Grid-based Environment for Mobile Operators

Antonios Litke¹, Rui Aguiar², Sotirios Chatzis¹, Nuno Inacio², Dimitrios Halkos¹, Kleopatra Konstanteli¹ and Theodora Varvarigou¹

Abstract—This paper discusses an architecture which exploits and advances Grid technology as a Service Support technology for Mobile Telecommunication operators. This architecture relies on both the OGSA framework, and on current trends for next generation mobile networks. Key advantages of this architecture are: complex service support capabilities, efficient resource management techniques possible and seamless access to the service infrastructure. The architecture potential is further enhanced by the introduction of the Mobile Dynamic Virtual Organization concept.

Index Terms— Grids, 4G networks, network architecture, OGSA, SOA.

I. INTRODUCTION

The Grid can be viewed as a distributed, high performance computing and data handling infrastructure, that incorporates geographically-and organizationally-dispersed, heterogeneous resources (computing systems, storage systems, instruments and other real-time data sources, human collaborators, communication systems) and provides common interfaces for all these resources, using standard, open, general-purpose protocols and interfaces [1]. It is also the basis and the enabling technology for pervasive and utility computing due to its ability to be open, highly heterogeneous and scalable. The Open Grid Service Architecture (OGSA) Working Group [2] defines, within a Service Oriented Architecture (SOA), a set of core capabilities and behaviours that address key concerns in Grid systems. OGSA is an open standard managed by the Global Grid Forum (GGF) [3] standardization body. These concerns include issues such as: service discovery and management, identity establishment, membership management within virtual organizations, service collections organization and monitoring, policy expression and management, and others.

Grids are typically highly dynamic infrastructures, since new resources can dynamically enter or leave at any moment. It does seem ideal to environments where multiple service providers cooperate and to offer increasingly complex services, an expected characteristic of future 4G networks. However, the existing Grid architectures show a lack in their ability to incorporate mobile resources, since they cannot handle the mobility issues effectively and are unaware of the network related issues or the users’ context. Taking this into consideration we can define the Mobile Grid as a full inheritor of the Grid with the additional feature of supporting mobile users and resources in a seamless, transparent, secure and efficient way. Such a Mobile Grid will be deployable over multiple mobile networks and provide a self-configuring system of mobile resources (both hosts and users) that are physically connected by wireless infrastructures, in arbitrary and unpredictable topologies, but retain a service logic overlaid by the Grid infrastructure.

Previous work on the mobile grids arena [5] [6] mostly addressed mobility in the context of grid technology for small spaces and personal devices. The Akogrimo project [4] addresses global service environment for global coverage, and aims to develop an architecture bringing OGSA-based service provision into a mobile operator environment, exploiting the fructuous convergence of both worlds, in technical, service and business terms.

This paper describes early results from the Akogrimo project. Section II will address the overall OGSA-based architecture, while Section III will detail the different architecture layers. Section IV will address QoS and mobility considerations at the network level, and section V addresses the problem of bringing Applications into this environment. Section VI concludes the paper.

II. AKOGRIMO OGSA – BASED LAYERED ARCHITECTURE

The basic set of objectives an OGSA compliant architecture aims to fulfil are:

• Manage resources across distributed heterogeneous platforms. With multiple resources, geographically widespread, it is essential to offer a common management infrastructure to the service provider.
• Deliver seamless access control and quality of service (QoS). The topology of Grids is often complex, with dynamic resources interaction. It is important that the grid should provide robust, behind-the-scenes services such as authorization, access control, and delegation, in function of desired QoS to be provided.
• Provide a common base for autonomic management solutions. A grid can contain many resources, with numerous combinations of configurations, interactions, and changing state and failure modes. Some form of intelligent self regulation and autonomic management of these resources is necessary, acting constrained by existing service agreements.
• Define open, published interfaces, in order to promote an open service market. For interoperability of diverse resources, Grids must be built on standard interfaces and protocols. OGSA is an open standard managed by the GGF.

¹Dept. of Electrical and Computer Engineering, National Technical University of Athens, 9 Heroon Polytechniou Str. 15773 Athens, Greece. Corresponding author: ali@telecom.ntua.gr
²Instituto de Telecomunicações, Universidade de Aveiro, 3810-193 Aveiro, Portugal.
standards body, and as such is one target standard body to follow.

The Akogrimo architecture is separated in four different layers: a logical and physical layer, a web services layer, an OGSA services layer, and the applications domain. Figure 1 depicts the overall architecture.

**Figure 1**: The focus of the Akogrimo layered architecture with respect to OGSA

1. **Physical and logical resources layer**

   The concept of resources is central to Grid. Resources comprise each and every capability of the Grid and can be classified in two categories: Physical resources, which include servers, storage, and network and Logical resources. The logical resources are above the physical and provide additional functionality by virtualizing and aggregating the resources in the physical layer. General purpose middleware such as file systems, database managers, directories, and workflow managers provide these abstract services, traditional to grids. Note that Akogrimo goes further in this issue, since at this very layer communication resources are also incorporated (see e.g. [7]), as well as their virtualization to the grid environment.

2. **Web services layer**

   All Grid resources, both logical and physical, are modelled as services on the basis of Web service implementations. However, for the purposes of the resources modelling in the Grid the Web service interfaces must frequently allow for the manipulation of state, that is, data values that persist across and evolve as a result of Web service interactions. To achieve this, the Web Services Resource Framework (WSRF) [7] defines a family of specifications for accessing stateful resources using Web services. Note that communication services are also described using this framework.

3. **OGSA architected grid services layer**

   The Web services layer, together with the WSRF, provides a base infrastructure for the next layer OGSA architected services providing the overall grid management functionality.

4. **Grid applications layer**

   These applications are the high – level applications a client might require. They might come from various scientific and business domains and include among others biomedics, construction, engineering, finance, etc. For the purposes of the Akogrimo system we have deployed an E – Health application offering a suite of E – Health services.

   In the proposed architecture of the Akogrimo system the Grid functionality of the whole infrastructure is provided by the “Grid infrastructure” level. This level consists of the Web Services Resource Framework (WSRF) level and the OGSA services layers, being placed in the middle of the Akogrimo architecture participating as the “Grid glue” to the functionality of the Akogrimo system.

**III. AKOGRIMO GRID INFRASTRUCTURE SERVICES LAYER**

The following figure gives a conceptual overview of the overall Akogrimo layered architecture depicting the main architectural components it consists of and their positioning within it.

**Figure 2**: Positioning of Grid Infrastructure Layer components in the overall Akogrimo architecture

The basic role of the Grid Infrastructure Services layer is to manage the execution of the services on request of the Application Service Layer, aiming to fulfil the Grid requirements, addressing the performance issues (job throughput, efficiency, maximum utilization of the resources) in a transparent way to the user and conforming to the determined Service Level Agreement (SLA). More precisely, the Grid Infrastructure Services layer interacts with:

- Application Services layer, to receive the jobs to be executed and to identify the corresponding SLAs and Policies that will regulate and influence this execution.
- Mobile Network Middleware layer, mainly in order to discover services and have access to AAA.
- Mobile Network layer in order to perform various actions that deal with the network (for instance QoS bundles of
services for having advanced reservation of network resources).

The basic functionality that must be supported from the Grid Infrastructure Services layer, is categorized in the following:

**Execution Management Services (EMS):** This category of services comprises all the functionality that is concerned with the problems of instantiating and managing tasks, such as assigning jobs to resources, creating an execution plan, balancing the workload, optimizing the performance, and replicating jobs to provide fault tolerance.

**Data Management:** This category comprises all the functionality that is concerned with the access to and movement of large data sets, as well as data sharing, replicating and archiving of data.

**Monitoring:** This category comprises the services that are focusing on monitoring and managing of the web services within the layer.

**Service Level Agreement (SLA):** Services related to the enforcement of the SLA contractual terms that especially influence the execution of jobs within the layer.

**Metering:** Services supplementary to the monitoring and accounting services, dealing especially with the measurement of resource usage.

**Policy management:** This category of services comprises the functionality concerned with the management of rules and the policies which apply in the execution of services within the Akogrimo architecture.

**Security:** This category comprises the services that are concerned with the security issues of the specific layer. It comprises the services that will deal with the confidentiality of the communications and the authorization for execution within the system.

IV. **QoS ENABLED MOBILITY SUPPORT**

Mobility is an essential aspect for our architecture. In order to provide users with a seamless environment, different types of mobility have to be addressed.

**Terminal mobility:** allows a mobile terminal to maintain its connection to the network when it changes access points. This is assured by the Mobile IPv6 protocol [9], enhanced with fast mobility mechanisms [10].

**User mobility:** allows the user access to personalized services independently of the user’s device. It is provided by a user-oriented security and authentication framework. The user has to perform its registration in the network – and in the Grid infrastructure – before using the network services. This registration process associates the user with the terminal [11].

**Session mobility:** enables the transfer of application sessions between different devices without interruption. This is achieved with the SIP protocol [12]. SIP can be used both by the user, and by the Grid infrastructure, to redirect communications (e.g. image display) to different devices, retaining the user association mentioned above.

To make an effective use of network resources and to assure that a user can utilize the services he is entitled to without disruptions, the Mobile Network Layer will implement end-to-end Quality of Service, for all types of mobility. Our QoS implementation uses a hybrid IntServ

[13]/DiffServ[14] QoS model in order to avoid IntServ’s scalability problems and DiffServ’s inflexible control system. The QoS system allow fine-grained QoS reservations at the access networks (following the Intserv model), but aggregates different flows with the same QoS requirements in the core network (the Diffserv model). For implementation and scalability reasons, the network supports well defined QoS bundles, strongly influenced by the existing models for mobile technologies (UMTS). Each of the three defined bundles is designed for a specific usage profile, audio, video and data. A QoS Bundle is comprised of several well defined services, which the user may choose from when using the Akogrimo network. Table 1 presents these bundles.

<table>
<thead>
<tr>
<th>Bundle 1</th>
<th>Bundle 2</th>
<th>Bundle 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mix audio + data</td>
<td>High data + video</td>
<td>Mostly voice</td>
</tr>
<tr>
<td>10 – Interactive</td>
<td>20 – Interactive</td>
<td>10 – Interactive</td>
</tr>
<tr>
<td>100 – Data</td>
<td>1000 – Data</td>
<td>1 – Priority</td>
</tr>
<tr>
<td>1 – Priority</td>
<td>200 – Priority</td>
<td>1 – Signalling</td>
</tr>
<tr>
<td>1 – Signalling</td>
<td>1 – Signalling</td>
<td>250 – Best Effort</td>
</tr>
<tr>
<td>250 – Best Effort</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1:** QoS Bundles, values and usage. Values are in kilobytes per second.

**Signalling:** This is traffic needed so as to maintain and support the network infrastructure; therefore it is the highest priority. It is time-critical and, in fact, essential to network operations as a whole. Typically its bandwidth requirements are very low. Critical grid infrastructure commands are also marked as such.

**Interactive real-time:** This is time-critical traffic that will be used mainly for video conferencing or audio communications. Interactive multimedia applications are very sensitive in regards to delays. The delays allowing for optimal functioning of interactive applications are less than 100ms. Latency, jitter and out-of-order packets also affect adversely voice communications.

**Priority:** This type of traffic is not time-critical, but it is important, such as multimedia streaming, or some grid application data exchange. It is higher priority than Data Transfer, but has lower bandwidth available typically.

**Data Transfer:** Data transfer is somewhere in between Priority and Best Effort. This type of traffic is not time-critical but may be loss-sensitive. Furthermore, out of order packets are typically not a concern for applications that use Data Transfer.

**Best Effort:** As the name implies this service offers best effort. If the network conditions are good, this should be fine for most applications. If the network is heavily loaded, BE will be the most affected. This is basically what Internet provides.

The network layer entity that handles QoS requests is the QoS Broker. It has information about the user and about the current status of the network. It knows whether a user is allowed to use a certain QoS level, and if the network will support it given its present traffic. For allowing the EMS to perform QoS requests, the QoS Broker has an interface designed specifically for interaction with the EMS. This interface is a cross-layer interface and is based on Web
Services standards. As a consequence, network services become part of the workflow, and are defined both according to the user profile (and subscribed services) and according to its current operations.

V. BUILDING GRID ENABLED APPLICATIONS

A. Application Services Layer

In this environment, each application service requested by a client is modeled as a business process instance. The business process represents a set of one or more linked procedures or activities that collectively realize a business objective goal, normally within the context of an organizational structure defining functional roles and relationships. Behind the business process there are one or more workflows representing the automation of the business process. Each workflow coordinates and manages component services or entities involved into the automation of business process. This coordination procedure of the operations carried out by the Workflows associated with a Business Process is called as Business Process Enactment.

In Grid, a Virtual Organization (VO) provides services and the means to manage and coordinate them. In the Akogrimo framework we extend the definition of the VO in the following manner: We consider the VO as a Mobile Dynamic Virtual Organization (MDVO), which is defined as a temporary or permanent coalition of geographically dispersed potentially mobile individuals, groups, organizational units or entire organizations that pool resources, capabilities and information, selected from the resources of an Enterprise Network, to contribute to the VO according to the dynamically established contracts typically driven by one or more business processes.

In order for a business process to be implemented, an Operational Virtual Organisation (OpVO) is created out of a base VO. The base VO is a Virtual Organisation that is not running a specific business process, but provides the means for creating and supporting it. The base VO provides the means to register users, services and other meta-data like Service Level Agreement (SLA) contracts and workflow templates. These repositories are used by the operational VO when a business process is instantiated and executed.

B. Web Services Resource Framework

The Web Services Architecture did not meet one of OGSA’s most important requirements, the statefulness. Although Web services in theory can be either stateless or stateful, they are usually stateless and there is no standard way of making them stateful. The Web Services Resource Framework (WSRF) [15] solves this problem, keeping Web service and state information completely separated. Instead of putting the state in the Web service, this is kept in a separate entity called a resource, which stores all the state information. Each resource has a unique key, so whenever we want a stateful interaction with a Web service we simply have to instruct the Web service to use a particular key (resource).

C. The e-Health scenario

For demonstration purposes the Akogrimo consortium has deployed a suite of e-Health application services providing to its users/clients a Heart Monitoring and Emergency Response Process. The organizational framework where the process is executed consists of a university hospital, regional hospitals, medical specialists, general practitioners, emergency medical services and an emergency dispatch center establishing a regional health network. The health network is headed by the university hospital and provides telemedicine services to the partners and the patients attended by partners of the network. These services are focused on particular diseases and risk groups.

D. From user to the service through Grid middleware

The Grid middleware consists of high-level services adequate for gridifying the various applications, and which are implemented through the WSRF specification, the most widespread implementations of which are the Globus Toolkit [19] and WSRF.NET [20].

Whenever the client application needs to use an e-Health service, it contacts the OpVO Manager Service by sending the proper SOAP request over HTTP. The OpVO Manager Service, after making the necessary negotiation with the client, invokes the EMS which is in charge of the initiation, management, scheduling and coordination of all the services. A notification mechanism, implemented upon the WS-Notification family of the WSRF specification is established among all the services that are part of the Grid middleware as well as between the EMS and the E-Health service. In Fig. 3 we present an overview of the aforementioned procedure, which encapsulates the physical and logical entities that participate in the e-Health applications scenario. The interaction with network services is done by specific contracts established between users and service providers, and the EMS. The network is informed of which bundle to provide to each service, and how to allocate network resources to the different information flows crossing the infrastructure.

Figure 3: Overview of the service access through the Grid middleware
E. Building Mobile Dynamic Virtual Organizations

The term Virtual Organization (VO) from its economic perspective has been described in [16][17]. It is an organizational model describing the rules of interaction between companies not limited to IT resources. Combining this with the definition of the Grid community [18], and the requirement for mobility support we define a Mobile Dynamic Virtual Organization (MDVO) as a temporary or permanent coalition of geographically dispersed potentially mobile individuals, groups, organizational units or entire organizations that pool resources, capabilities and information, selected from the resources of an Enterprise Network, to contribute to the VO according to the dynamically established contracts typically driven by one or more business processes.

In Fig. 4 we present an example MDVO for the Akogrimo project. The example is focused on the Authentication, Authorization and Accounting (AAA), whereby each participating organization maintains its local AAA management services and participates in the MDVO with specific physical resources (which can be mobile). In order to share physical resources and collaborate, the organizations have to provide the security credentials in a manner that will be valid not only within the logical boundaries of the organizations but also within the limits of the formed MDVO.

![Figure 4: Mobile Dynamic Virtual Organization example](image)

In this manner, the MDVO AAA Management services, located physically in the premises of a organization, form a middleware approach which interacts with the local AAA management services for reasons of authenticating the usage, the actions that the users are able to perform and to account the usage of the latter. Thus the grid infrastructure provides a service layer that takes in consideration the network aspects (mobility, QoS), and provides the service management layer for the traditional telecom services. This may be especially interesting in a complex multi-operator scenario, such as those depicted for 4G networks.

VI. CONCLUSIONS

The paper presents our architecture that incorporates grid concepts for mobile operator environments. This architecture follows the OGSA framework, and considers a layered approach. Mobile Communications aspects, such as the communication infrastructure, are integrated with high-layer grid infrastructure services. This infrastructure provides an interface for easy deployment of complex application services, including management of heterogeneous resources. The integration of mobile communications and grid-services framework may ultimately lead to the integration of grid technologies in the control of the mobile network services.

ACKNOWLEDGMENT

The authors would like to thank all partners of the Akogrimo consortium for their valuable contribution in the current study.

REFERENCES