Empowering Multimedia Applications in MAPGrid

Yun Huang and Nalini Venkatasubramanian
Donald Bren School of Information and Computer Sciences
University of California, Irvine, CA 92697
{yunh, nalini}@ics.uci.edu

Abstract

MAPGrid system exploits intermittently available grid resources to support mobile applications. In this poster paper, we address the problem of resource discovery and data placement in such a grid-based environment, specifically for running mobile multimedia applications. Our proposed efficient and adaptive resource management mechanisms ensure effective resource utilization and satisfy various QoS requirements of multimedia applications.

1 Introduction

We are observing a dramatic increase in the amount of multimedia content being delivered over the Internet and wireless networks today. Multimedia enhanced applications, e.g. distance learning, video-conferencing, news-on-demand, etc are resource intensive and consume significant network, storage and computational resources. They have diverse Quality of Service (QoS) requirements that determine the utility of the (perceived) information to the end-user. Also resource constraints in mobile devices and dynamic connectivity of wireless networks challenges these applications. Thus, efficient and adaptive resource discovery and data placement mechanisms are required to ensure effective resource utilization and satisfy QoS requirements of increasing number of users.

Global grid infrastructures allow the use of idle computing and communication resources distributed in a wide-area environment. A properly managed grid system can form the infrastructure upon which multimedia applications can be executed. Also because proxy-based techniques have been developed to provide localized computation and storage to enable distributed and ubiquitous service availability. In MAPGrid, available grid resources in the wired networks within close proximity to the mobile device are applied to support strategies such as proxy caching [1], proxy-based transcoding [2] or task-offloading [3], so that stringent resource needs of mobile multimedia applications can be alleviated. In fact, applications that provide delivery of multimedia information are especially well suited for grid based environments; often the content delivered is read-only, and coherence of multiple replicas is not an issue.

However, systems on a computational grid are available only intermittently. Effective load management in such a multimedia environment requires: (1) resource discovery and scheduling mechanisms, which select optimal grid proxies to service incoming requests and ensure user QoS satisfaction while grid resources are intermittently available; (2) data placement mechanisms, which determine proper numbers of data replicas that should be placed on each proxy, so that both data access cost and data replication cost can be reduced. We argue that our proxy-based resource scheduling and data placement approaches take into account of the intermittently availability of grid proxies [4, 5] that other proxy-based solutions fail to consider. Research efforts of grid computing technologies have focused on developing and implementing computation task scheduling and data replication techniques for computational-intensive and data-intensive scientific applications. Directly utilizing those techniques to implement MAPGrid system for multimedia applications (running on mobile users) will not lead to optimal performance [6, 7], because they do not account for providing adaptive continues service and hence are unable to support mobile multimedia applications.

2 Using Grid Resources as Proxies for Mobile Multimedia Applications

Using video streaming as a driving application, we describe techniques for addressing the key issues in resource allocation and data placement for QoS-based mobile multimedia applications. QoS needs can be expressed as user-perceived quality needs (e.g. video quality) that are translated into lower level application/system parameters. The QoS requirements e.g. required network transmission bandwidth, will be determined by streaming a certain video streaming object. QoS statements may specify constraints on timing, availability, security and resource utilization at various levels of abstraction. For instance, timing based QoS requirements can be specified using abstract properties such as correct/timely data delivery and uninterrupted service. These properties can be translated to concrete application parameters such as jitter, end-to-end delay, synchronization skew and/or concrete resource requirements such as network and disk bandwidth and buffer requirements [8]. Resources required to support these multidimensional notions of QoS in mobile applications can be in form of computation (CPU), storage, bandwidth, memory or services that must continue to be available as the user moves in the mobile infrastructure.

In MAPGrid, a client who may uses a mobile device initiates a multimedia request, R $<VID,T,Q_{MIN},Q_{MAX},E_R,itinerary>$, where VID identifies the requested video object, T represents the whole service period, optional information includes the lowest QoS level (Q_{MIN}) and the highest QoS level (Q_{MAX}), current residual energy E_R and user's mobility information ,*itinerary* (NULL, if no mobility information is available). A mediator utilizes the resource availability information stored in the context repository (maintained by the adaptive context collection module) to decide optimal proxy resources for servicing this incoming request. The mediator also selects proxies to replicate requested data objects, initiates data transfer from server storage to grid proxies and proxies to mobile clients. The selected grid proxies processes the request by transcoding the video segment and transmits the video stream via wireless links to mobile clients. Significant changes in the availability of the allocated grid resources will trigger the resource reprovisioning process to adapt the resource allocation accordingly. When the request terminates, the resources are reclaimed along the connection and the context collection module updates the resource availability status in the context repository.

Given these multimedia requests and information of grid resource availabilities, the mediator performs a static resource discovery, which aims to increase the overall acceptance of requests in the system by selecting optimal grid resources for each request, while satisfying users' QoS requirements. We proposed an approach [9] that divides the whole service period T into non-overlapping chunks (possibly of different sizes), each of which is mapped to an appropriate grid proxy e.g. the one that is geographically close and lightly loaded. A graph theoretic technique is proposed [10] for selecting an optimal set of grid resources to service each chunk. Decisions are made by taking into consideration all the following factors: (a) intermittent availability of grid resources, (b) currently allocated workloads and predicted future workloads on grid resources, and (c) user's distance to grid resources. The problem of discovering intermittently available grid resources is regarded as a maximum flow problem. A feasible maximum flow solution that meets resource constraints corresponds to a possible scheduling solution; the basic solution has been adapted to develop a family of policies that cater to various application QoS needs. Video objects are also divided into equal-sized segments. Corresponding video segments are placed onto selected grid proxies either on-demand or beforehand.

As unused replicas occupy premium storage resources of intermittently available grid proxies, data placement decisions directly affect request acceptance ratios for different multimedia objects; a bad placement policy will deteriorate system performance. We also developed intelligent data placement techniques for placing mobile multimedia data objects on intermittently available grid proxies. We proposed a two-tier architecture [7], where the upper tier captures grid related features, e.g. intermittent availability of grid proxies and the lower tier captures features associated with the mobile environment, e.g. data request patterns of mobile users. An interval tree-based data structure is applied to store information of these two tiers. We further devise intelligent data placement heuristics that effectively reduce data replication cost and data access cost by applying information maintained in the interval tree.

Grid proxies are participating machines on which applications can be randomly started or stopped, causing fluctuations in resource availability. Allocation mechanism must be capable of dealing with proxy failures and changes in proxy resources. For example, when a specific proxy becomes unavailable, the mediator retrieves information from the directory service about requests that are scheduled on the failed proxy, and triggers the re-scheduling process for each invalidated service. In order to reduce service failures and minimize service recovery time, the solution determines the order in which to migrate the disrupted services onto available proxies [11]. Our adaptation approaches significant decrease in the number of requests that fail to complete due to dynamic changes in proxy availability.

3 Prototype Implementation

We are implementing a system prototype over a network of volunteer machines at UC Irvine that serves as a testbed. The initial prototype of MAPGrid assumes a WLAN infrastructure. We address practical issues of implementing MAPGrid system in [12]. We introduce a comprehensive and versatile system design, where fine grained and well-defined meta-level services enable flexibility in easy plug-in of different policies. The underlying implementation of MAPGrid subsumes a multithreading model for multitasking the various models for resource discovery and data placement services, which effectively improves system performance. Currently the mediator runs on a Fedora machine. Proxies are both Windows and Linux desktops. Mobile client can use either a Sharp Zaurus handheld device or a Windows laptop. To allow for easy extensibility and re-use, the prototype MAPGrid implementation uses the Globus Toolkit, along with standardized messaging interfaces such as XML for metadata exchange. We implement a video streaming application for testing system performance. Users can use GUI to select preferred service type [4]. Video clips are displayed by JMF (Java Media Framework). When mobile user moves among wireless regions, data service connection will be changed to closer proxy machine as scheduled by the mediator, i.e. downloading file from nearby proxies during the whole service period.

4 Concluding Remarks

Our proposed techniques for MAPGrid system support diverse QoS requirements of mobile multimedia applications while improving overall system performance in terms of user QoS experience, client admission ratio, grid throughput and grid utilization. We illustrate the utility of MAPGrid system through video streaming applications executing on mobile devices. We also proposed an integrated solution that adapts to dynamic changes in device energy consumption and unpredictable grid resource availability without compromising application QoS [13]. We now are currently integrating proxy-based techniques developed by DYNAMO project [14] into MAPGrid. We believe our MAPGrid system will lead to promising multimedia applications for mobile grid environments.

Acknowledgement MAPGrid project is sponsored by Office of Naval Research (ONR) as part of the DoD Multidisciplinary University Research Initiative (MURI) Project CONTESSA. We thank Luca Clementi, Shivajit Mohapatra, Yang Wang and Bo Xing for their contributions to the design and implementation of MAPGrid system.

References

[1] Aameek Singh, Abhishek Trivedi, Krithi Ramamritham and Prashant Shenoy, "PTC: Proxies that Transcode and Cache in Heterogeneous Web Client Environments", *World Wide Web*, vol. 7, n. 1, pp. 7–28, 2004.

[2] S. Mohapatra, R. Cornea, N. Dutt, A. Nicolau and N.Venkatasubramanian, "Integrated Power Management for Video Streaming to Mobile Handheld Devices", in *ACM Multimedia*, 2003.

[3] Shivajit Mohapatra and Nalini Venkatasubramanian, ""PARM: Power-Aware Reconfigurable Middleware"", in *ICDCS-23*, 2003.

[4] Yun Huang and Nalini Venkatasubramanian, "QoS-Based Resource Discovery in Intermittently Available Environments", in *HPDC*, 2002.

[5] Yun Huang and Nalini Venkatasubramanian, "Data Placement in Intermittently Available Environments.", in *HiPC*, pp. 367–376, 2002.

[6] Yun Huang and Nalini Venkatasubramanian, "Supporting Mobile Multimedia Services with Intermittently Available Grid Resources", in *HiPC*, 2003.

[7] Yun Huang and Nalini Venkatasubramanian, "MAPGrid: A New Architecture For Empowering Mobile Data Placement in Grid Environments", Submitted for Publication.

[8] Nalini Venkatasubramanian, Carolyn Talcott and Gul A. Agha, "A formal model for reasoning about adaptive QoS-enabled middleware", *ACM Trans. Softw. Eng. Methodol.*, vol. 13, n. 1, 2004.

[9] Y. Huang and N. Venkatasubramanian, "Supporting Mobile Multimedia Services with Intermittently Available Grid Resources", in *Proc. of HiPC*, 2003.

[10] Yun Huang and Nalini Venkatasubramanian, "QoS-based Resource Discovery in Intermittently Available Environments", Jul 2002.

[11] Yun Huang, Shivajit Mohapatra and Nalini Venkatasubramanian, "An Energy-Efficient Middleware for Supporting Multimedia Services in Mobile Grid Environments", in *IEEE International Conference on Information Technology*, 2005.

[12] Yun Huang and Nalini Venkatasubramanian, "Design and Prototype Implementation of MAPGrid: A Grid System for Mobile Applications", submitted for publication.

[13] Yun Huang, Shivajit Mohapatra and Nalini Venkatasubramanian, "An Energy-Efficient Middleware for Supporting Multimedia Services in Mobile Grid Environments", in *ITCC '05: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume II*, pp. 220–225, Washington, DC, USA, 2005, IEEE Computer Society.

[14] "The DYNAMO project, http://dynamo.ics.uci.edu".