Distribution over Mobile Environments

Frédéric Le Mouël
IRISA / INRIA
Campus Universitaire de Beaulieu
35042 Rennes Cédex, France
flemouel@irisa.fr

Françoise André
IRISA / University of Rennes 1
Campus Universitaire de Beaulieu
35042 Rennes Cédex, France
fandre@irisa.fr

Keywords

Mobile Computing, Adaptive Systems, Distribution, Placement, Load Balancing.

ABSTRACT

The development of mobile computing, combined with the large-scale information systems growth, makes it possible to access an important volume of information and services anywhere anytime. However, due to this dynamic environment, applications are subject to important variations which must be addressed by an adaptive behavior. We propose in this paper an approach which consists in an adaptive distribution of applications or services over mobile environments. This distribution uses *contextual* resources to overcome the poorness of available resources on mobile entities, reduces and regulates variability effects, and takes benefit of the local environment when it is possible.

1. INTRODUCTION

The importance of mobile computing is rapidly growing, as well as the expansion of Internet use. These developments are the result of technological advances in several domains. High speed networks and their interconnections allow an easy access to lots of information and services, which are constantly renewed. Wireless-communication networks (GSM, satellites), combined with the development of mobile computers (notebook, PDA), allow users to move around with their own computing environment and to have a constant access to these information sources. Combination of these two aspects shows that supporting mobile users on information systems is essential, and this must be done by offering the user the best quality of service available in spite of variability constraints of the network (bandwidth, latency) or of the mobile entity (CPU speed, memory size, disk capacity, screen display, power consumption, etc).

In this paper, we propose an approach which consists in an adaptive distribution of applications or services over mobile environments. The paper is organized as follows. First, we explain in section 2 why and how our approach improves the QoS. In section 3, we examine other works in mobile computing. Then, we conclude and give future work in section 4.

2. DISTRIBUTION: A WAY OF IMPROVEMENT

We notice that most mobility adaptation approaches use external resources to the mobile entity to search for information or services, computation power, storage capacity or security. Most of the time, these approaches implement servers or/and proxies or agents on static networks. Each of these approaches handles with a particular problem of the mobility in a specific context: connection/disconnection, data consistency in mobile file systems for example. They are not generic nor extensible and try to remedy some disadvantages without looking to take benefit from the mobility.

From these observations, we adopt an approach which tries to improve the QoS by a well-suited adaptive distribution of applications or services over this mobile environment. Indeed, a mobile user is working in an environment made up of *data*, *resources* and *services*. Resources concern the hardware of mobile entities (memory, disk, CPU, etc) and of the environment (local networks, devices like DVD, scanner, printer, fax, etc). Services concern software components relevant for the application or the user such as a database or a reservation service. These data, resources and services, which correspond to the user's actual, local situation, constitute his *context*. Contextual information such as the number of hosts around, the workload of these hosts or links, or the distribution of services give the possibility to run applications in a distributed way.

This distribution improves the QoS in terms of:

(i) **User interactivity quality**: the user interactivity quality is measured through the response time of the system or applications. This response time depends on mobile physical constraints as the memory, the available CPU or the disk size especially if several applications are running. The response time also depends on the geographical network constraints such as the bandwidth or the latency. The response time must be kept optimum or above a given level according to priorities. Distribution, with its accessibility to resources of other hosts, allows applications treatments to be executed quickly. This adaptation can also be a way to prevent a decreasing response time when the user is moving and connecting to a less powerful environment than the previous one.

(ii) **Data quality**: different elements constitute data: text, colors, images, videos, sounds, etc. But for various reasons (response time, mobile computer characteristics, user's preferences, etc), data must sometimes be changed by different operations such as distillation of images, transformations from one media to another, removal of part of data, linguistic translation. These degradations or modifications of data can be prevented or improved in some cases by the use of multi-machines environment. Meanwhile, if transformations must be performed, their quality is better in terms of response time.
(iii) **Accessibility**: Distribution allows to transparently access to resources and services, and moreover, brings processes closer to relevant resources or services they need, reducing the communication cost. This reduction of communication cost is characterized by a reduction of the number of messages (processes are close to information) and a reduction of the size of messages (treatments are performed before sending).
(iv) **Security**: some services have security requirements: secure protocols, cryptographic operation, user or/and group registration and authentication. Their implementation as a set of distributed entities has the advantage of being incoherent, if considering one single entity, as regards to hostile environments with tampering attempts of the system or users. Distributed entities are only globally coherent.

Thus, distribution is a way to provide an execution environment for applications which allows them to run in "normal" mode where variations would have forced them to run in a degraded mode. It may even better satisfy the QoS when lots of resources are available.

3. RELATED WORK

Many architectures or systems supporting mobile computing and adaptiveness have been introduced providing adaptations to material variations such as CPU speed, memory size, disk capacity, screen display, network performance (bandwidth, latency) or power consumption. We distinguish between two approaches: *masking-based approaches*, which attempt to mask mobility to users at a network or system level, and *adaptation-based approaches* which explicitly implement adaptiveness within applications.
Masking-based approaches aim at providing a mobile device with the same set of services as on a desktop machine. They consist in masking mobile aspects at network or system level by emulating the behavior of network protocols of a static system. For example, Mobile-IP [5] is masking protocols, and then allows mobile devices to keep their own address in spite of mobility. This approach has also been used in the Coda File System [3] by offering users to work in disconnected mode, masking disconnection periods by caching and hoarding techniques. These approaches do not require existing applications modifications, and they allow applications not explicitly designed for mobility to run unmodified in a mobile environment. Nevertheless they have limitations. A complete masking can not be obtained: since they ignore the application behavior, they are not able to adapt to all variations. For example, user must give assistance to the Coda File System to solve update conflicts.
Adaptation-based approaches [2] partially solve this drawback. Some information on the environment are provided to the application in order to enable them to adapt their behavior and their quality to the changing conditions. For example, Odyssey [4], a software platform, monitors resources, detects the changes and notifies interested applications, which are in charge of reacting and adapting. Adaptation-based approaches allow to implement more relevant adaptations than masking-based approaches due to their integration with the application. For example, multimedia applications in [1] (an MPEG player) use software feedback to dynamically discover the appropriate data stream to use (resolution, color depth, palette, frame rate) and detect when it becomes inappropriate. However, because of this implementation within the applications, they are not flexible nor extensible. Moreover, they are often limited to degradation of applications behavior whereas improvements are also possible. In particular, they only deal with low-level resource monitoring and do not take benefit of the semantic of data handled by the application to influence adaptation.

4. CONCLUSION AND FUTURE WORK

In this paper, we have presented our approach which consists in an adaptive distribution of applications or services over mobile environments. This approach can be placed between masking-based approaches and adaptation-based approaches. On one hand, part of mobility problems are masked to overcome the poorness of available resources on mobile entities and variability effects by placement and load balancing techniques. And, on the other hand, the distribution allows to take benefit of the local environment by bringing processes closer to application relevant data or services needed.
We are currently developing a first prototype of our system using an object-oriented approach and we are planning to validate it with an electronic press application. This application is suitable for different reasons. It is an interactive application which uses an important volume of data (texts and images), contextual data and services (weather, cinema, regional information). It also needs an important calculating power for images treatment, pages layout, profile calculation, statistics, etc.

5. REFERENCES

[1] J. Inouye, S. Cen, C. Pu, and J. Walpole. System Support for Mobile Multimedia Applications. In *Proc. of NOSSDAV'97*, pages 143–154, St. Louis, Missouri, May 1997.

[2] A. D. Joseph, J. A. Tauber, and M. F. Kaashoek. Mobile Computing with the Rover Toolkit. *IEEE Transactions on Computers: Special issue on Mobile Computing*, 46(3):337–352, Mar. 1997.

[3] J. Kistler and M. Satyanarayanan. Disconnected Operation in the Coda File System. *ACM Transactions On Computer Systems*, 10(1):3–25, Feb. 1992.

[4] B. D. Noble, M. Satyanarayanan, J. Tilton, J. Flinn, and K. Walker. Agile Application-Aware Adaptation for Mobility. In *Proc. of SOSP'97*, pages 276–287, Saint-Malo, France, Oct. 1997.

[5] C. E. Perkins. Mobile Networking Through Mobile IP. *IEEE Internet Computing*, 2(1):58–69, Jan. 1998.