Service Composition Based Middleware Architecture for Mobile Grid

M.A. Maluk Mohamed and D. Janakiram

Distributed & Object Systems Lab,
Dept. of Computer Science & Engg.
Indian Institute of Technology Madras, Chennai, India

Abstract. Service Composition refers to the construction of complex services with the help of more primitive and easily executable services or components. With the proliferation of wireless communication and the mobile Internet, the demand for mobile data services has increased. In addition the recent spurt of e-services and m-services has increased the importance of service composition. We envisage service composition to play a crucial role in providing mobile devices access to complex services. The basis for our proposed approach is to virtualize the individual system resources as services that can be described, discovered and dynamically configured at runtime to execute an application. The paper discuss a novel paradigm which integrates mobile devices into the traditional grid, with the capability for composition of services. To accomplish such a global service composition in the new model we propose to add functional layers over the already proposed Anonymous Remote Mobile Cluster Computing model. The idea behind such middleware is to use the available resources efficiently and to hide the complexity inherent in managing heterogeneous services. This paper describes the unique capabilities of the proposed middleware and gives the layered view of the proposed architecture.

1 Introduction

The recent wave of innovation in computing, wireless communications, networking and electronics has embedded processing power, storage space and communication capabilities in electronic devices of day-to-day use, leading to the era of ubiquitous computing [1]. This trend has led to the pervasiveness [2], invisibility [3] and mobility of computing nodes. This has seen the availability of services also on the mobile devices in addition to the services available in the fixed network. This lead to the idea of integrating mobile devices into the grid, forming a new paradigm known as Mobile Grid.

The recent past has seen very fast growth in processing and networking capacity around the world. This has increased the need for software that can operate across multi-organizational resources at a higher level of abstraction than network protocols and independent interactive applications. Grid Computing [4] has emerged as a technology for coordinated large-scale resource sharing and problem solving among many autonomous groups. In Grids resource model, the resource sharing relationships are dynamic, attendants can join or leave multiple relationships at any time. However, Grid

A. Maluk Mohamed M. and Janakiram D. (2005).
Service Composition Based Middleware Architecture for Mobile Grid.
In *Proceedings of the 2nd International Workshop on Ubiquitous Computing*, pages 39-46
DOI: 10.5220/0002571400390046

requires a stable quality of service and the changing of sharing relationship can never happen frequently. There are many technical concerns that are addressed to produce effective grid computing. This may include information services, security, scheduling and resource management, performance, remote data access and archiving, and programming model. To this end, the model works for a conventional distributed environment but is challenged in the highly variational wireless mobile environment.

With the growing demand for mobile data services, novel value-added services and content provisioning will be the driving force behind the development and deployment of future communication networks and mobile Internet. As the Internet evolves and expands to include next-generation, data-enabled mobile devices, it is important to make data services available on these devices rapidly. It is essential to enable quick and flexible development and deployment of end application functionality. Such functionality should be enabled in the presence of a wide variety of heterogeneous access networks. Service mobility is the feature where the user can access the same set of services seamlessly, independent of the access network and the access device. To achieve this, it is necessary to personalize and adapt content to end-user devices with varying capabilities. Thus we consider the idea of composition.

Service Composition refers to the construction of complex services from primitive ones, thus providing rapid and flexible creation of new services. Fixed broker-based or fixed central-entity based service composition techniques cannot be applied as a solution to compose services on the fly as the clients can be mobile. The client may also use a service which may be mobile and connected through wireless network. This is due to the fact that such approaches presuppose the existence of a persistent central entity or coordinator in the clients neighbourhood, which is reliable and constantly available on request. Thus service composition can be defined as a dynamic integration of multiple services available in an mobile grid environment in response to a request from a client. The proposed research intends to study the key issues involved in building a middleware for service composition in a dynamic distributed mobile computing environment, and comes out with a layered architecture for the infrastructure of such a model.

The rest of the paper is organized as follows: section 2 gives a overview of the related work. Section 3 gives the implication of integrating mobile devices into the grid. Section 4 gives an overview of the proposed approach, that includes the motivation, design issues and principles. Section 5 describes the layered architecture of the service composition middleware architecture for mobile grid and in section 6 we conclude.

2 Related Work

In mobile environment, networks are unstable, mobile nodes can join and leave frequently. The quality of connection is unpredictable, even network itself can be ad-hoc. Moreover, mobile nodes have limited local resources and battery life. Although grid computing hasn't addressed these problems extensively, research in mobile computing provides solutions attacking these mobility issues. There are few infrastructures like Legion Grid computing infrastructure [5] that has come up to support resource heterogeneity ranging from supercomputers to small, handheld devices. There are many issues regarding mobile and wireless computing that are only solved in an application-specific manner and not by a general-purpose computing infrastructure. Nomadic nature of mobile devices which includes moving within and between environments poses some unique requirements. The foremost among them is the ability to locate available services and resources in or near a new location. In the vision of interconnecting diverse resources under a common framework, many research challenges arise, such as uniform ubiquitous connectivity, heterogeneity of sources, effective and precise resource discovery [6].Thus, we have investigated the relevant technologies to support building middleware for grid system that incorporates wireless and mobile devices.

To the best of our knowledge very few works such as [7] [8] [9] are done in mobile grids. But none of these papers address the issues due to mobility, which leads to availability problem. They had also specifically looked only at the computing nature of the mobile grid. [7] discusses about the challenges that arises due to integrating mobile devices into a grid. [8] provides an proxy-based approach for sharing of computing power among the participating mobile devices. The model considers only the mobile adhoc networks. [9] comes out with an architecture using mobile agent (MA) technology, where the mobile agents are used as a communication primitive. The paper discusses with the focus on mobile devices using the grid and not as a member of the grid.

3 Implication of Mobile Devices on Grid

The characteristics of the mobile devices such as the resource poor nature and low bandwidth connectivity (wireless) must be addressed if these devices are to be integrated as computational resources in a grid. The system must be capable to operate on constrained hardware. The sporadic and dynamic network environment must be handled gracefully. The grid should be flexible and reflective, to allow users to make trade offs and select the combination of services that is best suited for their purpose. This has led to the requirement of dynamic query and adjust, at runtime. The system should also be capable to support computation and service migration, replication, run-time monitoring, and application recovery. The disconnection of a given node in the grid should not disrupt service availability. In addition, the system should be able to offload computations if the device is likely to become unavailable soon.

4 Proposed Approach

Our aim is to integrate heterogeneous mobile wireless devices into the grid to form a new paradigm known as Mobile Grid. In addition we also aim to change the logic of adopting grid for single specialized high-demanding applications, and make it more general to allow many simple users to transparently access the distributed computing and storage resources, anyhow, anywhere and anytime. Thus increasing the user base of the grid. The mobile grid is visualized as cluster of clusters.

4.1 Motivation

There are many potential applications for our proposed mobile grid. Of particular interest is the earth science application. In earth science the geologist generally do fieldwork and mapping of some kind to study structure and dynamics of the earth. The work that is done so far by these field workers is manual, where they collect the data and return back to their office and do the collaboration at a central office. With the help of mobile grid, if the geologists are provided with some kind of mobile device, then collaboration with the globally distributed geologist becomes feasible. Thus the mobile grid helps in providing a integrated ubiquitous tool and environment to enhance collaboration among distantly located knowledge workers such as scientists and engineers. By collaborating they can work on solving problems assisted by the advanced collaboration and computation. This also helps in providing real time data gathering for collaborative field geology. This mechanism also helps to go out in the field as a group, collect data and share the data instantaneously with each other and with data fusion collaboratives hundreds of miles away. Can also have on demand access to large database, data summarization, and visualization resources at these centers to request site specific data. Provides group intelligence with the ability to monitor the data and the location of the data gatherers instantaneously. Gathered data can be fused immediately so that anomalies or missing information can be investigated while the data gatherers are still within the physical vicinity.

To achieve the above mentioned idea we attempt to identify key design issues, principles and layers of abstraction. Architecturally, the central idea of our approach is to virtualize the individual system resources of the mobile grid, such as data, computational cycles, wired and wireless bandwidth, or storage, as services that can be described, discovered and dynamically configured at runtime to execute an application. The system adds the available resources to the networked pool from which the resources needed to complete a given task can be aggregated dynamically. We propose to build a functional layer around the Anonymous Remote Mobile Cluster Computing kernel [10] to perform service composition for application execution.

An aggregated collection of resources operates as an assembly only for the period required to complete the collective task, after which the resources return to the network pool. We assume that nodes willing to share a certain subset of their resources use mechanisms to announce their availability, possibly stating the term of usage, to the rest of the mobile grid in which they are willing to collaborate. Individual nodes may be motivated to share resources for several reasons, such as to provide or gain access to unique data or to trade temporarily underutilized resources for profit or for the benefit of being able to draw upon the collective resources to handle their own peak loads.

4.2 Design Issues

We have identified the following key design issues for building the service composition middleware for mobile grid

- Handling hardware and software resources as a pool of services.
- Service Discovery - Optimal service discovery architecture must be able to find out all services conforming to a particular functionality, irrespective of its way of invocation.
- Service Coordination and Management - Must have some entity coordinating the different services involved in the composition.

– Verifying the performance of the components in composition.
– Uniform Information Exchange mechanisms to realize a heterogeneous service composition platform.
– Adaptability of the composed service to network and user dynamics.
– Compositions across independent service provider.
– Dealing with heterogeneity in protocols and data formats.
– Handling dynamic gluing of services at runtime, with capabilities to handle load balancing.
– Must be able to control and manage an aggregated set of services for completing a task.
– Fault Tolerance and Scalability

4.3 Design Principles

The two key design principles that have been identified for the design of the middleware architecture are: embedding intelligence in the network and create self-configuring, self-organizing network structures. It is seen from previous research experience that distributing intelligence tends to improve scalability. The self-organizing and self-configuring aspects contributes to both scalability and resiliency by creating ad-hoc networks of currently available nodes and resources. Improved scalability and performance result largely from the preference for use of local resources. Using local resources tends to shorten communication distances between the data source, its point of processing, and optional presentation to the user. This helps in reduced latency and a bias toward consuming edge bandwidth rather than the backbone bandwidth needed to communicate with remote centralized servers. Distributing intelligence into the network also tends to distribute the load across many processing points, as opposed to creating congestion at a few heavily used points.

5 Proposed Architecture

5.1 Layered Architecture

Figure 1 shows the key network services and abstraction layers of the proposed service composition middleware for the mobile grid. The central idea of the architecture is taken from [11], which debates on using the internet as distributed computing platform. The service composition middleware for mobile grid is proposed keeping the key issues identified and the design goals in mind.

The lower most layer represents the communication infrastructure, namely both wired and wireless network. This layer represents the underlying communication infrastructure being used to communicate between different services in our system. It allows the use of different protocols for transferring messages. Thus this layer can encapsulate any communication mechanism and allows seamless integration of different network protocols in our system.

The Resource Virtualization layer represents the mechanisms through which services express themselves and concentrates primarily on the use of resource sharing in

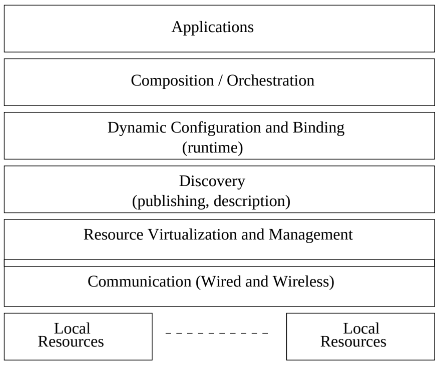

Fig. 1. Layered Reference Architecture for Service Composition in Mobile Grid

the grid. Resource Virtualization is thought of as an abstraction of some defined functionality and its public exposure as a service through an interface that applications and resource managers can remotely invoke. We consider the service to be a virtualized software functional component. Services can be advertised and discovered using directories and inspection. Once discovered, an invoking entity can bind to the selected service and start communicating with its externally visible functions, through platform independent protocols. Each such virtualized component can be abstracted, discovered and bound to. In the same way this can be extended to virtualize hardware resources. Thus a planetary-sized pool of composable hardware and software resources can be envisioned.

Discovery is the fundamental component, as the system must find a service before it could use it. Traditional systems often implicitly discover or fix it at configuration or compile time. In case of mobile grids which has to accommodate both mobile and static devices with dynamic aggregated environments, it should be able to support a flexible service advertisement and discovery mechanism. Applications can discover services based on their functionality, characteristics, cost or location. Dynamic discovery enables devices to adaptively cooperate and extend functionality.

Dynamic configuration depends upon the capability to bind components at runtime, in contrast to the binding at design time. Runtime binding can be implemented with the assistance of service discovery mechanisms. This helps in decoupling application design from the detailed awareness of the underlying system configuration and connectivity. Dynamic configuration facilitates application portability across a wide range of platforms and networks configurations. Runtime binding also enables load balancing and improved reliability.

Resource Orchestration deals with control and management of the aggregated set of resources for completing a task. It also includes the communication and synchronization necessary for coordination and collation of partial results. On completion of a task, the system releases the resources back to the pool for allocation to other users.

5.2 Service Coordination and Management

The proposed mobile grid is organized as a cluster of clusters as illustrated in figure 2. Each cluster maintains a cluster head, which maintains the required repositories associated with the cluster. The cluster head using trading services maintain the services available within the cluster. The cluster head communicates with each other in a peer-to-peer fashion in contrast to the usual hierarchical organization. The mobile devices using the exporter daemon sends the available services to the MSS, which in turn passes it to the cluster head.

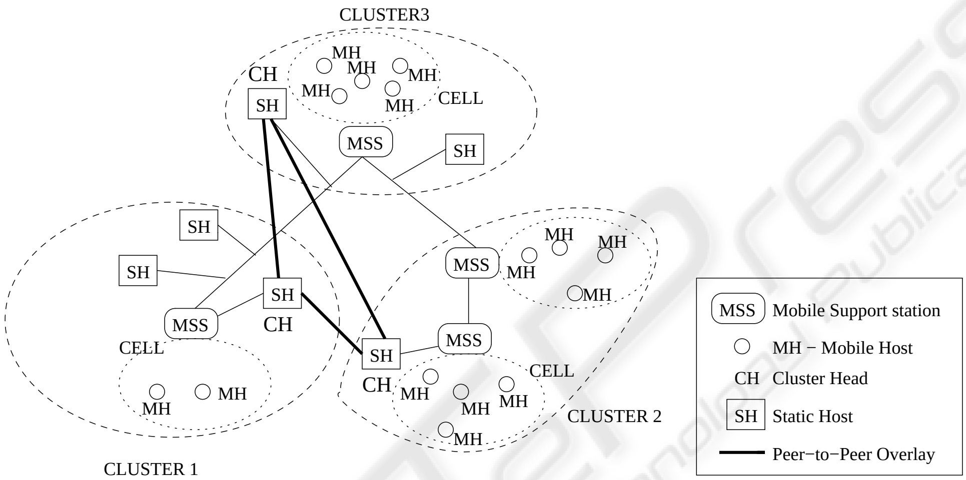

Fig. 2. Mobile Grid Architecture Clustered as p2p overlay

When client request the cluster head for a service, the cluster head checks for the availability of the service within the cluster. When the cluster head finds that services is not available within the cluster, the cluster head starts to search other neighbouring cluster through the peer-to-peer overlay. In order to optimize on the searching the cluster head stores the historical data related to the earlier requests. Thus the cluster head can take an intelligent decision based on the past request. Each cluster head maintains such information about each of its neighbour cluster heads. When a request for a composite service is initiated from a mobile device, the middleware should provide the service which is available in the near vicinity of the mobile device. The success of the composition depends on the number of hops that are taken to reach the desired service from the cluster head.

6 Conclusions

Service Composition architectures will become very important with the increasing growth of e-services. The proliferation of mobile devices during these years have started seeing the availability of services on different mobile devices. However the conventional approach is that mobile devices will be merely web browsers without significant computational capabilities unto themselves. This paper is proposed with an envision that mobile devices (embedded devices) can be capable to deliver much greater capabilities for grid computing. This has seen to the integration of mobile devices into the grid to form a new paradigm namely the mobile grid. Thus this paper proposes a new service composition middleware architecture for mobile grid. The paper also provides the design issues and principles that have been considered to build the layered architecture.

We are currently working on implementing the proposed model and experimentally validate the model. We are also looking at building a new model that could to some extent provide solution for the availability of the mobile device, which is due to the mobility.

References

1. Weiser, M.: Ubiquitous Computing, IEEE Computer - Hot Topics (1993).
2. Satyanarayanan, M.: Pervasive Computing: Vision and Challenges, IEEE Personal Communications (2001).
3. Norman, D.: The Invisible Computer: Why Good Products Can Fail, the Personal Computer is so Complex, and Information Appliances are the solution, MIT Press (1998).
4. Foster, I. and Kesselman, C. (eds.): The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann (1999).
5. Grimshaw, A.S., Wulf, W., and the Legion team: The Legion Vision of a Worldwide Virtual Computer. Communications of the ACM **40(1)** (1997) 39-45.
6. Bernstein, A., Klein, M.: Towards High-Precision Service Retrieval. Proceedings of the International Semantic Web Conference (ISWC 2002), Italy (2002).
7. Thomas Phan, Lloyd Huang and Chris Dulan: Challenge: Integrating Mobile Wireless Devices Into the Computational Grid, Proceedings of the eighth Annual International Conference on Mobile Computing and Networking, ACM SIGMOBILE (2002).
8. Junseok Hwang and Praveen Aravamudham: Middleware Services for P2P Computing in Wireless Grid Network, IEEE Internet Computing (2004) 40–46.
9. Dario Bruneo, Marco Scarpa, Angelo Zaia and Antonio Puliafito: Communication Paradigms for Mobile Grid Users, Proceedings of the third IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID'03), IEEE Computer Society (2003).
10. Maluk Mohamed, M.A., Vijay Srinivas, A., Janakiram. D.: Moset: An Anonymous Remote Mobile Cluster Computing Paradigm, Accepted to appear in the special issue of Journal of Parallel and Distributed Computing (JPDC) on Design and Performance of Networks for Super-, Cluster-, and Grid-Computing, 2005.
11. Milan Milenkovic, Scott H. Robinson, Knauerhase, C., David Barkai, Sharad Garg, Vijay Tewari, Todd A. Anderson, Mic Bowman: Toward Internet Distributed Computing IEEE Computers (2003) 38–46.