Innovation and the Knowledge Economy: Issues, Applications, Case Studies
Paul Cunningham and Miriam Cunningham (Eds)
IOS Press, 2005 Amsterdam
ISBN: 1-58603-563-0

Specific Challenges of Mobile Dynamic Virtual Organizations

Christian LOOS[1], Stefan WESNER[2], Jürgen M. JÄHNERT[3]
[1]*Universität Hohenheim, Information Systems II, Stuttgart, 70599, Germany*
Tel: +497114593786, Fax: +497114592961, Email: loos@uni-hohenheim.de
[2]*Höchstleistungsrechenzentrum Stuttgart, Nobelstrasse 19, Stuttgart, 70550, Germany*
Tel: +497116854275, Fax: +49711682357, Email: Wesner@hlrs.de
[3]*Rechenzentrum Universität Stuttgart, Allmandring 30, Stuttgart, 70550, Germany*
Tel: +497116854273, Fax: +49711682357, Email: jaehnert@rus.uni-stuttgart.de

Abstract: Mobility adds specific new challenges to Virtual Organizations (VO). The challenges for Membership Management and Service (Provider) Selection is significantly different compared to traditional Virtual Organisation (VO) in particular to static infrastructures typical in research Grid deployments. Having VO nodes that change their context such as location, device capabilities, connection quality etc. affect the way how tasks can be distributed. Furthermore nodes can even go offline without notice for an undetermined time. This paper discusses the specific new challenges of Mobile Dynamic Virtual Organizations (MDVO) aligned with a generic scenario from the e-Health domain and how the Session Initiation Protocol (SIP) in particular combined with SIMPLE (SIP for Instant Messaging and Presence Leveraging Extensions) is applied.

1. Introduction

Mobility adds specific new challenges to Virtual Organizations (VO). Based on a generic scenario this article will outline these challenges and describes how resulting requirements can be addressed using Grid and Network level technologies.

The first part of this paper will provide a definition of a Mobile Grid and how it is different from existing Grid infrastructures. In particular it will discuss new challenges arising from intermittent network connections, device and location mobility and other user context specific properties that need to be considered in this kind of environments.

The next part of the paper will put these properties into context using an emergency scenario worked out within the Akogrimo project (http://www.akogrimo.org). This mobile world scenario situated in the e-Health domain will clearly show the need for Mobile Grids and the properties discussed before, and will summarize the innovations needed in order to realize these new kinds of Virtual Organizations.

The third part of the paper will introduce existing mechanisms for the management of Virtual Organizations in Research Grids as well as recently proposed solutions in the area of Business Grids. Guided by the envisaged scenario their shortcomings are pointed out.

Based on this analysis the proposed solution from the Akogrimo project combining the session and device capability management functionality of the Session Initiation Protocol (SIP) with a Web Services based Grid is explained in the forth part of the paper. The paper is concluded by assessing how this novel approach could co-exist with non-SIP based Virtual Organizations.

2. What is a Mobile Grid?

The Grid is about hiding the distribution and the access of resources, the processing of tasks and the coordination of performing these tasks. Therefore in an ideal case the Grid is fully transparent to the End User providing a platform for efficient and effective cooperation. All these properties apply in the same way also to Mobile Grids. However, in addition to this, in a Mobile Grid, at least some of the Grid services are *mobile*. Mobile could either mean being moved together with the device that is hosting this service to another physical location (in a different environment and context) or being moved from one host to another service host. The latter case might happen if sessions (or parts of it) are transferred e.g. from a fixed computer to a mobile device. It is important to note that “mobile” does not necessarily imply "wireless" or "low bandwidth" or "can be carried by a person in the pocket". An example case could be to accept to join a session with a PDA while travelling and moving the session to the desktop computer on arrival at the office. Of course these changes of the capabilities (network bandwidth, display size, etc) must be reflected in the context of this service. The first case of geographical mobility requires at least additionally the virtualization of the location as part of the service context. This service context must be made available to other VO participants e.g. by exposing the state (using e.g. the Web Service Resource Framework), or a providing mechanism to query or receive events on these changes in order to allow other services to adapt to this context changes.

Definition – Mobile Grid Resource: A Grid resource is named *mobile* if the Grid resource and the respective Grid service interface are either not bound to a certain location and/or to a certain device.

Based on this definition we now approach a definition of a mobile Grid.

Definition – Mobile Grid: A mobile Grid is a Grid with at least one *essential* Mobile Grid Resource. Essential in this context means that the mobile Grid resource is actively participating in the Grid so that the resource can take the role of either a service consumer or a service provider. Consequently, the mobile Grid resource is an active member of at least one Virtual Organisation and is involved in executing workflows.

3. Mobile Grid Emergency Scenario

The application of Mobile Grid technology in the eHealth domain was influenced by two recent developments. At first, people are more mobile then ever before. Their health travels with them, a provision of health care services is expected everywhere. As a second point, the ongoing miniaturization of medical equipment for monitoring, diagnosis and treatment activities is of high relevance. Particularly, within the case of an emergency, both issues have to be combined at its best. Mobile Grid technology offers the possibility to make all available information, resources and knowledge available on site of an emergency. Historic patient data are collected throughout the patient records distributed in the health care provider’s databases. These data are prepared specific to the emergency. Joint with current patient data (e.g. ECG, ultra sonic, x-ray data) gathered from the respective mobile devices (monitoring equipment from the patient and devices within the ambulance car), they enable a comprehensive assessment of the emergency situation. Besides this information, paramedics, emergency physicians, medical experts and the emergency unit of the destination hospital are also parts of the established VO. This enables the hospital to be prepared when the ambulance is arriving. Crucial to this scenario, the access to these services may not hamper the treatment of the patient. Therefore, voice recognition and text to speech services are needed for interaction. For realizing this scenario virtual organizations should have the following particular properties:

- situation specific integration of available resources and access to services provided (on site of the emergency, in the ambulance, in the hospital)

- role specific (ambulance driver, paramedic, physician, expert, hospital admittance) access to services and information (privacy of patient data)
- concurrent usage to mobile resources (e.g. monitoring equipment)
- support of diverse network links and consequently, the adaptation of the quality of the provided services and the access methods to them
- ubiquitous access (e.g. through wearable devices on site, devices installed in the ambulance, desktop PCs in the hospital) to mobile and fixed services

As an example the following scenario can be considered. It is partially used in Section 7 in order to explain the role of the components additionally needed in Mobile Dynamic VOs.

A Spanish architect, on a European business trip, is on a building site when he suddenly feels strong pain in the left side of his chest. He uses his mobile phone to call for assistance and transmits his cardiology data with the help of the ECG equipment integrated into his phone. His data is received in the Emergency control centre. There, a virtual environment is immediately created, in which all information needed for optimal care of the patient is made available. The Grid service automatically discovered and embedded in the Virtual Organisation analysing the cardio data triggers a Notification to the Control Centre that an emergency team should be sent immediately. The emergency team is integrated into the VO with their on board equipment and audio/video devices. Upon arrival at the patient the high quality equipment replaces the mobile phone. This context change triggers actions on VO level. The mobile phone device is removed from the VO and replaced by the new equipment. The better data allow different analysis to be performed so new services analysing and storing the data are added. Furthermore the resulting data are displayed on the emergency car equipment and in the hospital at the same time…

4. Limitations of Existing VO Models

Initial models for Virtual Organizations have been built in order to meet the requirements of the research community. This kind of Virtual Organizations are often very static in its nature meaning that adding new users or even new organizations is a process that is not built in the operation model of the virtual organization process but is often a out-of-band process that might even involve a non automated process using paper forms. Beside the focus on resource sharing meaning that typically members of such a VO share the common goal and should be allowed to have access to all resources shared. Even if originally these kind of VOs have been very small because they have been limited to High Performance Computing users they recently reached a significant size through the DataGrid and EGEE (http://www.egee.org) projects aiming at setting up an European wide infrastructure for researchers. Similar to this approach the DEISA project (www.deisa.org) is aiming at a “virtual computing centre” combining HPC resource across Europe within a Research Grid.

In contrast to this, the notion of On-Demand created Business Grids has emerged recently. This concept as addressed by projects like GRASP [5] and TrustCoM [7] focusing not on long term fixed memberships but on on-demand extensions and reduction of membership combined with highly capable security systems that go far beyond what is implemented in current Research Grid infrastructures. As the Research Grids has reached a status where the problems are now more on administration and scalability rather than on the conceptual layer the Business Grids still need further research to be exploited in business to business interactions. Most of these Business Grids VO concepts combine the VO model from Grids with models from commercial contexts (see [8] [9] [10])

However, these models do not reflect the additional needs from mobile grid environments. The Research Grids with there static nature cannot accommodate as needed to context changes. As an example consider the need for utilizing local service such as an internet enabled display. This service is provided by a company that does not share at all any goal of the Virtual Organization. The motivation to participate in the VO is simply offering a display facility to customers. So in order to support this scenario it must be possible to add this service as a resource to the virtual organization on demand for a limited time. And of course this should mean that the provider of the display service is allowed to use other resources within the VO as needed. One could say that Business Grids are also addressing these kind of scenarios of ad-hoc business interactions but they are not designed (yet) to support service discovery and adaptation based on volatile contexts and furthermore the fact that VO participants may disappear without notice (e.g. simply getting disconnected in a tunnel) are not considered.

5. Potential Role of the Session Initiation Protocol

SIP (Session Initiation Protocol) is an application-layer control protocol which is specified in [1]. The protocol is used for creating, modifying, and terminating sessions with one or more participants. Within this context a session comprises a set of senders and receivers that communicate and the state that is kept in the senders and receivers during the communication. Most relevant examples of a session are Internet telephone calls, distribution of multimedia, multimedia conferences, distributed computer games, etc.

SIP is only intended to control the signalling in order to enable the communication; the communication itself must be achieved by another protocol. For that purpose the RTP and SDP protocols can be used. RTP [2] is used to carry the real-time multimedia data including audio, video, and text. The protocol supports the encoding and splitting of the data into packets and transports such packets over the Internet.

In addition, SDP [3] is used to describe and encode capabilities of session participants. Such a description is used to negotiate the characteristics of the session that all the devices can participate. This includes the negotiation of codecs for media encoding, and the transport protocol used.

SIP is an end-to-end oriented signalling protocol. Consequently, the whole logic is stored in the end devices. Within the SIP specification the following 5 basic components are defined:

User agent client (UAC) – is an application that initiates up to six feasible SIP requests to a UAS. These six requests are: INVITE, ACK, OPTIONS, BYE, CANCEL and REGISTER. When a SIP session is being initiated, the UAC determines the information essential for the request (protocol, port and IP address).

User agent server (UAS) – is hosting the application responsible for receiving/answering the SIP requests from a UAC. The UAS may issue multiple responses to the UAC, not necessarily a single response. The communication between UAC and UAS is based on the client/server principle.

Proxy server – acts as a mediator that serves the requests or forwards them to other UASs or UACs for serving. The Proxy server can use an intra-organizational configuration where all its SIP communications (routes) are based upon. Intra-organizational configuration can be described when user messages are routed through a proxy server before the messages are relayed to the destination SIP client. This occurs when initiating a SIP session to another user within the same organization. This can be useful for internal communication where security over an internet link can be a problem. Further functions of the Proxy server are e.g. name mapping and resolution where a proxy server can query a location service and map an external SIP identity to an internal SIP identity.

Redirect server – allows for redirection which enables users to temporarily change geographic location and remain contactable through the same SIP identity.

Registrar server – supports to alter the address at which users are reachable. This is possible through the SIP client sending a REGISTER request of change of an address to the registrar server, which then accepts the request and records the user's new address. There are two ways in which the SIP clients can contact the registrar server. The first way is through a direct approach, by utilizing information that is configured into the client. Secondly through an indirect approach, which users the multicast address to contact the registrar server.

Within the SIP IETF community the SIMPLE (SIP for Instant Messaging and Presence Leveraging Extensions) working group has been established working on an architecture for the implementation of a traditional buddy list-based instant messaging and presence application with SIP.

6. SIP/SOAP Integration Possibilities

As mentioned above, SIP is the de-facto standard protocol for session setup and session management. Furthermore, the SIP architecture provides services like localization and aligned with that, mobility management. SOAP as a higher level application protocol and is one of the basic building blocks for Service Oriented Grids. Within the Akogrimo approach, it was identified that both protocols, SIP and SOAP, despite of coming from different traditional areas, have some overlapping functions and are to be integrated in order to merge these two worlds. Within Akogrimo the options shortly described below has been investigated.

6.1. SIP as Transport for SOAP

The first option would be to implement an additional protocol binding to existing Web Services Toolkits that would allow messages between Grid nodes not to be transported over HTTP as most toolkits do but using SIP. The advantage of this approach would be that existing application do not need to be SIP-aware and the integration with SIP would be fully transparent. However this would mean in the same way that the Mobile Grid infrastructure cannot easily interoperate with existing services or other infrastructures in particular if a hybrid scenario is envisaged where also public available Web Services are expected to be part of workflows. Additionally, even in a Mobile Grid not all nodes are mobile and the benefit of the SIP transport layer would not be visible.

6.2. SOAP as Transport for SIP

The other obvious option is to use the SOAP protocol as transport for SIP messages. This would require changes to the message chain of the Grid services being able to understand and create SIP messages as payload of SOAP messages (e.g. as part of the SOAP header). With this approach the information typically transported using the SIP protocol is added to the SOAP messaging world and would have no impact on the SOAP toolkit or communication layer. However this would mean that applications must be explicitly designed for using a mobile Grid infrastructure.

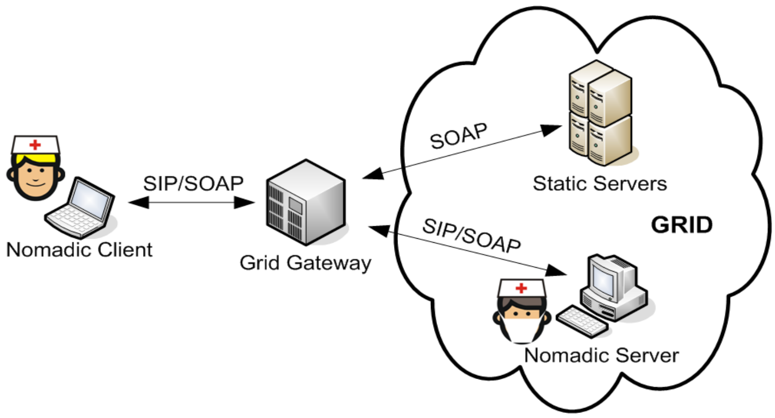

Figure 1: Basic View on SOAP/SIP integration

6.3. Using SOAP and SIP in Parallel

Of course another possible option is that a mobile grid node has two separate protocol stacks. One that is handling mobility, presence and multimedia sessions and another one responsible for the distributed application part utilising SOAP and Web Services.
Even if ideally integration between SOAP and SIP is achieved also this option has clear advantages. Existing Multimedia applications can be integrated easily and no interoperability problems with existing Grid infrastructures are expected. However following this approach strictly no integration of mobility relevant data is possible.

7. A VO model for Mobile Grids

So in consequence a model for Mobile Dynamic Virtual Organizations (MDVO) must have a lot of the properties of the dynamic flavour of Virtual Organization often referred to as Business Grids. However, there needs to be a mechanism for coping with a (possibly fast) changing context, discontinuation of the network connection (which can be seen as a specific context change) and the need to support location-awareness e.g. in the Service Discovery process.

Derived from this a MDVO needs to have additional components such as a Context Manager that keeps track of changing capabilities of VO participants using mechanisms such as detection of location changes e.g. through GPS but also the change of connection bandwidth and other network related properties combined with the device capabilities communicated through SIP. As outlined above the mechanism from SIMPLE providing information on availability (such as online/offline) is also captured here. As outlined in the previous section several possibilities for communicating this information on VO level are possible. Within Akogrimo a mixture between the possibilities described in section 6.2 and section 6.3 is used. For the sake of integration with existing SIP application such as audio and video communication systems all mobile nodes are required to support a SIP stack. Additionally all Mobile Nodes need to report their changing context to the Context Manager using SIP mechanisms with the already mentioned extensions coming from SIMPLE. However it is not planned to require SIP protocol support for all nodes. This is achieved by adding a function to provide a notification subscription mechanism for changing context to the Context Manager. The planned realisation is exploiting the specifications around WS-Notification [11] in order to allow all VO participants to subscribe to certain topics and react on the changing context. In particular, the VO Manager and BP (Business Process) Enactment components will need to react and trigger adaptation to this changing context. This approach can be seen as a realisation of the approach in section 6.2 as SIP information (and other context information such as location) are transported using SOAP.

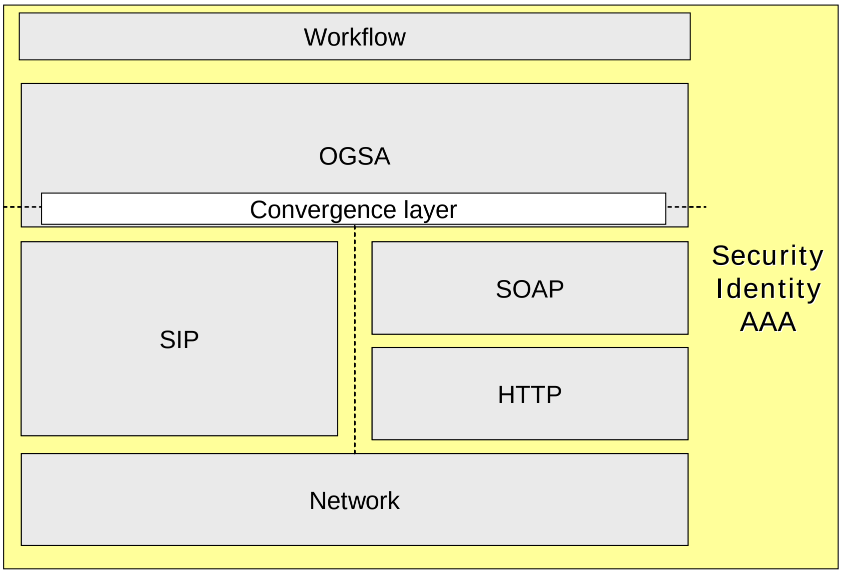

Figure 2: High Level Architecture View

The Context Manager Component can be seen as part of the convergence layer (Figure 2). In the example scenario above, the patient (VO member) using the embedded ECG device triggers the transfer of simple ECG signals through his mobile phone. For this situation the VO Manager has queried the Business Process Repository for the proper workflow. This workflow contains tasks enriched with metadata that allows the VO Membership Management component to identify appropriate services for this workflow using a Service Discovery Service. This workflow is given to the BP enactment engine. Part of the workflow the BP engine subscribes to context related topics at the Context Manager.

Component	Tasks
BP Registry	Contains workflows and workflow pattern that can be queried. It is assumed that these workflows do specify alternative paths to follow depending on context of the participants of this workflow.
BP Enactment Engine	This component is responsible for the execution of the workflow. It might not be performing itself the execution e.g. of a BPEL script but is subscribing to the Context Manager to notifications in order to trigger necessary actions (e.g. changing to an alternative path.
VO Manager	This component summarizes the functionality needed in order to trigger the discovery process of workflows and services, bookkeeping of membership and is also reacting on context changes (e.g. changing membership after a participant that is essential goes offline)
Context Manager	This component builds a crucial part of the convergence layer and is bridging the context information coming from the network layer (presence, location, network quality, etc) to the Grid layer.

Table 1: Summary of the VO Components

When the emergency physicians are arriving on site and they are adding a high quality electrocardiogram (ECG) device. Through SIP the new capabilities are communicated to the Context Manager which in turn generates immediately a notification to all subscribers. In our example we assume that the BP enactment engine has subscribed to this events and sets several conditions in the workflow to adapt to the better quality signal (e.g. introduce additional analysis services that get possible). Assuming that the VO Manager is also a subscriber of this event a VO level policy might define rules that after this event an external expert with certain device capabilities (e.g. display functionality for ECG signals and the derived data from the analysis services) needs to be added. This would trigger the discovery process for this participant, the whole set-up procedure and potentially the replacement of the existing workflow with another one. These components are summarised in Table 1.

8. Conclusions

The integration of mobility or more general context awareness into Virtual Organisations imposes new requirements on the infrastructure. Obviously existing High Performance Grids are designed for this purpose. Emerging Business Grid architectures are addressing a lot of functionality needed for this environment e.g. the dynamicity of membership. However they assume static properties of the VO participants or assume changing properties as a problem (e.g. bandwidth reduction) rather than being normal behaviour.

Context changes and adaptation to it is already a well known concept for technologies closer to the network layer. In this paper we have shown based on the example of SIP how such data can be integrated into a Grid architecture without imposing the requirement to build a completely new separate infrastructure but how these can be solved through a convergence layer between SIP and SOAP realised in the Context Manager component.

References

[1] J. Rosenberg et.al, *SIP: Session Initiation Protocol*, Internet RFC 3261, June 2002

[2] H. Schulzrinne et.al, Schulzrinne, H., Casner, S., Frederick, R. and V. Jacobson, *RTP: A Transport Protocol for Real-Time Applications*,RFC 3550, July 2003.

[3] Handley, M. and V. Jacobson, *SDP: Session Description Protocol*, RFC 2327, April 1998.

[4] I. Foster et.al., *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration*, 2002

[5] Theo Dimitrakos et. al., An Emerging Architecture Enabling Grid-based Application Service Provision, 6th IEEE International Enterprise Distributed Object Computing Conference (EDOC 2003), 2003

[6] J.Valles et.al., *The Grid for e-collaboration and Virtual Organisations,* eChallenges 2003

[7] T. Dimitrakos et.al., *Towards a Trust and Contract Management Framework for dynamic Virtual Organisations*, eAdoption and the Knowledge Economy: eChallenges 2004

[8] W. Saabeel et.al., *A Model for Virtual Organisation: A structure and Process Perspective,* Electronic Journal of Organizational Virtualness, 2002

[9] T.J. Strader et.al., Information Structure for Electronic Virtual Organization Management, Decision Support Systems, 1998

[10] L.M. Camarinha-Matos and H. Afsarmanesh, *A Roadmap for Strategic Research on Virtual Organizations, Processes and foundations for virtual organisations*, 2003

[11] Akamai Technologies, Computer Associates International, Fujitsu Laboratories of Europe, Globus, Hewlett-Packard, IBM, SAP AG, Sonic Software, TIBCO Software: *WS-BaseNotification 1.0 specification, WS-BrokeredNotification 1.0 specification, WS-Topics 1.0 specification,* March 2004.