description of possible traffic patterns is analysed. Users were given some possible traffic patterns as examples: stream using bandwidth for all the time, burst transmissions with some pauses, bursts and pauses duration, consideration about packet sizes, etc. Overview of given answers is presented in **Table 2.5**.

User group	Description of possible traffic patterns
EGEE	• mainly require bulk transfers of data over certain intervals alternating with pauses of transmission • stream traffic pattern and have real-time constraints • bandwidth range 10 Mbps – 1 or 10 Gbps • maximum expected packet size is 1.5 kB Notes: • for bulk transmissions it may be reasonable to perform an atomic couple of reservations in opposite directions with distinct parameters (data traffic, packet size ~ 1.5 kB, in one direction and control traffic, packet size ~0.1 kB, in other direction) • traffic pattern for bulk transmissions is expected to be flat or to consist of flat ranges • Globus/LCG applications use auxiliary buffers with default size of 1 MB for data exchange without additional SE. Thus, traffic burstiness equal to 1 MB are expected in some cases
Earthquake simulation project	• heavy demands for real-time video streaming • bandwidth range should be around 1 Gbps
SEE-GRID	• one or two burst transmissions with maximal bandwidth (order of 10Mbps to 1 Gbps) at the beginning of each session, followed by lower-bandwidth (around 1 Mbps) burst transmissions with real-time constraints and pauses; Bursts and pauses will fluctuate from several seconds to one minute • traffic generated due to crawling task induces long sessions of heavy bandwidth usage with no latency constraints (10 Mbps for quick download, 1Mbps minimal) • traffic between the application Web Portal and the grid nodes providing the crawling service may require vast amount of reliable data transfer (10 Mbps for quick download, 1Mbps minimal) • also see EGEE
CrossGrid	• permanent and temporal data transfers • bandwidth range from less than 1 Mbps to the order of 1 Gbps

User group	Description of possible traffic patterns
GRIDCC	• continuous traffic flow in the range of 80 Mbps, with bursts of unknown frequency (bandwidth range 80 Mbps – 800 Mbps) • continuous traffic flow in the range of 5 Mbps, with bursts of unknown frequency (bandwidth range 5 Mbps – 50 Mbps) • continuous traffic flow in the range of 4 Mbps, with bursts of unknown frequency (bandwidth range 4 Mbps – 80 Mbps)
CANARIE	• unspecified, all network traffic pattern (including bursts and bulk transfer) is possible • bandwidth range from 1 Mbps to 10 Gbps

Table 2.5: Description of possible traffic patterns

The service characteristics presented here should be classifiedi as 'required' (i.e. mandatory) since this is a description of possible traffic patterns. Having analysed the responses provided by the different user groups, it can be concluded that there is a wide range of traffic patterns. Most of the application cases exhibit traffic patterns in the form of bursty transmissions with some pauses. However, there are some cases of data streams with real-time constraints. There are not many answers regarding packet size (only one mentioning, 1.5 kB), so no conclusion can be made about that.

2.3.4 Description of network infrastructure resources

In this section, descriptions of the different network infrastructures and resources that are currently in use by the different user groups are presented. More specifically, the questionnaire asked the different projects and user groups about their current network connection (type, speed), the logical network topology (number of end points, single or multiple domains, point-to-point or point-to–multipoint), customer side equipment (L1, optical switches, L2, Ethernet switches, L3, IP router, PC) and physical interface types (gigabit ethernet, 10 GE LAN PHY, SDH, etc.). An overview of the answers that were provided is presented in **Table 2.6**

User group	Network infrastructure resources in use
EGEE	• 70 Resource Centres (RCs) are participating in EGEE. They are mostly located all over Europe except for a few that are outside of Europe (http://goc.grid.sinica.edu.tw/gstat/Region.html). These RCs are typically connected with 100 or 1000 Mbps ethernet to the national WAN. The national backbones are connected to GÉANT with speeds varying from a few 100s of Mbps to 10 Gbps. • EGEE is by nature a multiple end-points project which will need at least point-to-multipoint BoD services. • Customer side equipment consists mostly of PCs and PC-clusters.
Earthquake simulation project	The earthquake simulation project will be based on a connection with the US Network for Earthquake Engineering Simulation (NEES) (see http://www.nees.org/). However, it

User group	Network infrastructure resources in use
	will also develop it to suit the European needs.
SEE-GRID	The infrastructure for SEE-GRID is based on the SEEREN network, which connects the NRENs of Albania (INIMA), FYR of Macedonia (MARNET), Serbia & Montenegro (AMREJ) and Bulgaria (ISTF) with GRNET, the NREN of Greece. A backup link connecting GRNET to RoEduNet, the NREN of Romania, is also available. The link capacities are as follows: • Tirana - Athens: 2 Mbps • Skopje - Athens: 4 Mbps • Belgrade - Athens: 34 Mbps • Sofia - Athens: 34 Mbps • Bucharest - Athens: 34 Mbps A large variety of equipment and diverse technologies are used throughout the network (e.g. Gigabit Ethernet, ATM, MPLS, optical switches, IP routers, etc).
CrossGrid	In general there will be multiple domains.
GRIDCC	Two types of logical topologies: • Fast or gigabit ethernet (LAN), multiple domains, point-to-point, two endpoints. • Point-to-multipoint, multiple domains.
CANARIE	Mainly SONET/STS channels and wavelengths.

Table 2.6: Network infrastructure resources in use

Analysing the provided answers, it can be concluded that most of the examined user groups use quite similar network infrastructure. In general, most user groups or projects have fast or gigabit ethernet connections to the national WAN. The customer end site consists usually of PCs and PC clusters. Most national WANs in Europe are connected to networks such as GÉANT or SEEREN to provide worldwide connectivity. The connection uplink speeds of National WANs are varying between 2 Mbps (SEEREN) and 10 Gbps (e.g. GEANT). A large variety of physical interface types and diverse technologies are used (Gigabit Ethernet, ATM, SONET, optical switching, MPLS). An exception is the CANARIE project/network, which is using mainly SONET/STS channels and wavelengths.

It is worth mentioning that some user groups are or will span across multiple continents. Regarding the logical network topology, user groups require both point-to-point and point-to-multipoint types. Typically, across multiple (administrative) domains.

2.3.5 Description of future network infrastructure resources

In this section, future requirements on the network infrastructure resources for the different user groups are presented. The parameters that were considered are the same as in the previous section.

An overview of provided answers is presented in **Table 2.7**

User group	Future network infrastructure resources
EGEE	• Not expected to change drastically in the near future. • It is expected that low bandwidth access links towards the NRENs in the order of 100Mbps will co-exist with 10Gbps access links or direct links between resource centres. • The customer end site equipment will continue to be PCs and PC-clusters, but it is also possible that optical and ethernet switches are deployed at the customer side in a few cases.
Earthquake simulation project	N/A
SEE-GRID	• A network infrastructure upgrade is expected in the near future (possibly starting around September 2005). This will result in the SEEREN2 network. A first estimation is that the link capacity will vary between 2 and 155 Mbps and will be utilising MPLS in the network core. • In the long term, the acquisition of dark fibre will be pursued to serve the NREN interconnection needs.
CrossGrid	Multiple domains, mainly in Europe and the USA.
GRIDCC	N/A
CANARIE	N/A

Table 2.7: Network infrastructure resources – future developments

It can be concluded that most of the user groups do not perceive significant changes in the available network infrastructure resources in the future or they do not have information about it. An exception is the SEE-GRID project that uses the SEEREN network, which will be upgraded to SEEREN2. This network will be built in the near future (September 2005). The biggest change is mainly the increase of NREN connection speeds. In the longer term, the SEEREN network might use dark fibre to serve NREN interconnection needs.

Also, it is worth mentioning that the EGEE project indicated the possible deployment of optical switches at the customer end site.

2.3.6 Bandwidth usage monitoring

In this section, a description of the required or desirable monitoring for bandwidth usage is analysed. An overview of the given answers is presented in **Table 2.8**.

User group	Bandwidth usage monitoring
EGEE	EGEE will require the following types of BoD services monitoring: • Monitoring of the availability of the service • Monitoring of the QoS characteristics (loss, delay, jitter) along the end-to-end path • Dynamic real-time information about the traffic load between two or more endpoints where the service is provisioned • Estimating the achievable bandwidth between two or more end-points
Earthquake simulation project	Need real-time control.
SEE-GRID	Note: See EGEE. SEE-GRID will make use of the same performance metrics as EGEE. These metrics should be monitored and recorded for each BoD service transaction. The aggregate usage statistics at the BoD access point of each SEE-GRID site should include: The number and type of the established paths, the allocated bandwidth and the bandwidth utilisation. A usage break down by virtual organisation and by user would be highly preferred. The monitoring information should be available to assist the operation centre staff with troubleshooting.
CrossGrid	An owned GRID monitoring system is required. There are two alternatives, one of them is to have a permanent circuit for this purpose, the other could be an open channel during the interval in which the Resource Broker exchanges control messages in order to send the job correctly.
GRIDCC	Desirable: • a web application providing traffic charts, as well as averages per hour of the day, day of the week, day of the month, month, year. • a log of the BoD requests and responses as well as warning/error logs database is considered to be extremely useful.
CANARIE	Not needed.

Table 2.8: Bandwidth usage monitoring requirements overview

Some user groups (EGEE) provided a set of requirements for monitoring capabilities that have to be taken into account when designing a BoD service. These capabilities are: monitoring of availability of the service, monitoring of QoS characteristics (loss, jitter, delay) along end-to-end path, dynamic real-time information of the communication load between two or more endpoints where the service is provisioned, and estimating the achievable bandwidth between two or more end-points.

Besides that, one user group mentioned two highly desirable monitoring capabilities: a web application that provides statistics about the BoD usage (traffic charts, as well as averages per hour of the day, day of the week, day of the month, month, year), and a log of the BoD requests and responses as well as a warning/error logs database.

2.4 Reliability Issues

In this section the results from the Bandwidth on Demand (BoD) questionnaire regarding reliability are discussed. First, an overview of the different issues that may impact an application relating to reliability is provided. The reader is referred to appendix A for the exact questions asked to the different projects and in particular, to the section 'Reliability Issues'. Secondly, the outcome of the questionnaire will be presented and discussed. This section concludes with highlighting the reliability requirements that have to be considered during the development of a bandwidth on demand system in JRA3.

2.4.1 Overview

Reliability of a service is closely tied with the reliability of the network links and network equipment. There are two important parameters to consider in the context of reliability:

- Mean Time Between Failures (MTBF), is the average time between two subsequent failures.
- Main Time To Repair (MTTR), is the average time needed to repair.

Note that the reliability, and hence the availability, of a service is also impacted by planned maintenance if this maintenance operation respects the planned slot time and the planned impact. However, planned maintenance is often not taken into account for reliability statistics.

Protection and restoration mechanisms may be required to attain the necessary level of network service reliability. These mechanisms are often standardised for a particular network technology (e.g. SDH, MPLS).

Protection mechanisms are able to repair faults within very small time frames (e.g. 50ms in the case of SDH). This, however, comes at the expense of statically provisioning protection link requiring spare capacity and interfaces in the network. On the other hand, restoration mechanisms are able to repair faults within time frames of several 100s of milliseconds. Restoration mechanisms, however, do not require statically provisioned resources. It relies on dynamically allocating new resources on a different path to restore the connection. In case there are no network resources available the network service cannot be restored.

In addition, the behaviour after a fault has been repaired may be of concern to some applications. For instance, should a connection revert back to the primary path or should it remain on the protection/restoration path when the primary path has been repaired? A second question then becomes whether such reversion should take place under the control of the network operator, the application or may be just automatically after a pre-configured time period. Changing the path of a certain connection will change the one-way delay characteristic, which may have a negative impact on some applications. In addition, there may be some small loss and frame/packet reordering during this process.

Security is often deemed to be the responsibility of the endpoints. However, in the context of the questionnaire this issue is interpreted in a broader sense. For instance, the application may provide a hint to the BoD service provisioning system indicating that security is of concern. This hint may be used to constrain the routing of the path within a set of highly trusted networks. Alternatively, it may only use certain technologies for provisioning the service.

2.4.2 Results of the BoD questionnaire

The results from the BoD questionnaire are summarised in Table 2.9. It should be noted that information from the Earthquake simulation project was too short to be processed and inserted into the table as a full entry. Nevertheless, it is mentioned when referring to the latency issue. At a first glance, it is clear that the requirements for the different projects and applications vary widely. Hence, the JRA3 service request should provide the necessary options for requesting properties such as restoration.

From the responses regarding the reliability parameters it is clear that frame loss and reordering should be minimised as much as possible. Some projects have not explicitly indicated their need for a service with no loss and reordering. However, it can be assumed that most grid applications will require no loss and reordering. Therefore, a requirement for the BoD service is to provide a service with no frame loss and reordering.

Sensitivity to latency largely depends on the type of application. The CrossGrid project, for instance, indicated that they were mainly concerned about latency for signalling and control traffic. Also, Earthquake simulation project indicated that time delays are critical for control algorithms (must be shorter than 0.2ms which is determined by the sample rate, 5kHz). Therefore, a requirement of the BoD service is to provide an option that allows optimising the route of the physical path in order to minimise latency and round trip times.

Security can be implemented on the end host, which will provide true end-to-end security such as confidentiality and integrity. This approach was indicated in the response from CANARIE and GRIDCC. However, the network may implement security as well. This could be in the form of a security tunnel providing security on part of the path. Alternatively, the application could provide a flag in the service request indicating that security is of concern. This then can be taken into account by e.g. restricting the route of the path to a limited set of networks or technologies. However, from the responses it is clear that confidentiality (i.e. encryption) is the largest security requirement. Projects that need encryption are EGEE, CrossGrid and SEE-GRID. Therefore, a requirement for the BoD service is that it should be able to provide confidentiality.

Restoration is necessary for applications from the EGEE, GRIDCC and SEE-GRID projects. The response from the CrossGrid project did not indicate whether they require restoration or not. Only the response from CANARIE indicated that network restoration is not required. Their approach relies probably on establishing two lightpaths for resilience purposes. Note that this will require the capability to request diversely routed paths. The application can then decide which path to use by e.g. using routing protocols. However, it is clear from the responses that the BoD service should have an option to provide restoration.

Reversion is the process of switching from the protection path back to the primary path. This process can be controlled by the network using a switchback timer. This approach is preferred by the SEE-GRID project.

Another approach is to have it controlled by the end user or application. This approach is preferred by EGEE and GRIDCC. From the responses it seems that the BoD should be able to cater for both approaches.

The restoration/protection time required by the different projects varies between 50 ms and 5 s. This time to restore a path will determine the types of protection and restoration mechanisms that are required within the network. No indication for restoration times was provided in the response from the CrossGrid project. The resulting requirement is that an application should be able to indicate the required restoration time from the following values: 50 ms, 0.5 s, 1 s, 5 s and more than 5 s. This could possibly be a way to distinguish between different classes of BoD service provided, with the premium one being that with the minimum restoration time.

An overview of the responses to the questionnaire regarding reliability is provided in **Table 2.9**, **Table 2.10** and **Table 2.11**.

	Quality parameters			
	Latency sensitive	**Loss sensitive**	**Sensitive reordering**	**Other**
EGEE	No	Yes	Yes	N/A
GRIDCC	Highly	Depends on the application	Unknown	Bandwidth requirements are modest
CrossGrid	Yes, in particular for signalling traffic	Unknown	Unknown	N/A
SEE-GRID	Sensitive to changes in round trip time	Yes	Yes	N/A
CANARIE	No	No	No	N/A
Eartquake simulation project	Yes, critical for control algorithms	N/A	N/A	N/A

Table 2.9: Quality parameters overview

User group	Network security
EGEE	Network encryption is suggested (e.g. IPSec), however, it is not sure whether this is a network requirement.
GRIDCC	No security support is expected from the network.
CrossGrid	Information confidentiality is important. In particular for the medical visualisation application.
SEE-GRID	The application needs network encryption.
CANARIE	Security should always be at the edges with the application.

Table 2.10: Security issues overview

	Network path restoration			
	Restoration	**Reversion**	**Restoration time**	**Other**
EGEE	Yes	Should be coordinated by the user and service.	If possible, it can be as low as 50 ms in some cases.	Guaranteed capacity.
GRIDCC	Yes	The path re-establishment decision should be left to the user at the time of failure, with a timeout option. In case of a timeout, a pre-defined option (per application) should be used.	Restoration times depend on the application: (i) control and monitor of experiments: 50 ms - 0.5 s (ii) remote operation of accelerator facility: 0.2 – 1 s (iii) power grid: 0.5 – 5 s.	N/A
CrossGrid	Unknown	Unknown	Unknown	Request messages and signalling require a permanent connection.
SEE-GRID	Yes	Should be hidden from the application and end user. Best effort forwarding should be used until the path can be re-established.	Between 0.5 s and 5 s	N/A
CANARIE	No	N/A	N/A	It is an end user decision what to do in case of a failure.

Table 2.11: Reliability issues overview

2.4.3 Reliability requirements

In the previous paragraph the responses from the questionnaire have been discussed and the different requirements were identified. In this paragraph the requirements are summarised and a priority is assigned to each of them, which can be 'required' or 'desirable'. Note that the assignment of priorities may be subject to change during the development phase of JRA3.

The following six requirements can be derived from the responses on the reliability issues:

- The BoD service should guarantee no frame loss and reordering (required).

- The BoD service should provide an option to optimise the latency (desirable).
- The BoD service should provide an option to request confidentiality of information (required).
- The BoD service should provide an option to request restoration in the network (required).
- It should be selectable whether reversion is controlled by the network or by the end user or application (desirable).
- The restoration time should be configurable between 50 ms and 5 s (desirable).

Obviously, the feasibility to implement many of these service requirements will depend on the underlying network infrastructure and technology of the GÉANT2 network as well as the NRENs' networks. Because the number of responses has been small and level of detail was limited the requirements may be refined in the future depending on the needs of the different projects.

2.5 Other Issues

This section discusses the issues that are not related directly to the BoD service or the network infrastructure and its requirements. The subjects are threefold:

- Extended requirements, which do not fit into the network or control category. This includes future plans for the project evolution and the envisaged future BoD requirements,
- Interest in the BoD usage overview for statistic and accounting reasons,
- Interest in participating in the JRA3 BoD service pilot.

2.5.1 Overview

Some of the projects and organizations that were asked for input and requirements for the JRA3 BoD activity, are still developing and growing. The changes concerning JRA3, apart from the network infrastructure improvements,have mainly to do with the projected requirements on the BoD capacity and the additional BoD functionality in the future. Such issues are important for a more long-term planning and design specification of the BoD system in order to accommodate future requirements without major re-working.

Other important issues are accounting and monitoring features, which allow to investigate service usage. The reservation process and the client activities can be stored for monitoring and security purposes. Access to such functionality makes user quota management easier, may provide available resource tracking and will provide a statistical view of the service and the network utilisation. The latter feature seems to be the most useful.

Because the BoD service will be designed and implemented based on requirements collected from various NRENs, projects and institutions, it is crucial to test the implementation within their environments. The access to a diversity of network technologies, with real life limitations and political constraints, ensures a huge and diverse enough test environment for the final product. The experience gained from test participants will have a strong influence on the BoD service functionality. It also will evaluate the performance and provide motivation for future improvements. For these reasons, the contacted communities were asked to provide their availability in terms of acting as test-users of the JRA3 BoD service.

2.5.2 Extended requirements and future plans

This section presents the extended requirements that were stated by the project representatives, which were not mentioned in the control and network part of this document. A strong emphasis was placed on future project work, and future requirements, which should be taken into consideration during the design of the JRA3 service architecture and features definition. The collected information is shown in **Table 2.12**.

User group	Extended requirements and future plans
EGEE	It is envisaged that in the next year the requirements for the BoD capacity on behalf of EGEE will be more or less predictable and in the order of a few Gbps transmissions on a monthly basis. However in the two consecutive years, it is envisaged that the requirements in terms of bulk BoD capacity will increase exponentially with BoD connections of one or multiple GE being a daily requirement.
Earthquake simulation project	N/A
SEE-GRID	N/A
CrossGrid	N/A
GRIDCC	GRIDCC applications should have this test bed / infrastructure with the described characteristics available by the end of 2006, as the project ends in 2007 and it is expected to reach production-level results.
CANARIE	N/A

Table 2.12: Extended requirements and future plans overview

In general, the EGEE project foresees improvements in network bandwidth requirements and usage. Within a few years the bulk data transfer may consume one or multiple gigabit ethernet for a single connection for long term durations. The GRIDCC project is currently developing its network infrastructure, which should be ready by the end of 2006. The other projects did not provide any information about their future plans.

2.5.3 BoD usage overview

This paragraph discusses the requirement for accessing statistical information about the BoD service usage. Project representatives have declared whether they are interested in accounting and users activity monitoring. The results are represented in **Table 2.13**

User group	Access to BoD service usage information
EGEE	Yes, this is a very useful feature. It would be nice to see such data for each EGEE request. It would also be preferable to integrate the exchange of such information within the EGEE BAR service. The service should provide the information about currently available resources which are not allocated yet. BAR should provide the user with the information of currently available resources for his future requests in the context of his SLA.
Earthquake simulation project	N/A
SEE-GRID	This feature is not vital for the SEE-GRID project, although it may be useful.
CrossGrid	N/A
GRIDCC	A web-interface to provide such an overview is considered to be useful but not critical. In case it is provided, the total capacity requested, as well as the capacity per request and the total duration, including relevant timestamps is expected to suffice. Furthermore, failures to provide the service with the relevant statistics of the reasons for such failures and their [the failures'] time distribution over the period of a day, week, month, year, is expected to be very useful.
CANARIE	BoD usage overview is required.

Table 2.13: Access to BoD service usage information

Certainly the BoD usage overview is recommended for the BoD system. Most of the projects, especially EGEE and GRIDCC, claim that this feature is useful and they want to have access to statistical information about the users activity and service utilization. Moreover, the EGEE project pointed out that the JRA3 accounting data should be exchanged with their internal BAR service data. According to GRIDCC, information should be available through a web-interface and should include reservation constraints (like bandwidth capacity, reservation duration) and failures information. Earthquake and CrossGrid projects did not declare any interest in the BoD service usage information system. Summarising, a BoD usage functionality is considered quite useful but is not considered as being critical for the initial stages of the BoD system development.

2.5.4 BoD service test participation

This section contains the responses to the question whether projects are interested in participating in testing the JRA3 BoD service. **Table 2.14** contains an overview of the responses.

User group	BoD service test participation

User group	BoD service test participation
EGEE	Yes, EGEE would eagerly be a tester of the GN2 BoD service.
Earthquake simulation project	N/A
SEE-GRID	It is definitely an interesting option and they would like to be considered as candidate testers. But it would be necessary to understand what the testing would involve. e.g. what type of administrative rights would GN2 require on the participating site and NREN's network equipment.
CrossGrid	N/A
GRIDCC	Interested in being tester, involving all three parts of the GRIDCC project (control and monitoring of experiments, remote operation of accelerator facility and power grids).
CANARIE	If this service will be based on the UCLP system, then yes.

Table 2.14: BoD service test participation

Most of the projects have declared their interest in being a tester for JRA3 BoD service. SEE-GRID representatives have raised an important issue about administrative rights on equipment used for test, which may not belong to GN2 infrastructure. Due to the high level of interest, the test environment, permissions and test scenarios should be cautiously considered during further JRA3 activity. CANARIE declared their readiness for test participation, under the condition that UCLP is being used as a basis for a single domain BoD management.

3 Conclusions

A significant prerequisite for deriving qualitative BoD requirements is to acquire a representative sample of all potential user groups which would be involved in using a BoD production service. Unfortunately, although more than a dozen potential user groups were identified by the JRA3 working group and asked for contribution by filling out the prepared BoD requirements questionnaire, only five of them replied in the requested manner, or, more precisely, six with one that provided some basic information, but not in the required form. Of those six potential BoD users, four are typical representatives of the GRID community, one is more a successful example of implementation of a service with similar purposes, and one (which provided much summarized information) is a non-GRID representative. To conclude, it is highly recommended to try to get information from other potential user groups during initial project phases, especially during planned revision process of BoD requirements. Also, the user groups that did respond may be able to provide additional and more detailed information during the course of the JRA3 activity. Face to face interviews may be required to obtain this additional information. The information collected and analysed in this document has not been adequate in order to make a thorough comparative analysis of requirements. Therefore analysis was limited to the collection of information about particular topics that were asked, and less on comparison.

From the analysis carried out, there is a general conclusion that the need for BoD-like services exists today for a number of different communities/projects/users groups both in Europe and worldwide. It is expected that this need will become even more apparent in the upcoming years, when such a service should be mature enough to cater for BoD needs in a homogeneous and dynamic way. A number of Grid-related projects are in progress at the moment and it is expected that they will soon be mature enough to use the underlying networking infrastructure for their data transfer needs in a production environment. The usability of the planned BoD service will play a significant role for these projects.

In terms of control, BoD users/applications should be provided with standardized interfaces for resource reservations and service monitoring along the end-to-end BoD provisioning path. Signalling interfaces can be static/manual at the early stages of BoD service provisioning, but dynamic signalling will eventually be necessary. The interfaces should support more than a simple reservation request posting and processing. They should provide the means for the applications/Grid middleware to constantly interact with the BoD service modules in order to retrieve or receive event notifications and statistics of service usage. Grid applications requirements in terms of the required BoD service duration have proven to be quite diverse so it is estimated that it will be difficult to devise different classes based on time and frequency of BoD service usage. The BoD system should support a continuous space of duration intervals and frequencies for the service requests.

Bandwidth requirements range from 10 Mbps to 10 Gbps. Although 1Mbps requirements exist it is not recommended to incorporate them in the JRA3 BoD system, since they are likely not to take into consideration occasional bursts. In terms of the anticipated traffic patterns, the Grid application behaviour seems again to be quite diverse, ranging from a few high bandwidth burst transmissions to continuous traffic flows at low speeds often with real time constraints. Again here, the BoD system should support such different traffic types. It is worth investigating, during the design phase of JRA3, whether a set of different classes into which traffic patterns can be categorized is achievable and/or advantageous for the purposes of the BoD resource management implementation.

Concerning the network environment over which the multi-domain BoD service will be deployed, the requirements collection and analysis reveals a large variety of physical interfaces and technologies (Gigabit Ethernet, ATM, SONET, optical switching, MPLS-based overlay topologies etc.). This makes the work of JRA3 quite demanding in terms of the technical solutions for BoD that have to be supported in order to succeed in providing end-to-end BoD services to multiple projects/users. Besides that, it is a requirement from the JRA3 BoD service to provide the mechanisms and the visualisation components for monitoring the networking characteristics of the service, i.e. availability of each end-to-end BoD service instance, quality of the service provided, accurate estimated of the achieved bandwidth when compared to the promised guarantees etc. As the user perceived quality seems to be one of the crucial factors for the success of a service, monitoring is suggested to be a basic component of the BoD service from its early stages. In terms of reliability of each BoD service instance a number of priorities have been derived:

- It is required that no loss or reordering of frames is ensured, that restoration of a failing BoD service instance can be provided as an option and that confidentiality of transferred data is possible
- It is desirable to optimise the latency, to configure the restoration time from short to longer intervals and to control reversion in case of failure from the user/application environment

Finally, a number of the contacted projects/user groups have expressed their interest and/or offered to test the BoD system during its deployment phases in an effort to provide feedback to the work of JRA3 and benefit from the experience of using the service and tailoring their applications to use it effectively. JRA3 will remain in close cooperation with those groups for the purposes of its testing and validation activities.

Although it has already provided a first set of recommendations for the BoD system design, the analysis of BoD requirements of users and applications will be an ongoing task within JRA3 in order to incorporate the requirements of more projects and communities, that were not ready to provide such input at this stage, and to consider additional requirements and their refinements from the projects fully interviewed at this first stage.

Appendix A BoD requirements' questionnaire

This appendix provides the contents of BoD requirements questionnaire, as it was distributed to the targeted user groups.

A.1 Bandwidth on Demand service introduction

A.1.1 GN2 project and JRA3

The GN2 project is a four-year project funded by the European Commission that has started in September 2004. Partners in this project are nearly all of the European national Research and Education Networks (NRENs) and DANTE. The project will consist of a number of R&D activities (joint research activities) and a significant upgrade of the GÉANT network. A switched infrastructure will be deployed in addition to the routed IP network to enable reserved, point-to-point, high bandwidth services.

The Joint Research Activity 3 (JRA3) is focused on specifying, developing and piloting a Bandwidth Allocation and Reservation or Bandwidth on Demand (BoD) service. In particular, it will design and implement the necessary bandwidth reservation systems and signalling interfaces that will allow end users to make advance reservations with automated provisioning. The aim of JRA3 is to have a pilot service running at the end of the project.

A.1.2 Bandwidth on Demand service

A BoD service provides an end-to-end connection with a guaranteed capacity possibly spanning different administrative domains using connection-oriented technologies. This is sometimes also referred to as a 'lightpath' although it is envisaged that the BoD service developed in JRA3 will not be exclusively based on all-optical transmission. It is also envisaged that the service will be automated to a large extent the intention here being to give the end-user, middleware or application a certain degree of control over the network.

The European networking environment consists of a large number of networks, which are highly likely to deploy different networking technologies. Indeed, the BoD service has to operate in a technology heterogeneous environment. The following technologies are considered by JRA3:

1. Packet switching, MPLS preferably augmented with Quality of Service (QoS), Ethernet,
2. TDM switching, for example SDH or SONET,
3. Wavelength switching,
4. Fibre switching.

Packet or TDM switching are the most likely technologies to be available to implement BoD services in the nearer term. Wavelength and fibre switching technologies will be taken into account but are less likely to be widely deployed and available for use by an inter-domain BoD service during the course of JRA3.

Due to the variety of network technologies the parameters that are related to the service will differ. For instance, packet loss and packet jitter are parameters that are of concern for MPLS enabled networks but not for SDH networks. On the other hand, parameters such as delay and capacity (reserved bandwidth) are more universal.

A.2 Questionnaire

The questionnaire presented in this document has been created in order to gather input for the requirements of BoD services that will be best suited to serve the European R&E networking community and hence on the networks and systems that will be required to deliver such services. We would kindly request that the questionnaire be completed for each application (e.g. eVLBI, Grids, etc.) that can be considered to be a potential user of BoD services. The results of this questionnaire will provide valuable input to the requirements specification work currently being undertaken by JRA3.

A.2.1 General Issues

- Please briefly describe the nature of the application, what is its main purpose?

A.2.2 Control Issues

- Please specify what kind of signalling interface is required by the application toward BoD service?
  - Example:
    - Interaction with BoD reservation system - automatic–web service, API, manually-using web page or application

- Does the application need some kind of response from the service/user reservation interface?
  - Example:
    - Types of responses - e-mail, messages via the reservation interface, etc
    - Permanent status messaging - status reports or messages on usage of the BoD service
    - Event notifications - service start/stop, service failure, re-route, etc
- Please indicate the typical usage durations, frequency and day-time of the BoD service required by the application described above?
  - Example:
    - Usage duration - hours, days, weeks or months.
    - Request frequency - single transmission or repetitively
    - Daily usage - working hours (please indicate time zone), working days, non-stop

A.2.3 Network Issues

- Define bandwidth requirements for your application.
  - Example:
    - Minimum and maximum bandwidth size
    - Granularity magnitude of reservation bandwidth- 1 Mbps-10 Mbps-100 Mbps-1 Gbps-10 Gbps
- Please describe what traffic pattern is generated by application.
  - Example:
    - Traffic pattern - stream using maximal bandwidth for all the time; burst transmissions with some pauses, bursts and pauses duration
    - Packet size
- Please describe any network infrastructure resources that you are already using for the application/project
  - Example:
    - current connection – type, speed
    - Logical network topology - number of end points, single or multiple domains, point-to-point or multipoint
    - Costumer side equipment – Layer 1 (optical switches), Layer 2 (Ethernet switches), Layer 3 (IP router, PC)
    - Physical interface - GE, 10 GE, SDH, etc.
- Please describe network infrastructure resources you envisage or will require for the application/project in future.
  - Example:
    - Network infrastructure resources – data transport connection (type and speed), logical network topology (number of end points, single or multiple domains, point-to-point or multipoint)
    - Costumer side equipment – Layer 1 (optical switches), Layer 2 (Ethernet switches), Layer 3 (IP router, PC)
    - Physical interface - GE, 10 GE, SDH, etc.
    - Planned locations – organizations, countries, continents
- If needed, please describe what kind of bandwidth usage monitoring is needed?
  - Example:

— Types of monitoring - specific application, web

A.2.4 Reliability Issues

- Do you consider or use applications that need BoD service which have latency constraints? Please consider here also other reliability associated parameters.
  - Example:
    - Latency constraints - sensitivity to the one-way delay or the round trip time
    - Sensitivity on discontinuous changes in latency - e.g. when a protected service re-routes during a network failure condition
    - Reliability associated parameters - round trip time, loss, reordering, etc.
- Describe if any level of additional security from network infrastructure would be helpful for the application
  - Example:
    - Need for additional security provided by BoD service – e.g. application itself does not provide necessary security
    - Need for traffic separation – e.g. some level of traffic separation from other traffic
- Describe what level of network reliability your application needs. Do you or application need to control situations when optimal path is down, and the new, re-routed path is not within reliable network parameters?
  - Example:
    - Active path restoration – required or not
    - Path re-establishment decision – made by user or application
    - Restoration time preferred - <50 ms; 50 ms–0.5 s; 0.5 s–5 s; 5 s–1 hr; 1–24 hrs; 1 day–1 week
    - Path re-establishment scenarios - decision left to user; find another path; stop data transfer until path is ready; provide best effort; keep best effort to the transmission end or till path can be re-established; etc.

A.2.5 Other Issues

- Please, mention here some other requirements that application may have on BoD service and planned evolution of your needs - utilization projection for the next years, if such exist
- In order to manage access rights to BoD service an accounting system will be implemented. Please indicate if you would be interested in seeing an overview of BoD service usage concerning your account.
  - Example:
    - Parameters of interest – capacity, total duration, start time, end time, type of operation (create, cancel, etc.)
    - Types of interfaces - web interface, specific application, etc.
- Please indicate if you would be interested to become testers of the BoD service, sometime during its development

- End of Questionnaire -