Leveraging the Grid to Provide a Global Platform for Ubiquitous Computing Research

Abstract. The requirement for distributed systems support for Ubicomp has led to the development of numerous platforms, each addressing a subset of the overall requirements of ubiquitous systems. In contrast, many other scientific disciplines have embraced the vision of a global distributed computing platform, i.e. the Grid. We believe that the Grid has the potential to evolve into an ideal platform for building ubiquitous computing applications. In this paper we explore in detail the areas of synergy between Grid computing and ubiquitous computing and highlight a series of research challenges in this space.

1 Introduction

When Marc Weiser inspired researchers all over the world by sharing his vision of ubiquitous computing (ubicomp) [24], he expected a gradual movement towards the concepts of embodied virtuality within 20 years, leading to ubiquitous computing emerging as the dominant form of computer interaction. However, despite numerous lab-based prototypes, ubiquitous computing technologies have failed to become commonplace. Many reasons are cited for this slow progress [2,12], including the lack of a common distributed systems infrastructure to facilitate application and service development and deployment. This lack of infrastructure support has already lead to the development of a number of specialised middleware platforms, e.g. GAIA [18], Aura [19] and Cooltown [11]. However, such platforms have failed to become widely adopted as a common standard. This may be due, at least in part, to the fact that these platforms are proprietary in nature and it is notoriously difficult to convince researchers to base their work on third-party “home-brew” platforms that are not widely used and whose future may be uncertain. Even where these platforms have attempted to leverage existing technologies (e.g. Cooltown’s extensive use of Web protocols and services) they have not experienced significant uptake.

In contrast to the development of multiple platforms that tackle subsets of the challenges associated with ubiquitous computing, within the context of many other scientific disciplines we are currently witnessing the evolution of the vision of a global distributed computing platform, i.e. the Grid [5]. The Grid promises a world where access to (computational) resources across institutional boundaries is standardised, uniform, inexpensive, ubiquitous and reliable (see
section 2. While the origins of this vision might not exactly match that of ubiquitous computing, many common elements can be identified. We believe that a broad adoption of the Grid as a starting point for ubiquitous applications offers benefits to both the Grid and ubiquitous computing communities. We have already gained a significant amount of experience by using Grid technology for the realisation of a small number of applications within the ubiquitous computing domain. These applications, described in section 3, show how Grid concepts can be easily applied to this new target domain. Although we believe that the current instantiations of the Grid do not yet provide us with an off-the-shelf solution for ubiquitous computing, we argue in section 4 that the use of Grid technologies in ubiquitous computing will benefit both communities. Having presented the argument for applying Grid technologies to the domain of ubiquitous computing and highlighted the key research challenges that remain to be addressed, we conclude with a discussion of our plans for future work including the creation of a portal to support wide-scale cross-organisation experiments in ubiquitous computing.

2 The Grid Vision

At its inception, the Grid referred to the vision of a global networked computational resource for the execution of demanding jobs and experiments, comprised of numerous distributed processing and storage resources. Such a platform has great potential impact for many disciplines (e.g. high energy physics) and this vision has been widely adopted. It is anticipated that the Grid will be formed from a heterogeneous set of machines, ranging from simple workstations to high-performance vector parallel computers. The integration of these different kinds of nodes will be achieved by providing a standardised and uniform way to access computational resources. The overall aim is that the Grid will make processing power and data storage ubiquitous, analogous to the electricity distribution grid. Ian Foster, who is given most of the credit for coining the term Grid, describes a computational Grid in [5] as

“a hardware and software infrastructure that provides dependable, consistent, pervasive and inexpensive access to high-end computational capabilities.”

Over time, however, the Grid has moved away from the mere sharing of computational resources and has become a more generic platform for the sharing of any kind of networked resource. Foster therefore recently described the Grid as a system that

“coordinates resources that are not subject to centralised control . . . using standard, open, general-purpose protocols and interfaces . . . to deliver nontrivial qualities of service.” [4]

If we compare Foster’s vision of the Grid with Weiser’s articulation of ubiquitous computing, we find many similarities. In ubiquitous computing we speak of augmenting the environment with large numbers of devices and services. It becomes
clear that most of these devices will not be under centralised control. Moreover, in order to create valuable services to the human user, these services and devices will have to interact in a more or less coordinated fashion and these devices will be heterogeneous in nature, making it necessary to specify standardised ways for different devices to interact with each other. If we follow this path and redefine the notion of a resource to include services and devices, we can clearly see that the basic principles of ubiquitous computing and Grid computing are not very far apart. To further illustrate these similarities, we will use the term “resource” as a unifying notion, including both services and devices.

However, inter-organisational sharing of resources is not a task that can be achieved easily. While the vision of the Grid is widely accepted and maturing rapidly, a number of key issues remain to be addressed. Comparing these issues against the basic requirements of ubiquitous computing environments (as recently highlighted in [2,12]), reveals substantial synergy between the two paradigms. Below we present a brief analysis of the key challenges affecting infrastructures in both domains:

- **Heterogeneity and Interoperability**: Both ubiquitous computing and the Grid involve a large number of resources that are heterogeneous in their nature. Standardised mechanisms for inter-resource communication have to be defined. These mechanisms have to be extensible for the seamless inclusion of future resources.

- **Scalability**: Grid applications typically consist of highly parallel or massively replicated computations spanning 1,000s of computational elements. In ubiquitous computing, the environments themselves are comprised of innumerable devices, sensors and computational services. Both environments offer similar challenges of scalability, as the number of resources/services increases, adapts and evolves to encompass new nodes and users.

- **Adaptability and Fault Tolerance**: The Grid and ubiquitous computing environments are both highly complex distributed environments. As such, both kinds of environment will be required to cope with change, failure and the introduction of new components – gracefully and at run-time – it is not practical to reboot such systems to fix faults or return to a particular known-good state! While the semantics of failure are not well defined in ubiquitous computing, it is clear that platforms must offer the ability to adapt to changes in their underlying environment and offer dependable failure modes.

- **Resource Management and Service Composition**: Clearly related to scalability are issues of resource management and higher level service composition. As Grid services become commonplace and the range of services on offer diversifies, meta-level services and tools will be required to help automate and control the life-cycle, interaction, monitoring and dynamic composition of simple services to form higher level applications. Likewise, the ability to cope with the diversity of devices in emerging ubiquitous computing environments, and the difficulties in writing portable “applications”, that run across more than one such environment, present similar challenges for the supporting infrastructure.
– **Service Discovery**: In both application domains, effective and efficient ways of discovering services/resources need to be developed. This is also an essential requirement for the efficient deployment of resource management and service composition mechanisms.

– **Security**: As in most of today’s computing domains, the Grid and ubiquitous computing are both in need of a consistent security architecture that is sufficiently lightweight and scalable. There are however, many other non-trivial issues to address in this arena: Users must be authenticated and authorised to access resources. To support large numbers of users and resources, without relying on a centralised and integrated security infrastructure, authentication and authorisation might have to move towards the concepts of recognition and trust. The notions of ensuring data integrity, confidentiality of data and privacy of information are other aspects that have yet to be addressed. Although each of these notions can to some extent be tackled similarly in both domains, protecting users from unwanted disclosures of private or confidential aspects of their interactions within a ubiquitous computing environment offers some unique, and possibly intractable challenges (e.g. due to the pervasive use of embedded sensing and devices with extremely low capabilities). As a further challenge, any new security techniques need to be capable of integrating into organisations’ existing security infrastructures, with due consideration to both the technical and social issues this implies [17].

– **Communication**: Services/resources may be interconnected using a variety of different communication media. The characteristics of these different media clearly impact on higher layer protocol issues. More specifically the current Grid protocols assume plentiful bandwidth and reliable communications – features not typically found in wireless networks. In addition, some applications and services might demand specific qualities of service (e.g. timely delivery of significant events, or of streaming media). Means for specifying these requirements have to be supplied, as well as underlying mechanisms for enforcing these requirements. Already a challenging problem in conventional local and wide area networks, it is likely that ubiquitous computing in particular will raise additional issues due to the variety of device and the many paths of interconnection possible. In addition to dealing with heterogeneous communications media both the Grid and ubiquitous computing platforms must provide a range of communications paradigms to applications and services. In particular, there is a clear requirement for both synchronous and asynchronous (event based) communications support.

– **Audit Trails**: Applications in both ubiquitous computing and the Grid will involve interactions with large numbers of heterogenous resources. It is therefore vital to provide means for making internal system processes and the decisions of individual components visible to the user (e.g. for reassurance or simply to support application development) – making technology “calm” will inevitably involve the delegation of decisions away from the user into “the ubicomp environment” [25]. However, from time to time, users might still be interested in the steps and decisions that have lead to the initiation of a particular action, e.g. in the case of a “misbehaving” smart room. In the
Grid context, researchers might be similarly interested in the progress of a large-scale computation.

- **Payment**: In both computing domains, the deployment of infrastructure is an undertaking that can be very costly. Means have to found for financing these infrastructures, for example by directly imposing charges for accessing and using services and resources or via cross-subsidisations from other sectors [14].

As we can see, the set of requirements that have to be met for building (computational) Grids significantly overlaps with the requirements for supporting ubiquitous computing applications. We argue that future embodiments of the Grid have the potential to form “the” ubiquitous computing platform of the future. In this case, the issues that we have identified have to be dealt with in forms appropriate for ubiquitous computing applications. For example, one key assumption in traditional Grid computing, that no longer holds when applied to ubiquitous computing, is that of ‘always on’ high-bandwidth connectivity between all the participating nodes. As a result, a temporary phase of disconnection or weak connectivity between resources will be seen and handled as a severe fault.

However, as previously discussed, the Grid seems to be moving away from the classical view of a super-computing resource to embrace a wider range of emerging scientific disciplines (e.g. pervasive sensing, medical informatics [15]). Such domains will witness a more heterogeneous composition of devices and “services”, including mobile and sensing devices as first-class Grid entities. Clearly, these forms of applications could be viewed as harmonious with the objectives of ubiquitous computing.

This new vision still has to gain momentum within the Grid community. Recent documents, such as [76] still focus on computational resources and their possible exploitation within a global Grid infrastructure. To us this is a clear indication that so far there has been very little or no penetration of expertise from the domain of ubiquitous computing into the area of Grid computing. If we generally believe that this is true, the logical consequence is that there is little sensitivity within the Grid research community to problems that are specific and crucial to ubiquitous computing, such as the limitations of light-weight, possibly embedded devices. To the best of our knowledge there seems to be little consideration within the Grid community regarding issues that might occur when dealing with this class of resources, e.g. mobility, temporary disconnections and low quality links.

### 3 Using Grid Technologies

So far we have only been talking about the vision of the Grid, its similarity to the vision of ubiquitous computing and the common requirements that both approaches impose on their infrastructure platforms. But how far has the vision of the Grid already become reality? To what extent are existing Grid middleware platforms usable for building pervasive computing environments?
Although there already exists a variety of commercial and non-commercial Grid middleware platforms, e.g. Globus v2 \cite{9}, Legion \cite{10} and Avaki \cite{1}, currently one can identify a strong movement throughout both sectors towards one single architecture, the Open Grid Services Architecture (OGSA) \cite{6}. OGSA is a joint effort to agree on a standardised next-generation platform for future Grids, mainly driven by members of the Global Grid Forum (GGF) \cite{8}. With the adoption of OGSA as a common basis, the global Grid community has accepted and realised one of the main visions of Grid computing: to agree upon open standards that provide the means for interoperability between resources, in this case based on the concept of Web services \cite{23}. Although still undergoing a very dynamic process of standardisation, OGSA already provides some basic functionality for building computational Grids, including introspection, registration, eventing, lifecycle management, service creation and naming. Standard interfaces for these services and their behaviour are in the process of being defined as part of the Open Grid Services Infrastructure (OGSI) \cite{22}. More generally, OGSI is commonly seen as foundation, providing necessary infrastructure entities for building OGSA.

Being work in progress, OGSI and, even more, OGSA are far from complete. Nevertheless we have begun to explore whether these platforms offer enough functionality by building small prototype ubiquitous computing applications. In the remainder of this section we outline two such applications and discuss how these are realised using existing Grid technologies.

Our first scenario illustrates how we can build a classic ubiquitous computing vision – a context-aware room – using Grid components.
John is a researcher at the University of XXX who is currently visiting YYY University. For his coffee breaks he likes to visit the Foo-Bar Lab (FBL), a smart environment that includes facilities for room customisation. John uses an iris scanner to authenticate himself to the room and to gain access. After John has been recognised, the iris scanner sends a notification to a room customisation service that has been assigned to the FBL. As with all services associated with the FBL, the customisation service is a Grid service, i.e. a Web service that exposes as set of interfaces common to all Grid services. Upon receiving the notification, the room customisation service contacts a registry and retrieves information about John’s preferences. The preferences are analysed, to deduce the resource requirements of each customisation action. According to the current state of the services that are associated with the room and their availability, the customisation service picks a set of actions that can be performed. The service decides that it is possible to bring up a summary of the local news on one of the displays in the lab. It contacts a resource management service, requesting exclusive access to the display service for display 2. The service then contacts a factory service, initiating the instantiation of a display service and delegates the exclusive access rights for display 2 to the newly created service instance. The display service can now be used for displaying a personalised version of the local news for John.

Following our commitment to exploring the viability of Grid approaches in ubiquitous computing, a demonstrator showing the feasibility of this scenario has been built in early 2003 (see figure 1). Participating components have been engineered as OGSI-compliant Grid services, providing us with a standardised and cross-platform framework for communication between services. This scenario also demonstrates how the lifecycle management mechanisms of OGSA can be used for helping us to create a fault-tolerant ubiquitous computing environment. Instantiation and termination of a display service instance in our architecture is usually handled by the room customisation service. In order to keep these service instances alive, keep-alive messages have to be issued to these services at regular intervals. However, if a display service hasn’t received a keep-alive message for a specific interval, it will be automatically terminated. By enforcing this behaviour, we are able to make sure that forgotten services will terminated by the framework, releasing resources back to the environment.

In our second scenario we consider how the fusion between the ubiquitous computing and Grid domains can enable new forms of application, with potentially far reaching consequences.

Since his birth John has been suffering from chronic lung disease, which requires that he undergoes lifelong oxygen treatment. To help him cope with his condition, John has been given a small wearable computer that records his vital signs, such as his blood pressure, his pulse and, most importantly, the oxygen saturation in his blood. John’s wearable is
equipped with wireless communication facilities which are used for periodically transmitting the recorded data to a small number of software components. One of these services is responsible for storing the data in a large distributed database. The database can be accessed by clinicians to track developments in John’s condition. The database also contains information about the state of John’s surroundings, e.g., temperature and the quality of the air. This type of data gets automatically collected from all environmental sensors that are in John’s immediate vicinity. A second software component processes John’s data as it enters the database, looking for anomalies in John’s vital signs that need immediate action. Once an anomaly has been detected, an expert system assesses the severity of the event. In case of an emergency, John’s practitioner is automatically contacted and the system dispatches an emergency response team to John’s current location. In case of minor events that can be handled by John himself, a notification is sent to his wearable computer which will then inform its wearer.

This scenario is far from being science-fiction. We have built a proof-of-concept demonstrator that uses a prototype sensor jacket developed by researchers in medical sensing. The jacket can be worn by children suffering from chronic medical disorders. This work was conducted as part of a joint project with medical researchers who explicitly required interoperability with emerging medical Grid services. Our prototype uses an alpha version of Globus v3 (GT3), a Grid toolkit that aims to provide the first OGSI-compliant implementation. Most of the components of our demonstrator have been engineered as Grid services, providing standardised means for introspection and interaction. Our implementation relies heavily on OGSI’s eventing framework. Future evolutions of the demonstrator will be extended to use OGSI’s registry services for service discovery and the database access components of the OGSA Database Access and Integration framework OGSA-DAI [21], that is currently being designed by major academic institutions. OGSA-DAI will provide standardised access to existing heterogeneous database platforms, making it possible to store large amounts of data within the Grid’s numerous and widely distributed databases. If we try to envision the capabilities of future OGSA developments, one can also imagine making use of distributed computing facilities to process and visualise the data that has been collected from a possibly large number of people, e.g., for finding correlations, predicting epidemics and detecting trends in the data gathered from patients who are suffering from the same disease. One can also clearly envision how these repositories and sensor networks, which are to be deployed on a scale not possible in laboratory testbeds, can be re-purposed to support new experiments and, significantly, new ubiquitous computing applications.

In our prototypes, we have found GT3 to be a viable platform for rapidly prototyping some of our ideas in the context of future reactive environments and medical monitoring applications. The use of GT3 has enabled the construction of a testbed and a platform for our future research, without the effort of building our own middleware platform.
However, by the same token, we have have been forced to acknowledge that the functionality provided by GT3, OGSI and OGSA is far from sufficient when seen in the context of ubiquitous computing. The existing service discovery mechanisms are a good example: although OGSI comprises and defines a registry interface, it is unclear how clients will be able to discover registries themselves. This is clearly a bootstrapping issue that can probably be solved very easily, e.g. by using multicast announcements for advertising a registry’s presence. However, without such a bootstrapping mechanism, devices that have been freshly “plugged” into a Grid environment, won’t be able to discover other services, making the Grid inaccessible for them. This is a more significant issue if we are dealing with mobile or portable devices, instead of immobile desktop workstations that can be statically configured at installation time.

4 Discussion

As we have tried to show throughout the previous sections, the Grid can not yet be seen as the perfect ubiquitous computing platform. But we are convinced that the time has come for the ubiquitous computing community to become actively involved in the process of designing future Grid standards. The Grid seems to have reached an important turning point. Until now, most of the deployments of Grid middleware mainly dealt with computational tasks. However, many of the requirements that have been mentioned in section 2 have not been sufficiently addressed so far, making the Grid open to new solutions and influences. With the arrival of OGSA we have seen the Grid move towards a more service-oriented environment, already making it a more feasible platform for ubiquitous computing. Addressing and finally solving the Grid’s many open issues can be done in two different ways: with or without having the specific needs of ubiquitous computing in mind. The second option will leave us with a possibly global and widely deployed Grid infrastructure that can not be used for realising the visions of ubiquitous computing. The shortcomings of this approach are clearly illustrated by our medical scenario: without integration we will see duplication of many services and applications. Only by choosing the first option is it possible to form a global Grid that also suits the visions ubiquitous computing.

Inspecting OGSA’s roadmap [20] it is apparent that many of the issues discussed in section 2 will shortly be addressed by Grid researchers as part of their ongoing development process. The key challenge for the ubiquitous computing community is to ensure that the solutions developed to meet these challenges are appropriate for both the traditional Grid computing domain and ubiquitous computing.

In addition, participating in the process of defining the future of the Grid presents unique opportunities to the ubiquitous computing community. One of these opportunities is to leverage the existing Grid installations. A large amount of money is currently being spent for supporting research on Grid computing and to further promote the deployment of Grid middleware. The Grid, respectively
Grid middleware, is already widely deployed and this is expected to further increase as the demand for computational resources grows.

The Grid also offers a much wider scope for the deployment of ubiquitous applications. Using a Grid approach, applications could be tested using real world components rather than simple lab-based tests under artificial conditions. Applications could be dropped into the Grid, requiring a new sensitivity regarding their behaviour in respect to things like scalability and security within this global environment. Moreover, using the Grid, formerly separate testbeds can be linked together. Currently, research in the field of ubiquitous computing has often been carried out in a very isolated fashion. For example, researchers working in the areas of security, service discovery and HCI have all created their own test-environments to show the feasibility of their solutions. However, as research efforts are often very focused, the interchange of results between these different areas of research seems to be very limited. For example, new service discovery approaches seldom pick up recent developments in the area of security. Among other reasons, the limited re-useability and accessibility of developed services and devices play an important role. Solutions are often tied to a specific proprietary platform, leading to limited interoperability between existing solutions. Agreeing on a single standard platform could help us to overcome these issues and bring significant synergistic effects.

Closely related to the issues of interoperability, picking up a common platform for research interests in the area of ubiquitous computing might help many ubiquitous computing applications to escape their current testbed existence and become widely deployed in the “real world”. So far, it is arguable that the deployability has suffered from two major problems. Firstly, the current computing infrastructure hasn’t been able to provide an environment for these applications and their needs. Secondly, these applications have often been tailored to a specific ubiquitous computing environment, making a deployment outside of their lab environment impossible. A widely-spread introduction of a ubicomp-friendly Grid infrastructure could help to cross these hurdles and make more and more technology transfer from our labs into the “real world” possible, bringing us one step closer towards the pervasive computing world we are all keen to see one day.

In order to make this first step towards the goal of having a globally accepted, open platform for supporting future deployments in the domain of ubiquitous computing, we have to start using Grid technology for realising research ideas in our area. It is our strong belief that by doing so, more contributions to the Grid community will be made, turning the Grid into a viable platform for our research efforts. Contributions from the ubicomp community will hopefully enhance the usability of future Grid embodiments for many different purposes, helping it to evolve into an even more accepted and widely spread middleware platform.

5 Creating the Ubi-Grid

Following our commitment to the promotion of Grid technologies in ubiquitous computing, we are currently in the process of creating tools to facilitate the
The emergence of a global Grid dedicated to ubiquitous computing, the Ubi-Grid. As we have pointed out throughout this paper, existing Grid approaches still provide very limited support for building ubiquitous systems. For example, many of the issues highlighted in section 2 are not addressed by current instantiations of the Grid. Therefore we believe that it is vital to provide the research community with an open forum that can serve as platform for discussing and addressing these issues.

One of the biggest drawbacks that we have encountered so far is the lack of support in Grid technologies for dynamic service discovery, as we have pointed out in section 3. We have therefore started to create a Web-based service directory that can be used by researchers to advertise existing ubiquitous computing applications that are already available as Grid services. At present our portal is simply a database of service URIs that are available to researchers. These URIs must be discovered and applications configured to use these services either manually or via the simple SOAP interface we provide. We anticipate that as research into Grid technologies progresses, this form of service discovery will be replaced by a standardised solution. Our Ubi-Grid portal also aims to be a valuable tool that can be used for discussing new ideas and solutions within the domain of using the Grid as a platform for ubiquitous computing, e.g. by hosting mailing-lists and discussion forums.

Inevitably, our current set of services is small and restricted to our local domain. In more detail, we offer services to support tracking of entry to a smart space, video recording of events within this space and display of information. These services could, for example, be used by researchers interested in creating ambient displays who do not themselves have access to a smart space. As a specific example, an experiment could be created that used our iris scanner to identify researchers entering the smart space, the displays in the lab to present information and the video recording service to monitor researchers’ reactions to the information displayed. Clearly, such a scenario has significant privacy implications and we are developing a number of solutions in this area [16].

In order to realise the benefits offered by the Grid, wide participation is required and we encourage researchers to visit the portal (URL removed for anonymous review) and add references to Grid services that they develop.

6 Conclusion

While significant progress has already been made in the area of ubiquitous computing, there still is no standardised, uniform platform for the wide-scale deployment of applications in this area. The development of such a platform could provide substantial benefits for both the research process and the applicability of its results. We have argued that the Grid could provide a viable platform for ubiquitous computing: the Grid and ubiquitous computing share many common requirements. Furthermore, we have pointed out that Grid research is currently at a turning point that will influence the usability of future Grid platforms for the purposes of ubiquitous computing. Although current Grid toolkits do not yet
allow us to build sophisticated applications, our first applications have demonstrated that it is already possible to use Grid technology for rapidly prototyping small ubiquitous computing applications. We will therefore continue to use and explore Grid technologies for building ubiquitous computing applications, e.g. as part of our medical project. In addition we have developed a portal to support other researchers interested in using the Grid to support ubiquitous computing research and enabling us to offer services to the community. By pursuing this avenue of research we hope to be able to make significant contributions to the Grid research community and bring benefits to developers of future ubiquitous computing applications.

References

14. Removed for review: Removed for review.
15. Removed for review: Removed for review.
16. Removed for review: Removed for review.