Grid and Cloud Computing Integration with NGN

TATIANA KOVACIKOVA
on behalf of ETSI STF 331*
Department of InfoCom Networks
University of Zilina
Univerzitna 8215/1, Zilina
SLOVAKIA
tatiana.kovacikova@fri.uniza.sk http://www.uniza.sk

Abstract: - Over the past few years, Grid computing has become an increasingly hot topic in the world of new technologies. It is gradually also becoming familiar to the mass public through scientific and industrial field trials. Grid computing has been adopted and is used by academia and industry due to its high potential for improving manageability, utilizations, and performance of geographically distributed and diverse resource centers. During 2008 and 2009 commercial interest shifted from “grids” to “clouds”, with the availability of several on-demand compute and storage resources. Recently, interest has been growing to integrate grid and cloud computing with the Next Generation Network (NGN) in the telecommunication domain. Telecom operators are bound to become key players in a grid and cloud computing value chain as they provide connectivity and own computing resources. Moreover they have established customer relationships and accounting/billing experience, essential for business/commercial grids and clouds. Defining the best ways to integrate existing as well as future telecommunication equipment and network infrastructure with grids and clouds assumes the availability of interoperable Grid solutions built by IT in conjunction with the Telecom industry. This paper describes the approach taken by ETSI Technical Committee for Grid computing (TC GRID) and STF 331 (ICT GRID Technologies Interoperability and Standardization) to support the integration of grid and cloud computing with the NGN architecture.

Key-Words: - Integration, Standardization, Grid, Cloud computing, Next Generation Network (NGN)

1 Introduction

Grid and cloud computing may have many application fields. The first may be the improvement of performance and the reduction of costs thanks to combining of resources. The possibility of creating virtual organisations to establish collaboration between teams with scarce and costly data and resources is another option. Scientists who use applications that require enormous resources in terms of computing or data processing, are large consumers of grid and cloud computing. Grid computing is largely used in particle-physics experiments. Nor are leading industries staying behind: Grid computing is massively present in the automobile and aeronautical business, where digital simulation plays an important part. More recently, grid and cloud computing have emerged in other areas with the purpose of optimising company business. The aim is to combine resources for several services by reallocating them in a dynamic way depending on performance peaks. This strategy offers considerable cost cutting thanks to better management of resources, administrative tasks and maintenance.

For telecom operators, the future lies in converging fixed, mobile and IP telephone services into a platform known as the NGN. Integrating new services at attractive rates is further made difficult by the lack of interoperability. Grid and cloud computing therefore represent a considerable stake. Potentially, they could enable telecom operators to manage all of their resources in a dynamic and optimal way through one single platform. Their importance on the one hand therefore lies in the possibility to economise on a large scale by optimising resource management, and on the other hand in operators’ ability to offer innovative services.

The European Commission, in its 2006 ICT Standardisation Work Programme invited ETSI to carry out a study on Interoperability & Validation of International Open Grid Standards and to consider the impact of these standards on the NGN. This activity on the global convergence and evolution of IT infrastructure and electronic communications transformation is aligned with the i2010 initiative, NESSI ETP (Networked European Softwared & Services Initiative European Technology Platform) and the Next Generation GRID (NGG) vision of the SOKU (Service Oriented Knowledge Utility). The goal is to actively support and involve GRID stakeholders in the standardization of GRID test specifications in the IT-Telecom converged world.

ETSI GRID Technical Committee (TC GRID) is seeking to contribute to an improvement in worldwide co-operation in the ICT-Telecom collaborative Grid standardization efforts. The objective of the ETSI STF 331* (ICT GRID Technologies Interoperability and Standardization) [1] funded by the EC/EFTA and working under the control of the ETSI TC GRID is to address, in general, IT-Telecom (Information Technology -Telecommunications) convergence and, in particular, the lack of interoperable Grid solutions built by IT in conjunction with the telecom industry.

In this paper, the focus is on the integration of grid and cloud computing with the NGN architecture. The paper is structured as follows: Section 2.1 provides a brief introduction to grid and cloud computing. In section 2.2, a grid conceptual model is presented. Some reasons for telecom industry's interest in grid/cloud computing are discussed in section 2.3. Section 2.4 identifies standardization gaps between grid and NGN by comparing use cases deployed in both systems. Scenarios for integration Grid with NGN are outlined in section 3. Finally, we conclude with a summary and an outlook in section 4.

2 Grid / Cloud Computing and NGN

In this section, we will discuss how grid and cloud computing technologies, concepts, and capabilities can be integrated with NGN architecture to benefit from its capabilities.

2.1 Grid vs Cloud Computing

During 2008 and 2009 commercial interest shifted from "grids" to "clouds", with the availability of several on-demand compute and storage resources. What is the relationship between a "grid" and a "cloud"? The concepts are independent and complimentary. The similarities are that both aim to provide access to a large compute (CPU) or storage (disk) resource. Beyond that, a cloud utilizes virtualization to provide a uniform interface to a dynamically scalable underlying resource, with the intention that the virtualization layer conceals physical heterogeneity, geographical distribution, and faults. The cloud environment only provides direct support for single user or single organization access, and current models typically have a high cost to integrate computing, data, or network transfers from outside of the cloud. This model suits environments where compute and data resource needs can be isolated to a single location and rapid scaling (up or down) of compute, network, and data availability are important. Pricing models are variations on normalized CPU-hours, GB/day storage, and MB network I/O, or are based on a "cloud" product that can be licensed and used with local physical resources. Existing cloud systems use proprietary command line or browser (web) based interfaces. In contrast, grid computing aims to provide a standard set of services and software that enable the collaborative sharing of federated and geographically distributed compute and storage resources. It provides a security framework for identifying inter-organizational parties (both human and electronic), managing data access and movement, and utilization of remote compute resources.

Grid computing can benefit from the development of cloud computing by harnessing new commercially available compute and storage resources, and by deploying cloud technology on grid-enabled resources to improve the management and reliability of those resources via the virtualization layer. Cloud computing can benefit from grid concepts by integrating standard interfaces, federated access control, and distributed resource sharing. The current state of the art favors cloud computing for single organization commercial applications that can be deployed in their entirety onto a cloud environment. The dynamic provisioning of storage, compute power, and network bandwidth allows rapid scaling for intensive utilization either directly by the organization or by the public via Internet-based interfaces. Grid technology continues to dominate public sector scientific computing environments due to the collaborative nature of this work and the need to manage existing data sets and computing resources across organizational boundaries. The more advanced state of interface standardization within grid technology allows some degree of choice between various software and hardware systems. Deploying data and applications into a cloud environment, however, limits an organization to a single cloud provider or requires duplicated effort to repeat the deployment process for additional cloud environments.

2.2 Grid conceptual model

To discuss grid infrastructure in a telecoms context, a conceptual model has been developed [2]. This can be depicted as a layering of services which can be utilized independently or together. In Figure 1, these are grouped by the type of services they deliver. The lowest level represents the foundation of the infrastructure: networking, storage, computing power, and pre-existing software applications. These are wrapped and presented as software services. The next layer represents services that are central to the operation of the grid, while the outer layer provides user-focused services. These services are utilized by consumers, customers and providers. The consumer models the individual or organization using a grid services. The customer models the entity responsible for contracting the grid services, and pays for usage by consumers they have authorized. The provider models the entity providing grid services.

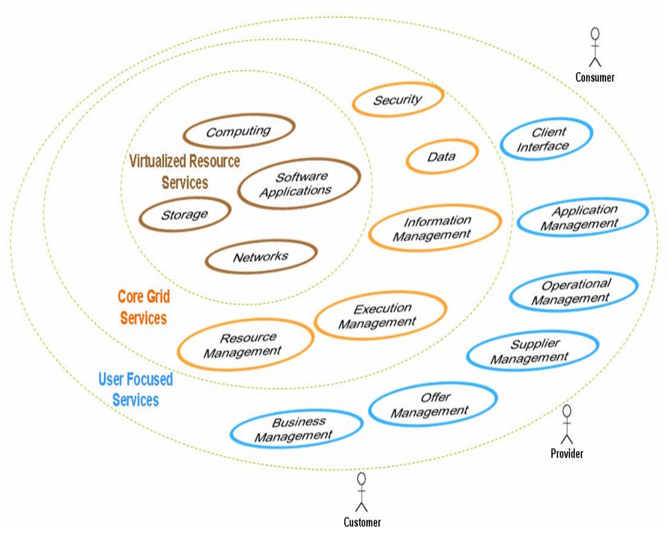

Figure 1: Conceptual model of a grid and associated roles.

2.3 Why Grid/Cloud computing is of interest to the telecom industry

Next-Generation Network (NGN) represents a fundamental change in telecommunication core and access networks, separating different service-related functions from transport-related technologies. NGNs are in many cases intended to provide consistent, ubiquitous and reliable information and communication services to users. In other cases, NGNs just specify that transport is separated from services using the same mechanism [3]. NGN (as defined by ETSI [4]) differs from the Internet. The Internet is a packet-based network, but it is an open network developed through interconnection of networks; it has no guarantee of QoS and it depends on applications for security and authentication. On the other hand, NGN which is also packet-based, enables to make use of multiple QoS mechanisms. NGN provides its own authentication and authorization mechanisms as well as security mechanisms at various protocols/logical layers [5].

NGN continues an evolution of “the network” to include newer technologies, to enable additional capabilities and to support an increasing array of applications. For example, the film production business is no longer only about making them available in cinemas, but also involves publishing content on official websites, blogs, chatrooms, social networking spaces and, in some cases, the launch of new video games. It is expected that in the near future there will be a huge increase of consumer-generated content, the spread of Machine to Machine (M2M) software as a service and grid and cloud computing.

Grid and cloud computing offer many application fields that can be used by telecom operators. Improvement of performance and reduction of costs due to the combining of resources is one of them. Another one is the possibility of creating virtual organizations to establish collaboration between teams with scarce and costly data and resources. More recently, grid and cloud computing has emerged in other areas with the purpose of optimizing company business. The aim is to combine computing resources for several services by reallocating them in a dynamic way depending on performance peaks. This strategy offers considerable cost cutting thanks to better management of resources, administrative tasks and maintenance.

Nowadays, telecom operators are more and more driving SOA (Services Oriented Architecture) implementations in order to:

- decouple applications via middleware from IT server/storage/network resources,
- flexibly compose new services using standards-based technologies and protocols,
- reuse architectural components to lower costs, time-to-revenue,
- rapidly manage configuration, change management and services assurance with automation.

It should be noted that grid and virtualization fit naturally into SOA-based infrastructures and can support services provided by NGN across different NGN layers. As an example, for IPTV service recently standardized by ETSI [6], [7], they can provide IT resource optimization and billing in the management layer, high performance content rendering and encoding in the service layer; co-allocation and cross optimization of network resources and grid resources (computer, storage) in the control plane).

There are more challenges for the use if grid and cloud computing within the NGN, for example, with respect to prediction of network behavior with the rise of IP-based services augmented by grid, recognition of new services and loads to arise with adoption of Web 2.0, now barely anticipated by telecom operators, or matching application/infrastructures to grids in a more generalized method.

2.4 NGN and Grid Gap Analysis

In [8] the STF 331 identifies barriers to interoperability of grid technology from the perspective of gaps in existing standards. Five areas are covered in detail: Architecture, Service Level Agreements (SLAs), Charging, Security, and Service Discovery.

In the presence of a multiplicity of network technologies and the resulting integration problems, the vertical integration of network layers is increasingly gaining in importance. These architectural elements are critical to the NGN and to the formation and operation of dynamic large-scale grid infrastructures.

The architecture must support end to end services, where the quality of network services requested by the grid layer need to be independent of the underlying networking technologies. It is the responsibility of the network service provider to map and enforce the required quality of service. Currently neither NGN nor grid domains provide suitable interfaces or models to manage this relationship.

Horizontal integration of grid and NGN architectures needs to support the co-existence of multiple network service providers for widely distributed grid applications. These providers need to allow collaborative mechanisms for end-to-end service establishment. Current cross-network standards focus primarily on network provider interfaces and relatively static topologies. To realize the full potential of an integrated NGN and grid environment it is necessary to expose the cross-network routing and QoS (Quality of Service) interfaces to third-party applications for real-time dynamic service provisioning.

A major shortcoming of the grid standards landscape is the lack of a widely agreed upon architectural reference model. While the OGF (Open Grid Forum) have produced OGSA (Open Grid Services Architecture) and adopted the Enterprise Grid Reference Model from the short lived Enterprise Grid Alliance, neither of these have found much practical use in the development of current grid infrastructures.

In the domain of SLA and QoS contracts and control, there is yet to be an established protocol or contract standard. The OGF have produced the WS-Agreement (Web Services Agreement) draft specification which is seeing gradual adoption. One challenge in establishing SLAs is the formation of an agreed set of SLA properties. While CIM (Common Information Model), GLUE, and JSDL (Job Submission Description Language) provide starting points for such a set of properties, there has been little success in finding widespread adoption of any single ontology, thereby impeding the application of an SLA. A similar lack of agreed SLA properties exists in the NGN domain. SLAs will need to be formed dynamically, implying automated systems for matching resource requests with resource providers and forming contracts with usage, QoS, and charging terms attached. Satisfying the QoS specified by an SLA implies an ability to monitor the various system layers in an integrated way and at an appropriate level of granularity. Current monitoring systems often do not provide suitable end-to-end service monitoring facilities or are unable to differentiate resource utilization by multiple users.

In terms of security, X.509 based Public Key Infrastructure (PKI) has been one of the great successes of grid technology, and the standards surrounding X.509 PKI have been widely adopted. These have also been integrated into some GSM-based SIM (Global System Mobile-based Subscriber Identity Module) card for mobile devices, and there are plans and standards describing the widespread use of PKI in the telecoms domain, with device and end-user PKI identity tokens. The issue of key distribution, binding devices to particular users, and trust of device identity tokens has many similarities with user, host, and service identities in Grids, and both lack standards to guide development.

Furthermore, there are significant standards gaps around the issue of authorization in the grid domain. SAML (Security Assertion Markup Language) and XACML (eXtensible Access Control Markup Language) are the only two broadly applicable authorization standards, however their complexity makes them difficult to use in practice. There is a need for a simplified authorization policy language. Furthermore, there is significant scope for standards concerning authorization policy management: sharing, merging, and updating security policies in an efficient, clear, and secure manner.

A standard model which defines a Virtual Organization (VO), membership, capabilities, and policies would provide an operational framework to improve VO-centric services, in contrast to the current focus on either user- or site-centric services. Finally, there are no standards for data provenance, an important issue in many domains: science, medicine, and financial services, to name a few. Only bespoke solutions for auditing data provenance are available.

Charging is an area where NGN and the commercial nature of network operations has significant experience, while the open collaborative origin of grid computing has resulted in little attention being given to this important topic. The complex and dynamic nature of grid usage patterns makes it a significant challenge to establish pricing models that can be implemented in practice to recover costs from the various providers involved in grid and NGN resource usage. The NGN model allows Billing On Behalf of Others (BOBO), such that a customer of an NGN service provider could elect to enable end users to purchase many products via an NGN account. Some effort has been made in the grid domain to produce Usage Records which could usefully be aligned with NGN Charging Detail Records (CDR) specifications. No standards exist for on line charging for Grid Services. There is scope to develop these jointly with the evolving NGN Standards. It is important that the use of these records for billing customers not limit the flexibility of Service Providers to develop custom pricing models.

Lastly, Service Discovery is a key issue for grid users as the discovery of resources to satisfy a particular need must be done quickly and efficiently, taking into account several characteristics and requirements of the particular usage pattern. Service registry mechanisms can be adapted for this purpose, however these typically do not support a high degree of dynamism (i.e. updates to the state of existing services in the registry, or the addition and removal of entries). Various grid service registries such as Monitoring and Discovery Service (MDS), Grid Resource Information Service (GRIS) and the associated Grid Information Index Service (GIIS), Relational Grid Monitoring Architecture (R-GMA), and Berkley Database Information Index (BDII) and others have been utilized by various grid infrastructures to varying degrees of success, but replication of such services indicates the lack of a single successful standard for resource registration and discovery. Resource and service coordination is also important given the mix of components which are provisioned and accessed during a typical grid usage scenario, however the complexity of standardizing this led the OGF working group Component Description, Deployment and Lifecycle Management (CDDLM) to abandon efforts to form a standard, despite some progress in the area. Currently the LDAP-based (Lightweight Directory Access Protocol) BDII system gained the highest level of adoption.

3 Scenarios for integration Grid with NGN

Several possible scenarios for integration of grid with NGN have been identified [9]:

- Grid-enabled NGN application
- NGN subsystems offering grid and cloud services
- Combining grid and networking resources in a new architecture
- Grid and cloud technology for implementing NGN functionality.

In the first scenario, as depicted in Figure 2, a grid-enabled NGN application is deployed as an Application Server (AS). The NGN is already designed to support a wide range of Application Servers and is capable of incorporating additional types developed by other groups.

Application Servers in NGN are available via a standard interface (ISC/Ma, where ISC is a reference point between Serving Call Session Control Function (S-CSCF) of the core IMS and AS, while Ma is a reference point between Interrogating Call Session Control Function (I-CSCF) of the core IMS (IP multimedia Subsystem) and AS (Application Server)). It needs to be further investigated if there is any impact on pre-defined NGN reference points and which type of NGN Application Server would be appropriate. The grid AS would access interfaces from the Core IMS subsystem. In this scenario, the NGN could provide IP connectivity, authorization, security, QoS, and charging.

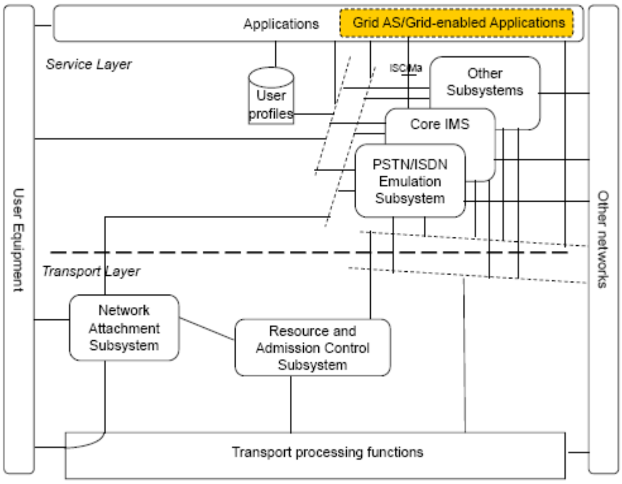

Figure 2: Grid-enabled NGN application.

In the second scenario, as depicted in Figure 3, a new NGN subsystem is added to the NGN Service layer to support the provisioning of grid or cloud services. This Grid Services Subsystem (GSS) gives access to a virtualized grid-enabled cloud resources, through a new service interface to the virtual resources and grid services. Such an integration of a GSS presents three different options:

a) Grid-enabled applications as defined in the first scenario. The GSS interfaces would provide common grid functionality and resource access, thereby accelerating development of higher-level applications.
b) Direct access by end-user applications, providing them with resources managed by the GSS in the same way that non-IMS Internet Protocol Television (IPTV) services are offered.
c) Updating other subsystems to leverage new capabilities offered by the GSS. This option requires the current NGN service layer subsystems to be enhanced to support grid services in the same way that IMS based IPTV services are offered.

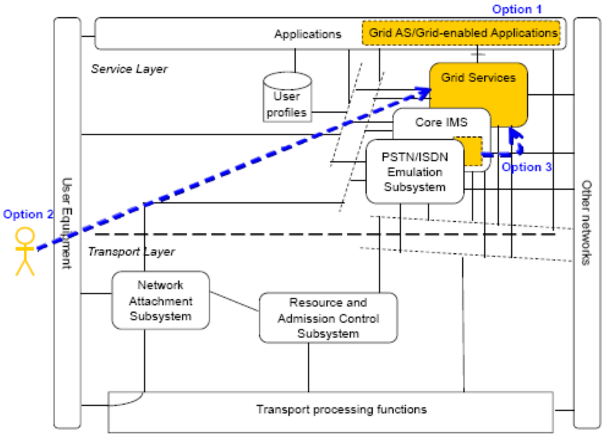

Figure 3: NGN subsystems offering Grid Services.

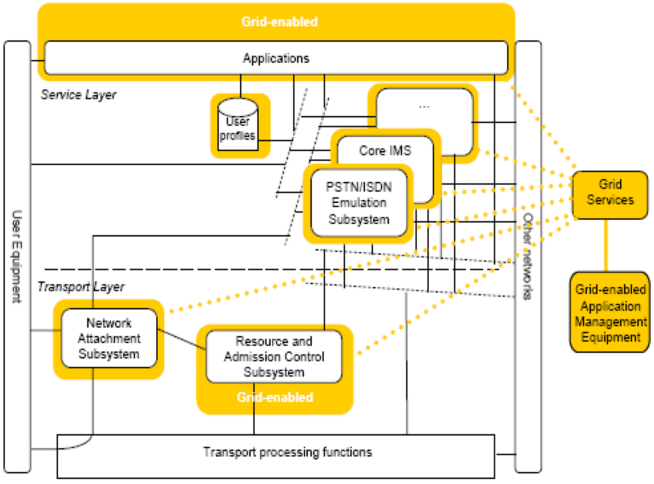

Figure 4: Grid technology for implementing NGN functionality.

A third scenario is to propose a separate "Resource Management" grid service that manages shared resources such as computing power, network and storage. This would enable the assignment of these resources to the grid or the NGN in a flexible, generic way.

The final scenario enhances the entire NGN architecture with capabilities for harnessing virtualized cloud resources and grid-enabled services, as depicted in Figure 4. This is the most disruptive and ambitious scenario, where logical NGN functions and entire NGN subsystems are refined in light of grid and cloud capabilities. This allows the optimization of the resources used by these functions and provides greater flexibility in the deployment and operation of various NGN subsystems.

3 Conclusion

Telecom operators are bound to become key players in a grid and cloud computing with the objective to improve their internal network operation as well as extend the services they offer to their customers. For this, integration of grid technology and cloud computing with telecom networks has to be achieved. ETSI and its TC GRID have a key role to play in establishing priority scenarios, standards, and testing mechanisms. This paper has considered several possible scenarios for integration of grid and cloud computing with NGN. Such integration requires interoperability between the both technologies. This can only be achieved when there are clear standards for interfaces and an environment that supports multiple implementations of architectural components. The lack of a widely agreed grid architecture, encompassing software, hardware, and services, impedes the development of a consistent set of standards. The telecom industry is expected to gain experience with third party services with the roll-out of NGN. We expect that this roll-out will lead to increased efforts for the interoperability of grid, cloud and telecom systems.

Acknowledgements

The authors would like to thank to the ETSI Technical Committee Grid for the guidance and support. We would also like to thank the European Commission for the funding provided under the contract SA/ETSI/ENTR/000/2006-10 which has made this paper possible.

About ETSI

ETSI produces globally-applicable standards for Information and Communications Technologies (ICT), including fixed, mobile, radio, converged, broadcast and internet technologies and is officially recognized by the European Commission as a European Standards Organization. ETSI is a not-for-profit organization whose 700 ETSI member organizations benefit from direct participation and are drawn from 60 countries worldwide. For more information, please visit: www.etsi.org.

About ETSI Specialist Task Forces (STF)

STFs are teams of highly-skilled experts working together over a pre-defined period to draft an ETSI standard under the technical guidance of an ETSI Technical Body and with the support of the ETSI Secretariat. The task of the STFs is to accelerate the standardization process in areas of strategic importance and in response to urgent market needs. For more information, please visit: http://portal.etsi.org/stfs/process/home.asp.

About STFs Funded By EC/EFTA

The work carried out here is co-financed by the EC/EFTA in response to the EC's ICT Standardisation Work Programme.

References:

[1] http://portal.etsi.org/Portal_Common/home.asp.
[2] ETSI TR 102 659-1 V1.1.1: GRID; Study of ICT Grid interoperability gaps; Part 1: Inventory of ICT Stakeholders. European Telecommunications Standards Institute (ETSI), Sophia-Antipolis, France (2008)
[3] WTPF (World Telecommunication Policy Forum): THE SECRETARY-GENERAL'S REPORT (28.2. 2009), http://www.itu.int/osg/csd/wtpf/wtpf2009/sg-report.html
[4] ETSI ES 282 001: Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Functional Architecture. European Telecommunications Standards Institute (ETSI), Sophia-Antipolis, France (2008)
[5] ETSI TS 187 003: Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Security; Security Architecture
[6] ETSI TS 182 028: Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IPTV Architecture; Dedicated subsystem for IPTV functions
[7] ETSI TS 182 027: Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IPTV Architecture; IPTV functions supported by the IMS subsystem
[8] ETSI TR 102 659-2 V1.1.1: GRID; Study of ICT Grid interoperability gaps;Part 2: Study of ICT GRID interoperability gaps
[9] ETSI TR 102 767: Grid Services and Telecom Networks; Architectural Option. Tech. rep., European Telecommunications Standards Institute (ETSI), Sophia-Antipolis, France (2009)