Towards the Realization of a Mobile Grid

Konstantinos Katsaros and George C. Polyzos
Mobile Multimedia Laboratory, Department of Computer Science
Athens University of Economics and Business, Athens 104 34, Greece
ntinos@aueb.gr, polyzos@aueb.gr

ABSTRACT

The introduction of mobile devices and wireless communications in the context of the Grid computing paradigm has recently drawn the attention of the research community. We point out fundamental issues and problems emerging from this introduction and further describe our ongoing work for the investigation of key design decisions and optimizations.

1. INTRODUCTION

A grid computing system is essentially a large-scale distributed system designed to aggregate resources from multiple sites. In this context, the availability of an enormous number of mobile computing devices such as laptop PCs and PDAs, has naturally led to the extension of the idea of resource sharing to mobile and wireless communication environments.

There are various, quite different approaches on the exact character of this extension. Their differences mostly rely on whether mobile devices should exploit the traditional fixed Grid in order to overcome their resource constraints and thereby act as *resource consumers* (e.g. [1]) or they could also act as *resource providers* (e.g. [2]) because of their increasing computational power and their ubiquity. Mobile devices indeed face resource limitations, at least in comparison to their stationary counterparts, but the vast number of available mobile devices and the fact that their computational power is constantly increasing, lead us to the assessment that the aggregated sum of their resources could be exploited in order to overcome exactly these limitations. In effect, resource-sharing collaboration between small or large scale communities of mobile users appears as a promising research direction towards the alleviation of the inherent resource constrains present in mobile computing environments.

We have proposed a hierarchical, campus-wide computational Mobile Grid system architecture and studied it as a case-study [3]. In the proposed architecture, mobile nodes (MNs), willing to offer their computational resources, move between APs of the campus. This willingness is based on the expectation of *reciprocity*. In this direction, we have devised an *incentives scheme* which allows each MN to collect *credits* for each task it processes and exchange them for the execution of a job of its own on the Mobile Grid [3].In our work we have considered *divisible load* applications in which a job can be divided into tasks that can be carried out independently of each other. For example, distributed query processing can be performed by the members of a mobile workforce on partitions of large data sets. All computing tasks are distributed by the system's schedulers to the collaborating MNs, which process them and return the results back to the scheduler. In the following, we describe our current research approach towards the realization of a Mobile Grid system.

2. SCHEDULING AND PERFORMANCE

In a wireless and mobile networking environment noise and intermittent connectivity impose delay and other overheads on the communication between the scheduler and the MNs. Therefore, the question is how to overcome this barrier in order to achieve high performance. In this context, we evaluate several fundamental scheduling approaches based on performance metrics such as the *Response Time (RT)* i.e. the time required for the resulting data to be available to the scheduler, the *Resource Waste (RW)* i.e. the amount of resources wasted in the effort to *hide* intermittent connectivity and *speedup* i.e. the comparison of the achieved *RT* with that of a single node execution, which reveals the actual degree of parallelism achieved. These policies refer to whether a task executed by a mobile node must be aborted upon disconnection, in the sense that it will be rescheduled, or it must be carried out by the mobile node until a new point of attachment to the network is found and the results can be returned back. On the one hand, the resources spent by a MN for the execution of a task are obviously wasted if the assigned task is aborted. On the other hand, it may be preferable, in terms of RT, to abort and reschedule a task if disconnection peri-

ods are long and MN arrivals are frequent. What is implied here, is a trade-off between efficient turn-around times and resource utilization. Hence, our study focuses on the investigation of the underlying network characteristics affecting this balance. In the same context, we further examine the effects of *task replication* and the notion of *installments* on the observed system performance. Based on the fact that not all MNs present the same networking behavior at the same time, we replicate tasks in order to conceal disconnection events by picking the earliest response among the replicas. As resource utilization is an issue here too, we study the impact of redundancy on the observed RT and the resulting RW. In the case of *installments*, we consider further sub-dividing the tasks into smaller, possibly successive sub-tasks with lower requierements in resources but higher fragmentation and re-assembly overhead. We are currently investigating the role of the installments size on the system performance.

2.1 Future work on scheduling

Our experimental work on scheduling policies has been based on the WLAN traces collected at Dartmouth College. However, even though these traces give a realistic view of the performance of each policy, they rather confine our results to a specific networking environment. For this reason, we intent to utilize various models, e.g. those developed by Papadopouli *et al.* [4] which describe several important mobility parameters such as session arrival, session length etc. Our target is to produce *synthetic traces* for various networking conditions (e.g. high/low mobility) so that our investigation reveals the suitability of each policy.

At a later stage, we also plan to extend our research to mobile ad-hoc networking environments (MANETs). In this context, our investigations will involve resource discovery, scheduling and load balancing under the more stringent mobile ad-hoc networking conditions in which the scheduler is also mobile. We shall focus on the discovery of sets of mobile nodes satisfying the resource requirements of the Mobile Grid user taking into account the underlying network conditions such as ad-hoc network density, mobility and failure rates, hops-to-providers etc. This process may vary subject to the resource requirements and the nature of the application. Issues such as load division and distribution, task inter-dependencies, heterogeneity in task and/or resource characteristics show the interplay between resource discovery and scheduling and call for the investigation and the development of proper resource discovery and scheduling protocols and algorithms. In this effort, the implicit resource provision by intermediate nodes in multi-hop networks must also be considered. To this end, possible *cross-layer* scheduling schemes will be investigated with respect to information available from the underlying resource discovery and routing protocols. The embedment of scheduling requirements into the resource discovery protocols (e.g. mapping of task inter-dependencies onto specific routing paths) is another research direction under consideration.

3. SECURITY ISSUES

In a distributed job execution environment, potential risks rise both for the integrity of the application and the resource provider i.e. the MN. First, the integrity of the processing results is compromised insomuch as a malicious MN attempts to deliberately alter the results possibly in a goal-oriented way. Second, the *free-riding*[1] problem may reduce the overall system utility. Furthermore, *data confidentiality* may also be a requirement. Promising research directions for the confrontation of these risks include the use of code and data *encryption* techniques [5] , the introduction of *redundancy* in the system and the employment of proper incentive schemes and *reputation mechanisms* (e.g. [6]).Moreover, several serious security considerations emerge in case mobile code is dispatched to resource providers. To name few of the risks, malicious mobile code may destroy files and applications, divulge confidential information, perform actions on behalf of the device owner. Countermeasures under investigation include the *authentication* of the mobile code dispatcher e.g. with X.509 certificates, *code verification* and *access control*. In each situation, our target is to balance between functionality and secureness while achieving a lightweight operation of the aforementioned mechanisms that will not impose any restrictions (e.g. increased power consumption) on the personal use of a mobile device.

4. CONCLUSIONS

In this paper we have tried to point out the directions of our research on Mobile Grid computing with respect to importand aspects such as scheduling, performance and security, in various networking contexts. Issues such as device heterogeneity and impact of human factors introduce further complexity in the path towards the realization of a Mobile Grid and certainly require proper investigation.

5. REFERENCES

[1] M. Migliardi, M. Maheswaran, B. Maniymaran, P. Card, and F. Azzedin, "Mobile interfaces to computational, data, and service grid systems.," *Mobile Computing and Communications Review*, vol. 6, no. 4, pp. 71–73, 2002.

[2] S. Kurkovsky and Bhagyavati, "Wireless grid enables ubiquitous computing.," in *Proc. of ISCA PDCS*, 2003.

[3] K. Katsaros and G.C. Polyzos, "Optimizing operation of a hierarchical campus-wide mobile grid," in *Proc. of PIMRC'07*, Athens, Greece, 2007.

[4] E. Raftopoulos M. Karaliopoulos, M. Papadopouli and H. Shen, "On scalable measurement-driven modelling of traffic demand in large wlans.," in *Proc. of the 15th IEEE LANMAN Workshop*, Princeton NJ, USA, 2007.

[5] T. Sander and C. Tschudin, "Towards mobile cryptography," in *Proc. of the IEEE Symposium on Security and Privacy*, Oakland, CA, USA, 1998.

[6] S.D. Kamvar, M.T. Schlosser, and H. Garcia-Molina, "The EigenTrust Algorithm for Reputation Management in P2P Networks," in *Proc. of WWW '03*, Budapest, Hungary, May 2003.

[1] A MN may attempt not to actually provide its resources while benefiting from the contribution of others.