A prototype and demonstrator of Akogrimo's architecture: An approach of merging grids, SOA, and the mobile Internet

Jähnert, J; Mandic, P; Cuevas, A; Wesner, S; I. Moreno, J; Villagra, V; Olmedo, V; Stiller, B


Postprint available at:
http://www.zora.uzh.ch

Posted at the Zurich Open Repository and Archive, University of Zurich.
http://www.zora.uzh.ch

Originally published at:
A Prototype and Demonstrator of Akogrimo’s Architecture: An Approach of Merging Grids, SOA, and the Mobile Internet

Jürgen Jähnert¹, Patrick Mandic¹, Antonio Cuevas¹, Stefan Wesner², Jose I. Moreno³, Victor Villagra⁴, Vicente Olmedo⁴, Burkhard Stiller⁵

¹ Rechenzentrum der Universität Stuttgart, Germany, [jaehnert|mandic|cuevas]@rus.uni-stuttgart.de
² HLRS, Stuttgart, Germany, wesner@hlrs.de
³ Universidad Carlos III de Madrid, Spain, jmoreno@it.uc3m.es
⁴ Universidad Politecnica de Madrid, Spain, [villagra|volmedo]@dit.upm.es
⁵ University of Zürich, Switzerland, stiller@ifi.uzh.ch

ABSTRACT

The trend of merging telecommunication infrastructures with traditional Information Technology (IT) infrastructures is ongoing and important for commercial service providers. The driver behind this development is, on one hand, the strong need for enhanced services and from the other hand, the need of telecommunication operators aiming at value-added service provisioning to a wide variety of customers. Furthermore, this is driven due to the fact that existing value chains in the telecommunications infrastructure are disappearing, while the IT sector could bring added value to the telecommunications network. The last step is highly demanded in order to increase the willingness to pay for new services. Such an increased willingness to pay can justify huge investments to be made by operators within their infrastructures. In the telecommunications sector, the IP Multimedia Subsystem (IMS) is a promising service platform, which may become a “standard” for supporting added-value services on top of the next generation network infrastructure. Since its range of applicability is bound to SIP-enabled services, IMS extensions are being proposed by “SIPifying” applications.

In parallel to these developments within the traditional IT sector, the notion of Virtual Organizations (VO) enabling collaborative businesses across organizational boundaries is addressed in the framework of Web Services (WS) standards implementing a Service-oriented Architecture (SOA). Here, concepts for controlled resource and service sharing based on WS and Semantic Technologies are deployed. Since the telecommunications sector has become, in the meantime “mobile”, all concepts brought into this infrastructure must cope with the dynamics that mobility brings in. Therefore, within the Akogrimo project the VO concept has been extended toward Mobile Dynamic Virtual Organizations (MDVO), considering additionally key requirements of mobile users and resources. Different scenarios for MDVOs show a clear demand on integrating multimedia applications, in particular conversational service. This work here describes major results of the Akogrimo project, paying special attention to the overall Akogrimo architecture, the prototype implemented, and the key scenario in which the instantiated Akogrimo architecture shows a very clear picture of applicability, use, and an additional functional evaluation.

1 INTRODUCTION

The term Service Platform brings to the reader’s mind concepts or platforms such as i-mode, IMS (IP Multimedia Sub-system) [2], or OSA/Parlay (Open System Architecture) [23], which are controlled by network operators. They can be regarded as a way to regain the central position in the telecommunications business value chain. These platforms potentially enable operators to play a new role termed service brokers, which manage the relationship between consumers and service providers.

Operator-controlled service platforms are business-oriented and adapted to mobile environments. For instance, IMS is designed by the 3GPP (3rd Generation Partnership Project), in co-operation with the IETF (Internet Engineering Task Force), for 3rd Generation (3G) UMTS (Universal Mobile Telecommunications System) mobile networks. IMS is a service platform, that uses open Internet protocols respecting the Internet’s paradigm of data transport and application separation, but still building links between these two layers [15], [9]. Due to its design and being based on Internet protocols, IMS may become a service platform for Next Generation Networks (NGN), also called 4th generation.

Nevertheless, service platforms must entice users to employ them, offering a rich service environment. IMS is a promising service platform, but it targets only SIP-based (Session Initiation Protocol) [3] services, which use the SIP protocol and SIP proxies to help to control sessions between participants. These peer-to-peer applications allow users, among others, to make “traditional phone calls”, conceptually replacing services offered in the POTS world, which are, still today (although with a negative trend), the network operators’ main business. However, many of these applications with a broad user acceptance do not use the SIP protocol, and thus they are not fully integrated in the IMS framework. Efforts are being done to broaden the panoply of services that can fit under the IMS control framework. OMA (Open Mobile Alliance) [21] is the leading actor in enriching services offered over the IMS platform.
Beyond IMS and other operator-controlled service platforms, other infrastructures, adhering also to the concept of service environments, have been developed by the ICT (Information and Communication Technologies) sector. In the Grid computing field, based on Web Services (WS), solutions and concepts have been developed for the specific use of resources from different administration domains. Furthermore, the Service-oriented Architecture (SOA) paradigm has provided means to ease the reutilization and composition of services to a much broader range than telecommunication areas have envisaged. Of special interest is the concept of a Virtual Organization (VO) in combination with several execution environments providing a promising approach for the delivery of complex services. The key communication protocol in this area is SOAP (Simple Object Access Protocol), used both for signaling and for service execution [19].

Web Services-based Grid concepts, however, lack efficient and flexible mechanisms supporting large scale commercial operations, which might establish contracts even dynamically as needed in a mobility enabled telecommunication infrastructure, since they were designed for the “fixed Internet”, where hosts are desktops computers or large servers being totally network-agnostic, and therefore unaware of the end (commercial) user, who is mobile. Additionally, the origin of Grid computing from the High Performance Computing domain has driven solutions towards scenarios where the number of resource providers is small, their services are easily comparable and closely related to their physical resources and the number of users is just a few hundred.

IMS and Service-oriented Grids both claim to deliver service environments: IMS, for commercial scenarios, controlled by network operators, and, in the “mobile Internet” environment, its range of services is bound to SIP-based applications. Service Grids may offer services from the network up to the application in a more open environment, but have not proven their commercial liability in large settings or for mobile environments.

Thus, to overcome these drawbacks, the driving idea behind the Akogrimo project [12], [13], [27] is to merge large scale properties of telecommunication platforms and their support for multimedia applications with the cross-organization collaboration concepts of service-oriented Grids into a coherent and integrated service platform. The overall Akogrimo architecture [14] extends the definition of the traditional Grid computing toward a mobile collaborative business Grid. The proposed architecture covers the dynamic collaboration across organizational boundaries, cross-layer cooperation from the network up to applications, and considers the impact of a potentially very large number of mobile users using a wide range of devices and access networks. The type of mobility considered in Akogrimo embraces two types of mobility: nomadity, where users and resources can sign on and off from different locations or terminals, and true mobility, where the user and resources can change the access network without interrupting a session. Consequently, Akogrimo is targeting a public telecommunications network architecture beyond 3G towards a semi-walled All-IP service provisioning.

This work describes the final step of those phases comprised in such a project: the integration and demonstration phase. Previously, results relating to other phases of the project, such as the concept development or the architectural phase had been published. This article describes the Akogrimo project with a special focus on the final and successfully implemented, integrated and demonstrated Akogrimo architecture and is divided into the following sections. Section 2 describes the key Akogrimo architecture from a global perspective showing the key building blocks and concepts. Section 3 describes major concepts of the Akogrimo architecture, which are an integrated signaling framework overcoming the traditional gaps between the telecommunication sector and the IT/Grid world. Furthermore, it contains the integration of network-related context and its provisioning to the intelligent Grid “layer” enabling the integration of workflow-driven intelligence and the control of an execution environment. Finally, an integrated accounting concept comprising all involved “layers”, which serves as a must for commercial telecommunication providers, is described. While Section 4 details the instantiation of the Akogrimo architecture being implemented, which has been successfully demonstrated in public events, Section 5 provides a summary of the evaluation of this integrated architecture. Finally, Section 6 draws conclusions.

2  BASIC ARCHITECTURE

The fundamental assumption driving the Akogrimo architecture was to deliver a business proposition for telecommunication operators. However, within Akogrimo, a fair collaboration of the following three different business entities is considered: (a) network operators essentially offer an IP network and connectivity, (b) service providers provide their services on top of this network to their customers, and (c) the customer itself. The provision of non-network-related services is not limited to specialized service providers, but customers and network operators can similarly provide such services. It is assumed that the cooperation of these three different business entities is controlled by pre-negotiated contracts that build the basis for dynamically established session-based agreements. The “semi-walled garden” business model, which privileges contracts that put network operators in the “center” of the customer-to-service provider relationship, is the most immediate model that can be instantiated in the Akogrimo architecture. While in this model the network operator establishes the initial contact with the customer, this does not automatically mean that the major revenue is generated by this service broker role.

While Akogrimo remains partially network operator-centric, it borrows from the Grid domain the Virtual Organization
(VO) concept, which has been merged with the telecommunication operator’s business world. Early service-oriented Grid research projects, such as TrustCoM [24] have defined a VO as a temporary or permanent coalition of geographically dispersed individuals, groups, organizational units, or entire organizations who pool resources, capabilities, and information to contribute to the VO according to the established contracts. These contracts are typically driven by one or more business processes. VOs can provide services and thus participate as single entities in the formation of further Virtual Organizations (VOs). This enables the creation of recursive structures with multiple layers of “virtual” value-added service providers.

Akogrimo extends the GRID VO concept in several aspects. Firstly, Akogrimo considers static contracts, linking network operators, service providers, and customers. These contracts form what is called a “Base Virtual Organization” (BVO), identifying all actors (including millions of customers) that may interact and define aspects such as the payment flow. BVOs are static and do not allow any recursion. Secondly, Akogrimo delivers (and consumes) a service, chooses dynamically several actors from the BVO, and selects the most appropriate ones to deliver the service. Composition techniques from the GRID framework are employed in this task. These techniques outperform composition techniques from current mobile service platforms. Essentially, WS are employed, which form the most utilized GRID framework. Once the workflow is composed, it can be instantiated and different parties involved dynamically set-up specific agreements. As a result of these specific negotiations, an “Operational Virtual Organization” (OpVO) is created. OpVOS are dynamically created, destroyed, and excursiveness is allowed. Any actor can participate simultaneously or sequentially in different OpVOS [14], [25], [26].

The degree of dynamicity inside OpVOS considered in Akogrimo is very high and is one of the substantial differentiators to other existing approaches. The dynamicity supported covers local resource discovery, “prosumer” (producer and consumer) concepts, where Akogrimo users can offer themselves “services” using their mobile devices. Additionally, users change their devices, change their access network, or change other context parameters up to temporarily loosing access network coverage. Such a degree of dynamicity is not supported by other Grid frameworks; it is only supported partially today by platforms used by mobile telecommunication operators.

For the design of interactions and interfaces between different building blocks of the Akogrimo architecture, a layered approach was deemed too rigid and the current approach uses “areas” instead of layers as shown in Figure 1. An important feature includes direct interaction possibilities between those areas, for example network requirements can be communicated directly from the Application Services area to the Network area.

![Akogrimo high level conceptual architecture](image)

**Figure 1** Akogrimo high level conceptual architecture

From the bottom to the top of these areas, the following key definitions hold. In Akogrimo all services (including traditional network services offered by the network layer) are virtualized as (Web) Services. It is important to notice the information flow between different areas regarding infrastructure processes, such as context or identity, which will be explained below:

- **Network**: The network area interfaces with all other areas, but above all with the Network middleware. The network area defines and implements the mobile network infrastructure, the network policies and enforcement, and the network Quality-of-Service (QoS) support.
- **Network Middleware**: The network middleware provides services, such as Authentication, Authorization, Accounting, and Charging (AAAC), session control, identity management, and user context.
- **Grid Middleware**: The Grid middleware provides a framework for the utilization, reservation, and discovery of resources and data.
- **Grid Application Support Services**: In Akogrimo, all services (including traditional network services offered by the network layer) are virtualized as (Web) Services. It is important to notice the information flow between different areas regarding infrastructure processes, such as context or identity, which will be explained below. This area interfaces very closely with the Grid Middleware and handles the interaction higher level interaction
between the different components, the management, discovery, and instantiation of workflows for each application and the creation of the different VOs.

3 AKOGRIMO KEY INNOVATIONS AND OPERATION

The key innovation of Akogrimo includes the VO operation, the dynamicity and mobility support for Web Services, the cross-layer execution framework, and a seamless integration of charging and accounting support for mobile services.

3.1 VO OPERATION

As mentioned, Akogrimo adopted the notion of the Virtual Organization (VO) and extended it to the Base VO (BVO) and the Operational VO (OpVO). BVO and OpVO are the two main Akogrimo building blocks and their operation, depicted in Figure 2 defines the key basis in the Akogrimo Framework.

![Figure 2 Akogrimo fundamental operations](image)

The customer is allowed to perform two main actions inside a BVO using its base services: identifying best candidates for service provisioning and creating an OpVO.

The OpVO is the run-time environment or the private and dynamically created service provisioning platform for Akogrimo applications. The OpVO maintains relations to other building blocks during its operation as follows:

- **The Network Operator (NO)** provides the infrastructure to access OpVO services and the OpVO leverages the trusted NO for the Identity Management and authentication for all incoming requests.
- **The Customer** can use administrative services to manage its OpVO (e.g., subscription of new users or termination of the OpVO). Beside these administrative services, application services can be accessed by the customer and other users.
- **The Service Provider (SP)** will be contacted by the OpVO in order to negotiate the instantiation of a service in the SP domain to be used during the workflow execution. Upon successful negotiation, a Service Level Agreement (SLA) as an electronic contract will be concluded.

The Customer domain is the representation, in the home domain, of a customer buying services from the Service Bundling instance and offering them to a user or consuming them himself (in that case the customer is also the user).

In the SP domain, services manage and supply resources as well as enforce policies and SLAs. These tasks are supported by services located in the service provider domain and include:

- **The Execution management** is decomposed into multiple tasks, such as preparation, initiation, and managing of the execution.
- **Discovery-advertising services** enable the SP to create advertisements of services and/or resources that are for sale and register them to the index service of the platform, serving as a SP-domain wide index service.

On the NO domain, all related functions are represented, which are required to provide network services being compliant with Web Services, which in turn make this domain Akogrimo-specific and unique. Basic network functions — as currently provided in other architectures as well — are not described necessarily in a WS-compliant way. Additionally, available accounting and charging functionality in this domain are provided to account and charge for any service available. This enables the network domain to be commercially deployable as a Grid service on their infrastructure in a way, which is fully compliant to current IETF standards. This also comprises an identity concept,
which offers Single-Sign-On (SSO)-based services, WS, and QoS-enabled network services.

3.2 DYNAMICITY AND MOBILITY FOR WS

The dynamicity capabilities of OpVOs as highlighted above are provided by the network middleware “area” of the Akogrimo architecture (cf. Figure 1) and utilizing a SIP-based infrastructure. This area interacts with Grid areas, whose infrastructure uses the SOAP protocol [19]. This means that two worlds are merged: the traditional networking world and the world of SOA as well as grid computing. In such situations, one can always encounter redundant parts and other missing parts need to round up the architecture as a whole. In this case, two different signaling protocols come into play: on one hand SIP from the network side and, on the other hand, SOAP. Thus, Akogrimo combines these two protocols in order to provide a maximum flexibility, while allowing legacy-based development in an integrated manner.

SIP targets the establishment, modification, and termination of sessions (cf. Figure 3). In the context of SIP, a session is defined as an association created between two or more participants to interchange data. SIP’s specification defines signaling messages and the operation of the protocol, decoupling it from the actual type of data to be interchanged once a session has been successfully established. Despite this decoupling, SIP has been used traditionally to manage multimedia sessions, such as Voice-Over-IP (VoIP) or telephone-like calls.

SIP has mobility and dynamicity capabilities and this is one of the main reasons for 3GPP to use it as its main protocol in IMS. SIP signaling functionality is applied to locate ubiquitous and nomadic users and to manage sessions between them.

![Diagram of SIP signaling protocol for applications](image)

Figure 3 SIP as signaling protocol for applications

SIP can be leveraged in order to manage signaling for other applications beyond traditional audio-video conferencing. This is the case of the Mobile Grid scenario: WS-based applications provision and consumption among mobile, nomadic, and ubiquitous users in the scope of an OpVO. In such a scenario, there is a need to manage sessions for traditional SOAP-based applications. Thus, Akogrimo proposes SIP as the signaling protocol for SOAP-based applications.

In general, the SIP-SOAP interaction, to enable an integrated approach, can be implemented in two ways:

- To provide a SIP API (Application Programming Interface) to Grid applications. Thus, Grid applications will issue SIP messages; i.e. a SIP-UA (User Agent) is integrated into the Grid application, just like for any other application applying the SIP protocol. While, in terms of the architecture, this would be straightforward, it implies a costly (sometimes impossible) modification of Grid and WS applications.

- To provide intermediate modules, which will be invoked by applying the SOAP protocol from Grid applications. These modules will issue SIP methods (e.g., REGISTER, INVITE, or BYE respectively) on behalf of the Grid application, establishing “Grid sessions” on behalf of them as needed. This approach does ease the implementation of Grid applications from WS providers, since they can use SOAP messages as usual. They are directed only to intermediary modules as for SIP-related methods.

There exist other proposals, which address the SIP and SOAP interactions: [1] and [4] express SOAP requests inside the SIP protocol in order to invoke WS from the IMS environment (where SOAP might not be supported by terminals). However, this proposal does not address the problem of mobile WS, only the access from IMS terminals. The solution that was adopted in Akogrimo was different, similar to the solution proposed in [5]: using SIP for controlling SOAP sessions. In this way, the approach covers the inclusion of SIP support in a SOAP-enabled terminal, by including a SIP-UAI the WS applications of a terminal (first option of the above-listed) and implementing a SOAP-SIP Gateway ready to be used by WS providers (second option) [20]. The resulting sequence of message exchanges for this interaction is depicted in Figure 4. In IMS terms, the SIP Server is equivalent to an S-CSCF and the gateway is an SIP-SOAP AS.
The SIP session control service is presented as a WS and is provided by the NO via SIP proxies. Thus, it can be integrated and composed with other WSs in the OpVO. E.g., the workflow manager orchestrating all WSs in the OpVO can decide to setup SIP-controlled calls between users’ devices and servers hosting WS applications. The SIP proxy infrastructure can be used by the workflow manager or WS for many other aspects, while exploiting its dynamicity, mobility support, or user status tracking. These technical advances determine the first key Akogrimo innovations.

Within the first step the system has to register with the SIP proxy infrastructure all WSs being provided by users and their terminals that are logged in (e.g., medical first-aid for a doctor). Using SIP, the status of these services is published, which, from the user’s terminal side, concludes this initial process (cf. Figure 5). Service providers must also register and publish those WSs they offer. It is important to note that, during registration, the service in the SIP proxy infrastructure uses the SIP-URI (Uniform Resource Identifier) as the basis for identifying this service.

By utilizing EPRs (End Point Reference) Grid services and SOAP reference these services, while, within SIP, this is done using URIs. All services (Grid-based and SIP-based) offered within the IMS service platform are indexed in the SIP-Registrar/Server. This indexing issue is addressed in Akogrimo during the session setup. The main result of the
session setup negotiation is that session participants will come to know those EPRs of services they are offering and using in this session. The session setup is performed completely by using SIP. SIP allows for carrying any SDP (Session Description Protocol) data, and thus, tailored SDPs data can be defined for particular applications. Therefore, Akogrimo utilizes SDP to exchange EPRs. [20] provides more insights and details on this process.

During the session setup, the SIP-SOAP proxy performs a subscription, with presence information, to the service SIP. A URI is offered by the terminal. Thus, the domain will be aware of the availability status of the terminal and may notify changes to the Grid Service Provider (SP). Note that the terminal may also be interested in knowing the status of services offered by the Grid SP, but the Grid SP status is static compared to services offered by the terminal. Therefore, Grid SPs subscribe to the terminal status (via the SIP-SOAP proxy), but the opposite is not necessarily needed. If the information is related to service changes, the Mobile Terminal will send a status change and, eventually, a reINVITE message containing new parameters for this session. The Grid SP will be informed on this situation by a WS-NOTIFY SOAP message, whose content indicates a change in the service.

Since the SIP-SOAP proxy is aware of the terminal status, it can use this information, for example, to inform the Grid SP, if (a) the user disconnects from one location and reconnects from a different one or (b) with a different terminal (or, simply, if the user just disconnects). Using this feature, the Grid SP can take an appropriate decision, like achieving services’ EPRs at the new service location. Eventually, a new session can be setup with the user’s new terminal or with the new terminal’s location (typically an IP address). This design, used in Akogrimo, provides WS with an unprecedented dynamicity, profiting from these SIP features.

One of this method’s applications is to bring mobile users into the WS’s infrastructure. Therefore, mobile users can be used as the “logic” – in the sense of the “mechanical turk” – behind a WS interface, providing corresponding services in their area of expertise, e.g., first-aid.

The SIP-SOAP proxy provides a WS interface, which can be complemented with several operations, beyond the approach presented, which leverages it to allow for the creation of Grid sessions. While other proposals rely on specific WS-based signaling (like the Suspend-Relocate-Resume for the WS framework [11]), in the Akogrimo demonstrator, this WS interface was enriched with a WS control of other SIP-based applications. The best example is the utilization of the Third Party Call Control (3PCC) mechanisms to allow for Grid-controlled multimedia sessions between users. Finally, for a third party to setup a SIP session between two users, two options are available: using SIP core methods, as described in [6] or using the SIP REFER method. In Akogrimo the latter is used. When 3PCC operations of the SIP-SOAP proxy are invoked, utilizing SOAP signaling, the domain produces REFER messages, which reach the requested users. As a consequence, a media flow between those users will be established. Additional SIP signaling is applied to inform the SIP-SOAP proxy on the status of the 3PCC session. Upon a successful session establishment, the domain will send a SOAP message to the Grid SP, indicating the success of the operation.

### 3.3 The Network as a WS

The Akogrimo network architecture follows key ideas form the “Next Generation Networks” field, where IP (version 6) is native and any access technology is supported. Terminals can reserve network resources using RSVP (Resource Reservation Protocol) [22]. For Akogrimo, as well, core routers are only DiffServ-enabled [7], while IntServ and RSVP are only supported by access routers at the network edges, which are used by customers’ terminals to gain network connectivity. In addition, network routers are controlled by a bandwidth broker. This infrastructure builds the Network “Area” of the Akogrimo architecture (cf. Figure 1).

The bandwidth broker, mainly controlling routers, exposes interfaces to service providers controlling the provisioning of the service. These WSs run by service providers would need to be modified to include the Diameter protocol [10] to interact with the Bandwidth broker. Instead, Akogrimo offers a “QoS Gateway” (also named QoS proxy) solution, which exposes network features as being “WS ready”, to be invoked using SOAP [19]. In other words, Akogrimo exposes the network infrastructure (controlled by the Bandwidth broker) as another WS. The actual service offered transports information throughout the network with different QoS. This service can be integrated and composed with other WSs in the OpVO.

As explained below, the design and instantiation of an OpVO has to be made according to dynamically negotiated SLAs. Thus, an Execution Management System (EMS) is needed that should include the functionality of reserving network resources according to the SLAs established (cf. Section 3.4). To do so the Execution Management System has to interact with the Network QoS subsystems in order to check the availability of network resources and to request for these resources during the entire lifetime of a service execution.

Thus, the QoS gateway is exposing these QoS management functionalities as WSs, which is invoked by the EMS. The functionality is abstracted with the following two WS:

- A WS for checking the network availability is used in the negotiation and discovery phase of the execution management system for checking that the SLA of a specific service can be fulfilled.
- A WS for making the network resources reservation is run, as needed by a particular SLA of a service.
The QoS proxy is offering different network QoS bundles to the EMS, which can be checked and reserved, specifying QoS parameters for different combinations, e.g., data, audio, and video traffic. These WSs are translated to the Diameter protocol for interacting with the Bandwidth Broker, which, according to predefined network policies, will operate the QoS management interaction with network resources under consideration.

### 3.4 Cross-layer Execution Management Framework

The design and implementation of a cross-organizational collaboration infrastructure cannot be based on a best-effort approach, since the design decision to run without a central control instance needs the support of loose and dynamically changing coalitions of different service providers and potentially a large number of consumers. The approach chosen within Akogrimo, as within other service-oriented Grid research projects, such as [16], [17], and [18], is to safeguard all interactions by dynamically negotiated SLAs. These SLAs typically refer to existing paper-based contracts in order to be legally binding. The agreed upon SLA needs to be expressed on a business level, understandable to customers and providers, in order to enable a supervision of their fulfilment by both contracting parties and potentially a trusted third party acting as a clearing house for this interaction. The abstraction of SLAs from the underlying infrastructure is needed. At the same time they have to provide those services as essential for achieving a decoupling of resources and services. But this leads to the challenging problem that agreed upon SLAs, on a business level, have to be mapped to the level of the concrete infrastructure for their execution. Additionally, services to be managed might be provided by mobile entities with a frequently changing context. Therefore, the specific challenges on an EMS within Akogrimo are as follows:

- External guarantees toward the service consumer are expressed on an abstract level allowing the consumer or a trusted third party to validate the fulfillment of these SLAs.
- Services used to aggregate a service offered to the consumer are not only built from services hosted on resources operating in a fixed environment and are not necessarily available from services within a local network or administrative domain.
- Service consumer and providers might be mobile and may transfer their sessions between devices, they change their location, access network, or even partially go offline.

Compared to other service-oriented Grid solutions, these requirements impose the ability to extend the reallocation capabilities from the monitoring viewpoint and also to violations on the network layer and to treat changes in that context beyond simple Boolean conditions such as “operational” or “failure” to a more continuous space of changing capabilities and delivered service quality.

The mapping problem of these high level SLAs to infrastructure level metrics is addressed within Akogrimo in a way that limits the number of pre-defined SLA templates: Those are defined with a mapping to the resource level beforehand. Such an assumption limits the management space and is valid for a good number of cases only. As the underlying provisioning infrastructure remains volatile, a close interaction of the EMS with the network provider and the SIP-SOAP proxy is needed (cf. Figure 7). During an initialization phase of the service provisioning, not only an appropriate reservation of resources, such as computing and storage, is needed, but also required capabilities on the network level, such as bandwidth and latency, need to be ensured. Additionally, availability and capabilities of mobile providers and consumers need to be collected via a subscription to the SIP-SOAP proxy.

The EMS needs to know how to react on SLA violation messages from different sources in order to provide a new solution, while applying additional internal resources or by communicating supporting services of the BVO, such as the Service Discovery module, which identifies potentially new service providers that can replace the faulting service provider. Besides the status information of mobile service providers communicated via NOTIFY messages from the SIP-
SOAP proxy, it needs to be considered that appropriate actions need to be taken. *E.g.*, if the SLA between a mobile user and the service provider under the control of the EMS has agreed that the data stream is consuming between 60% and 80% of the overall available bandwidth of the device and a change to the access network leads to a significant change of the available bandwidth, the EMS needs to make sure that providing services are receiving new set-up messages, changing their behavior or an internal re-allocation of resources is necessary to accommodate these new conditions.

**Figure 7 Execution Management System Interactions**

3.5 *INTEGRATING GRIDS IN A MOBILE AND COMMERCIAL FRAMEWORK*

Authentication, Authorization, and Accounting (AAA) for users and services, independent of their nature, is assured by the AAA System part of the “Network Middleware Area” of the Akogrimo architecture (cf. Figure 1). An authentication to the network implies that the user can be authorized to use network services such as connectivity, but also other services following the SSO paradigm. This approach requires that services have to be modified in order to be able to allow access to network-authenticated users. The rational of choosing the network-based AAA system for authentication instead of other mechanisms from the Grid or WS world is driven by potential customers, contracts, and information that telecommunication operators already have installed, which would not need to be rebuilt, if a new architecture and protocol was to emerge. Although not treated in depth, Akogrimo also identified the eventual necessity of an external identity provider, which could offer this specific role in lieu of the network operator.

The overall AAAC architecture for Akogrimo [30] is based on the IETF AAA Architecture proposed by the Internet Engineering and Research Task Forces (IETF, IRTF) [10] and has been enriched with a “Charging” functionality. Thus, AAAC is a generalization that supports the Internet Protocol version 6 (IPv6). The fact that the AAAC service is used in a QoS-enabled Mobile IPv6 environment has been considered in this architecture to provide features, which enable new functionality, as well as performance optimization of the overall system.

Figure 8 generically depicts these enhancements to the AAA Architecture. The new AAAC System includes the traditional AAA Server with those extensions named, explicitly covering charging in a new module, metering in a new module, and auditing functions beyond the AAAC system [28]. Auditing enables further functionality, necessary for commercial applications, mainly with respect to the evaluation of audit trails generated by the AAAC System and other entities under audit. Such audit trails are based on SLAs negotiated and enable an automated compliance check between service parameters delivered and contracted [29]. Within this context the Policy Repository is considered as being part of the policy-based AAAC System. In support of mobility, an AAAC attendant handles the interface with the Mobile Terminal and an Application-specific Module (ASM) communicates with the Bandwidth Broker. Furthermore, interactions with additional AAAC systems residing in different or the same domain are run similarly to the traditional AAA Architecture, in Akogrimo’s case applying the Diameter protocol.
The Akogrimo project successfully deployed an integrated demonstrator, which is based on those key innovative concepts described above. This integrated demonstrator has to be considered as a functional proof-of-concept not addressing global scalability issues nor performance optimizations, but as a validation platform addressing the functionality of the new and advanced Akogrimo architecture.
4.1 APPROACH AND GOALS

All architectural and protocol innovations have been validated in the scope of several test-beds, which have been deployed and operated during the project’s lifetime. The design of the test-bed was made, while taking into account the business modeling of the Mobile Grid environment, in order to identify most adequate areas for the Mobile Grid technology. As a result, several business areas were identified, each defining a potential scenario for the operation of the Akogrimo architecture:

- **eLearning**, where students and teachers can access and provide services dynamically in the scope of an eLearning base VO
- **eHealth**, where patients and doctors can compose a dynamic scenario for accessing and providing healthcare monitoring services
- **Disaster Handling and Crisis Management (DHCM)**, where a scenario (Virtual Organization) is created dynamically in the case of an emergency, composed by the most adequate actors available and suitable for that emergency (e.g., in terms of firefighting, paramedic, and police)

These scenarios have been designed and the basic Akogrimo architecture was instantiated in terms of its application to these specific scenarios. The specific configuration of the architecture for the test-bed includes:

- The definition of the specific BVO, SLAs, profiles, policies, and all administrative issues for defining roles and configuration of involved parties in this scenario.
- The integration and deployment in the generic architecture of specific test-bed applications (clients and services), while the generic platform provides clients for video conferencing and WS-based services.
- The definition of a specific application workflow, including all interactions needed for OpVOs.

The Akogrimo architecture was designed in order to provide the infrastructure needed for application-specific clients and services. Thus, mobile and wired clients run SIP-based Akogrimo mechanisms, which also contain all interfaces needed for accessing the Mobile IPv6, the AAAC, and the QoS infrastructure. Additionally, generic interfaces for a login and video conferencing are included. Furthermore, application-specific services are operated on servers, which run the Akogrimo infrastructure and which enable access to them, including their interfaces to the EMS. Finally, metering modules integrated do monitor SLAs parameters. To describe in all necessary levels of detail the scenario demonstrated, the DHCM scenario has been chosen. This case determined the final complex demonstrator, integrating all key features designed and developed in the context of the Akogrimo architecture.

4.2 PHYSICAL MAPPING OF COMPONENTS

The architecture and all applications were mapped to the physical level onto separate machines in the form shown in Figure 10. Each light blue box denotes a software component implemented and applied to architectural as well as application components. All architectural components are those, which are part of the Grid Middleware Services (cf. Figure 1), whereas application components are applications utilizing that middleware, such as the eLearning, eHealth, and DHCM. These applications components reside within terminals and the service domain.

As shown in Figure 10, the overall architecture can be split into different areas:

- **Terminals**: Software in mobile and fixed terminals is classified into common software, base software, and application-specific software. The base software contains everything needed for terminals to function in the Akogrimo infrastructure, such as SIP components used to handle SIP signaling involved in many operations, e.g., context information propagation or multimedia conferences. Between SIP components, the component named “gSDP Server” (gSDP stands for Grid-enabled SDP) can be highlighted, since it is in charge of all SIP-based Grid sessions, allowing the mobility of WS offered by all terminals. In order to achieve a successful authentication, terminals run a PANA client [8], which is supporting the SSO functionality, e.g., the utilization of SIP is granted based upon a token linked to the previous authentication at the network level. Credentials are provided by the user by means of the login application. Thus, the network provider manages the user’s identity. The packet metering and QoS module provides necessary hooks for the terminal to manage the reservation of network resources, taking into account network mobility, which is carried out by the MIPv6 module. Finally, the application-specific software on terminals changes depending on the role of the terminal in the demonstrator. Apart from a specific graphical user interface, each role needs a different set of services and applications. Terminals run application software in order to collect information provided by RFID tags (Radio Frequency Identity).
- **Access Networks**: All access networks are used by mobile and fixed terminals to access the Akogrimo infrastructure and consume services offered by SPs. Beyond their basic functionality, access routers define also the point of authentication for terminals and apply policy enforcement as well as QoS management tasks.
• **Core Network**: Along with access routers, the core network incorporates the key for the Akogrimo infrastructure and is identical for all scenarios defined and applications used. The core network hosts several elements, which are essential for the operation of the Akogrimo infrastructure. These elements support the key functionality needed, such as the orchestration of different services and resources, to solve a certain situation, such as a workflow management, the deployment and execution, an OpVO brokerage, or a participant’s registration. The GrSDS (Grid Service Discovery Service) and the context manager support the workflow creation and execution processes by providing service discovery information and context aggregation as well as event notification, respectively.

Other important tasks are related to the exposition of the network functionality, such as SOAP interfaces, directly usable by the Grid WSs. These tasks are supported by the QoS Grid Gateway and the SIP-SOAP proxy, which allow Grid WSs to configure network resources depending on service and application requirements. Additionally, these tasks interact with the SIP infrastructure whenever needed. Finally, core servers are also located in the core network. Examples of these include the SIP Server, the central element of network signaling, and the Policy-based Network Management (PBNM) Server, which handles policy-based aspects of the network’s behavior, the AAAC Server, and the SAML (Security Assertion Mark-up Language) Server, which the user identity management relies on.

**Service Domains**: Two service domains are needed to meet all requirements of the final DHCM demonstrator. The service domain represents the infrastructure of the SP. This infrastructure comprises elements related to the planning, scheduling, execution (EMS), monitoring, and management of services offered.

Along with the provision of these services themselves, the creation and management of SLAs is also of importance to the SP. This includes the monitoring of SLA compliance as well, by applying the AAAC functionality. Thus, several SLA-related components are found in the service domain.

4.3 **SCENARIO**

The operational setup of the demonstrator was distributed between Madrid (Spain) and Stuttgart (Germany). While Akogrimo’s core platform and specific services were run at the University of Stuttgart’s Computing Center, user terminals and access routers were located at the premises of the Technical University of Madrid. These two test-beds were connected using IPv6. Since an external network was used, QoS was only guaranteed locally between terminals in Madrid, although it was managed from Stuttgart. Such an infrastructure distribution shows the viability of the Akogrimo’s architecture instantiation as an operational scenario, where the Akogrimo operator does not own the entire network or a subcontracted network does not offer QoS services.
1) Story Line
The story board for the DHCM scenario works on a simulated dirty bomb causing both, infrastructural damage and human casualties. Both problems need to be identified and treated as soon as possible. In the first step an observing policeman alerts the emergency system by contacting the emergency control center, which evaluates the situation with a risk assessment manager. It triggers the corresponding workflow to deal with this new situation. The workflow is understood as a BPEL (Business Process Execution Language) workflow, which manages a logic and automated process of events based on WSs. In the second step, fire-fighters identify casualties and feed the risk assessment manager. In the third step, a paramedic is guided to the location of the each casualty, identified by the emergency system, which also provides the paramedic with an assessment of information on prioritizing injuries and their treatments. Applications involved take over the uploading and downloading of the multimedia information obtained, the risk assessment undertaken, and all results of injury assessments performed.

2) Set-up of Virtual Organizations
Akogrimo builds up a dynamic collaboration following the paradigm of the VO. As explained in Section 3.1, the BaseVO is taken for granted as a first step of contact, where the user becomes a member of the platform, exchanges credentials and profiles, and policies are created together with the User Agent. For the second step, when actions are required in order to accomplish a concrete task, suitable participant members of a BVO are searched for and any matching players available are created and placed in contact via the private OpVO. In this DHCM scenario three different OpVOs exist and are called: Bronze, Silver, and Casualty Treatment OpVOs.

3) Operation of the OpVOs in Terms of the Workflow
The Bronze OpVO is a long-term running OpVO used by the policeman when patrolling. In order for the OpVO to be activated, the policeman has to login to the system on his terminal. The system provides adequate network QoS characteristics, single-sign-on – an authentication in the network is propagated to upper-layers –, and takes over the propagation of presence information to the system. All these processes have been described in Sections 3.2, 3.3, and 3.5. The Bronze OpVO is static and the Workflow engine is not involved. Invocations are done directly between different participants and services. Once the policeman witnesses the bomb explosion, he contacts the emergency call center by means of his SIP-enabled device and uploads graphical information (pictures). After a risk assessment, the emergency call center starts two OpVOs to handle the situation: the Silver OpVO and the Casualty Treatment OpVO. These OpVOs are created, taking into account SLAs and specific policies requirements that have been defined previously. In particular, during the creation of these OpVOs, it is required to interact with the Network QoS subsystems in order to check and reserve appropriate network availability and quality parameters as stated in the SLA (cf. Section 3.3).

The different ingredients of the Silver OpVO are supported by a workflow, which orchestrates the collaboration between them. All communications with mobile and nomadic users is done by WS using the SIP-signaling approach (cf. Section 3.2). In this OpVO, a silver command alerts available fire-fighters and paramedics and transmits information.
about the disaster. The silver command establishes a video connection with available surveillance cameras in that area (by using SIP-based Service Discovery) and after an analysis of the scene, it instructs fire-fighters to enter the disaster area. Upon entering the connectivity is kept, even while changing access networks due to Mobile IPv6, fire-fighters send their location to the system by applying a SIP-based context and presence publishing approach. The system runs a risk assessment of that area and automatically triggers notifications of nearby risks to all fire-fighters. Upon determining casualties fire-fighters send information on those (including the identification of RFID tags, which have been attached to these discovered casualties). The Silver OpVO workflow contacts in turn the health data and patient information services from all available services and, finally, triggers the Casualty Treatment OpVO.

Within the Casualty Treatment OpVO a new player is involved: the paramedic. The workflow automatically sets up a SIP-based multi-participant conference between fire-fighters, the paramedic, and surveillance cameras involving all processes as described in Sections 3.2 and 3.4. The connection is kept alive despite any mobility due to MIPv6 [31], whereas a stateful change of terminal is possible due to SIP. The access to Grid resources in mobile terminals is also solved using the SIP and SOAP interaction. Upon a context change in which the system identifies that a paramedic is close to a casualty, the information provided by fire-fighters and other records are conveyed to the paramedic’s terminal. Meanwhile, the workflow continues to check bed occupancy at near-by hospitals and to notify the paramedic about the hospital for the casualty.

5 Evaluation of Key Elements Demonstrated

The evaluation of such an overall platform and architecture does contain a number of different dimensions. As indicated the key purpose of this paper is to describe, in which way the integration of those new innovations have been performed. Therefore, the functional evaluation of those key elements developed fits suitably in a platform design paper.

Automatic Collaboration

Different virtual organizations are created for a scenario in order that those can be put into contact to each other and collaborate. The base contract inter-relates all available organizations to the same pool of organizations in terms of the Base VO. Upon a scenario problem to be solved, appropriate organizations are brought together in the OpVO. The DHCM scenario demonstrated this by launching three different types of Operational VOs. Therefore, the use of VOs for automatic collaboration of mobility-enabled users in such a dynamic scenario and for bootstrapping multimedia collaboration is a novelty contributing a large potential for beneficial IT support of the society.

Service and User Mobility

The Akogrimo architecture allows for nomadicity and mobility not only of users, as traditional architectures envisage, but also for services. In particular (stateful) WS can be searched for, independent of their physical location, due to the use of SIP infrastructure capabilities. Apart from nomadicity, MIPv6 with its fast-handover functionality enables a transparent mobility between different networks. In the DHCM scenario fire-fighters and the paramedic are able to enter the area, while constantly receiving relevant information due to the integrated MIPv6. Thus, MIPv6 enables seamless terminal mobility, while SIP supports user and session (changing a terminal without terminating an ongoing session) mobility. Mechanisms still need to be developed in order to provide the user with a real end-user mobility, where the user can change its end device on-the-fly and the context or state is transferred from one device to the other allowing seamless terminal changes.

Service Abstraction

Services, independently of their nature, are abstracted as WS. This enables network resources such as connectivity, QoS characteristics, network information, multimedia services, and network mobility to become part of a generic workflow. The scenario shows how multimedia conferences are triggered from a WS–based workflow (such as the one amongst fire-fighters, paramedic, and surveillance cameras), together with the corresponding QoS parameters required by the multimedia conference. A QoS level is also enforced upon sending or retrieving large amount of data, e.g., in the case investigated, the bandwidth. The service abstraction can be extended with knowledge-based techniques by applying ontologies, which are setting relations between different components and which would set the grounds of automatic service composition.

Identity and Federation

Different administration domains come into play in the DHCM scenario: a network provider, a hospital, fire-fighters, and the police. In order for the cooperation between VOs to be possible without security leaks the need for identity federation will enable authorizations on different levels of information granularity, such as the one provided by the hospital to the paramedic with relation to a certain casualty. In Akogrimo this is accomplished by means of the previously mentioned AAAC server/client system, which utilized extended SAML attributes. This method enables the integration of the network roaming and SSO at those two different layers of network services and WSs.

Workflow

The utilization of the system is based on a series of workflow templates, which orchestrates the abstraction of services and coordinates users that are involved. Upon an event the best suited workflow is chosen, either manually or semi-automatically. In this case, the operator at the emergency call center runs the risk assessment and the respective workflow
is chosen: Silver OpVO and Casualty Treatment OpVO. Once these templates are selected, the service discovery procedure finds best matches and instantiates the workflow and its services. At this time, the set-up of these OpVOs takes time, which is depending on the complication level of the scenario and those interactions required. Optimizations are possible by applying the service discovery mechanism and the BPEL implementation used.

**Context-aware Execution of Workflows**

A running workflow can also take decisions on events that happen at lower layers. Bringing dynamic context information into workflow-driven executions is a major advance in the state-of-the-art. In the DHCM scenario the running workflow takes decisions upon context changes, such as the location of a participant, the type of device used by him/her or the state of the connections. In the DHCM scenario, the fire-fighter approaches an area of risk, which is detected by the Context Manager and the Workflow Engine is notified through the monitoring daemon. The BPEL workflow processes this information and notifies the fire-fighter. Therefore, Context information is integrated in the workflow. One solution for a versatile integration independent of the specific workflow has been studied, which uses different workflows running in parallel and their information basis.

**Cross-layer Integration**

One of the most important key points demonstrated and a main motivator is the viability and gains achieved with a cross-layer integration, in turn providing a consistent and optimized platform. Cross-layer integration elements have been shown especially through the following concepts:

- **Security and identity:** Users are able to access different services following the SSO paradigm that only requires one authentication at the network level. The network’s AAAC is considered as the identity provider. After a successful network authentication the AAAC authorizes the access to different services according to corresponding associated policies.
- **SOA-driven multimedia:** Parties involved in an emergency deployment can be put into contact automatically by means of SIP hooks to the SOA infrastructure. In the DHCM scenario, this was applied to the police, paramedic, and fire-fighter and the set-up of the multi-party conference by means of a 3rd party call control.
- **Network-Service integration:** The abstraction of the network as a service allows the application of service-related mechanisms to the network itself in a seamless manner. In Akogrimo, this enables bandwidth allocation, control, and WS mobility. In the final demonstrator, bandwidth reservation is used in order to upload pictures from the policeman’s device addressed to the emergency call center to obtain detailed information on the situation and to optimize decisions. QoS reservation is also used for the transfer of large files, such as patient records, as well as the multimedia conference mentioned. WS mobility is demonstrated, when the fire-fighter and the paramedic are presented to the system as services, which allows the platform to include them in a workflow to be orchestrated along other services.
- **Fault tolerance:** Services that may show no current network connection or simply stop working are recognized and a new service is searched for by means of the service discovery mechanism that connects the new service provider to the VO. In the DHCM scenario, as soon as a fireman went off-line, the EMS looked for a new service substitute, e.g., another fireman logged into his device, and directed all information to him.

6 **CONCLUSIONS**

This work describes the comprehensive and successfully implemented, integrated, and demonstrated Akogrimo architecture merging large scale properties of telecommunication platforms and their support for multimedia applications with the cross-organization collaboration concepts of service-oriented Grids into a coherent and integrated service platform. This architecture defines the first approach to extend the definition of traditional Grid computing toward a mobile collaborative business Grid. It covers the dynamic collaboration across organizational boundaries, cross-layer cooperation from network to application, and considers the impact of a potentially very large number of mobile users using a wide range of devices and access networks.

The key problems solved by this architecture address successfully the seamless integration of both mobility and Web-“Servicification” of any kind of resources, including the network. Within this Akogrimo architecture two types of mobility are considered: (a) nomadicty, where users and resources can sign-on and off from different locations or terminals and (b) true mobility where the user and resource can change the access network without interrupting an operational session.

The Akogrimo architecture is flexible, open, and provides a clear separation between technology, administrative domains of a network operator and a service provider, the customer domain, and the two different instances of Virtual Organizations (OpVO and BVO). The Akogrimo architecture has the potential to introduce sophisticated services in a “beyond IMS” infrastructure providing added-value to customers generating an increased willingness to pay.

The full integration with WS could enable, e.g., service discovery operations, which are important from a commercial point of view and which are based on a standardized procedure. The Akogrimo project has shown that these concepts are highly applicable in up-to-date scenarios in order to find a successful entry into a newly emerging WS market. In order to achieve added-value – based on WS – within the mobile network, the concept developed for the SIP–SOAP integration is ready to be applied. Thus, the designed, implemented, and successfully integrated solution shows a very promising
start. Since the Akogrimo architecture never targeted to replace an IMS infrastructure, but to enhance it, key investments already performed in the IMS area can be protected and an operator as well service provider sees the open flexibility to seamlessly integrate those mechanisms proposed into his service platform for various service enrichments.

7 ACKNOWLEDGEMENTS

The work presented was funded by the EU Project IST-2003-004293 “Akogrimo” (Access to Knowledge through the Grid in a Mobile World). All authors would like to thank all their colleagues that contributed to the project.

8 REFERENCES


