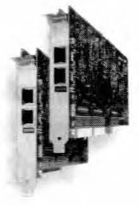


VIDEO STREAMING WITH QUALITY ADAPTATION USING COLLABORATIVE ACTIVE GRID NETWORKS

Juan Carlos Peláez López,
Jorge Mata Díaz
Departamento de Ingeniería Telemática. Universitat Politècnica de Catalunya
Jorge Girona 1-3. Campus Nord, Mód. C3, UPC. 08034 Barcelona
Teléfono: +34-934016028 Fax: +34-934011058
E-mail:{jpelaez,jmata}@entel.upc.es

ABSTRACT

Due to the services and demands of the end users, Distributed Computing (Grid Technology, Web Services, and Peer-to-Peer) has been developed rapidly in the last years. The convergence of these architectures has been possible using mechanisms such as Collaborative work and Resources Sharing. Grid computing is a platform to enable flexible, secure, controlled, scalable, ubiquitous and heterogeneous services. On the other hand, Video Streaming applications demand a greater deployment over connected Internet users. The present work uses the Active Grid technology as a fundamental platform to give a solution of multimedia content recovery. This solution takes into account the following key concepts: collaborative work, multi-source recovery and adaptive quality. A new architecture is designed to deliver video content over a Grid Network. The active and passive roles of the nodes are important to guarantee a high quality and efficiency for the video streaming system. The active sender nodes are the content suppliers, while the passive sender nodes will perform the backup functions, based on global resource control policies. The aim of the backup nodes is minimize the time to restore the system in case of failures. In this way, all participant peers work in a collaborative manner following a multi-source recovery scheme.

Furthermore, Video Layered Encoding is used to manage the video data in a high scalable way, dividing the video in multiple layers. This video codification scheme enables the quality adaptation according to the availability of system resources. In addition, a buffer by sender peer and by layer is needed for an effective control of the video retrieve. The QoS will fit considering the state of each buffer and the measurement tools provide by the Active Grid on the network nodes.

Keywords: Peer-to-Peer Grid Architecture, Services for Active Grids, Streaming Media, Layered Coding, Quality Adaptation, Collaborative Work.

1. INTRODUCTION

The users of Internet demand more and better multimedia services. The number of users who request multimedia information in Internet is increasing. This rising demand involves two challenges: 1) Internet offers only best-effort services, whereby intelligence techniques are required to deal with the user requirements of multimedia applications; 2) Streaming multimedia demands higher and ensured resources availability in the network, in the content supplier and distributor, and moreover higher capabilities in the receiver user system. The challenges mentioned have been reached by the distributed computation evolution and development. This evolution has changed the centralized models for the distributed architectures.

Grid computing [1] and Peer-to-Peer models [2] arise from the last evolution in the development chain of the distributed computation and communication. Grid technology creates secure and efficient Virtual Organizations (VO). Grid is the novel enabling technology to transparently access computing and storage resources anywhere, anytime and with guaranteed QoS. The new application field for the Grid Computing is the multimedia services management. These services are actually the more requested by Internet users. File Sharing P2P applications emerge as the best alternative for new Internet services. File sharing systems, such as Napster [2], is an example of the large current acceptance of P2P architecture. Hence, the contribution of P2P models is the flexible and scalable support of VOs with a fast discovery of accessible services.

A general classification of multimedia applications is needed to clarify the type of multimedia services addressed in this work. Two groups of multimedia applications are: 1) Real-Time: video-on-live, or videoconference. They receive and visualize the video at the same time the capture is accomplished. 2) Not-real-Time: Video-on-Demand (VoD). Here, a user decides the moment that wishes to receive and visualize the video; moreover, the video is pre-captured and pre-stored. Streaming video is classified within the second group. The current file sharing services allow to retrieve a multimedia content in the download-playback mode. The present work is focused in video streaming services which follows a Playback-during-download mode.

Different architectures have been proposed to support decentralized media streaming services. Caching Server [3] was the first attempt to decentralize the content. It consists in storage the content among points that have already visualized the video. Nevertheless, caching technique transfers the bottleneck to several points but does not eliminate them completely. The Content Distribution Networks (CDN) [4] provide multimedia content recovery with an acceptable degree of scalability. But the high implementation cost of powerful equipment servers deployed in different geographic points limits this option. Multicast technology emerged under the idea to minimize the number of packets transmitted through the network. Real-time multimedia transmissions have been addressed by both IP multicast or Network Level Multicast [6] and Application Level Multicast or also called Overlay Multicast [5]. But multicast schemes are not able to support not-real-time solutions. Nowadays, Grid architectures are the emergent solutions for real-time and non-real-time multimedia services. Grid architectures provide a fast deployment of these services by means of interoperability, the easy management of specific QoS policies, and the organization scalability.

Home entertainment and shopping, digitals video library, movie-on-demand, distance learning, medical information services, interactive games, news/services-on-demand and, emergency and disasters management are some of the most outstanding applications of a media streaming service. Push and pull services are very attractive in the current market of Internet users. For instance, a push service (initiated by the server) could be the transmission of a descriptive video of the tourist site to some user with a mobile device triggered by the user location and current time. On the other hand, a pull service (initiated by client) could be the multimedia message transmission from an isolated site where communicating is only possible by wireless networks.

Parallel technologies like third generation mobile communications, wireless Internet, video compression, among others, provide an improved structure of communications for the media streaming services. A controlled Active Grid architecture is the best alternative for supporting VoD applications. Moreover, a global efficiency of the V.O. will be obtained applying collaborative work concepts over a common interaction environment. This efficiency is measured in terms of used bandwidth, experimented latency and loses. The Active Grid QoS framework [7] is based on OGSA (Open Grid Service Architecture), where the Grid Security Infrastructure (GSI) protocol and Global Resource Control techniques are included.

In conclusion, an efficient and scalable VoD Streaming Service using the P2P Grid technology is introduced in this work. The rest of the paper is organized as follows. The problem definition is addressed in Section 2. The designed model and architecture is explained in Section 3. Section 4 concludes the paper and describes future works.

2. PROBLEM DEFINITION

There are two ways to recover a video file over Internet. The first one is called 'Download-and-play' mode, and the second one is the 'Play-during-download' mode. Video streaming is a service that demands the application of the second style. The users of video streaming expect to watch the video few seconds after their request. Likewise, the users employ different terminals to access to multimedia services networks. Hence, the ubiquity in user terminals must be considered. Also, the

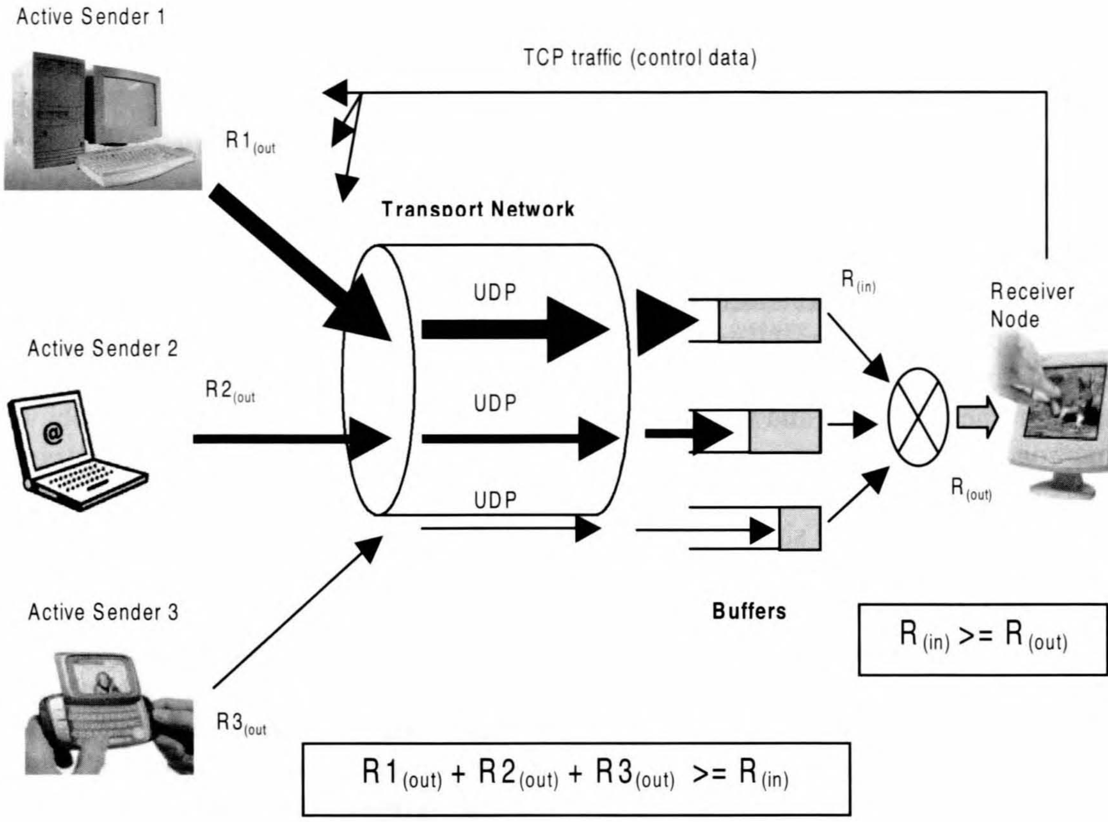


Fig. 1: Multi-source recovery

network conditions to transmit the video over the networks are variable and unpredictable. These conditions should be controlled by the active Grid in order to guarantee the QoS in the audiovisual services.

The video recovery over Grid networks in a play-during-download mode includes the following processes and characteristics: 1) search of nodes with the stored video; 2) selection of the best content suppliers; 3) congestion control mechanism in the active network nodes; 4) buffering techniques in the receiver side; 5) adaptation of the video playback rate in the receiver; 6l retardonodo emisor, as nodos y la ruta a seguir hasta el nodo receptor, midiendo el ancho de banda disponible y ofrecido) system recovery after a failure; and, 7) security mechanisms within the V.O.

The Active Grid architecture solves the search, selection and classification of nodes that store the video content required. The selection of the best content suppliers will be the results of the analyzing resources of each candidate node and transport network resources. The delay time and packet lost probability are also considered to select the best peers in the admission control function of the video services. This selection will be upgrade each time the network conditions change. A Grid QoS mechanism is required to evaluate the changes in the network conditions. A congestion situation in the network will imply to fire a reconfiguration of the video streaming recovery system. This mechanism is based on global resource control policies of the Grid Network. Buffering techniques in the receiver side are used to manage the playout video during an irregular system condition. The content in the buffers should be controlled every time, adapting the quality of the video.

Video quality adaptation will be supported by the encoding technique called 'Video Layered Encoding'. There will be multiples layers, since the base layer to enhanced layers. This scheme of layered encoding will allow adapting the quality video in a gradual manner. A buffer per layer and per peer is proposed in this work. The model of multiples buffers is more efficient and more reliable when a system failure or a network congestion happen. Quality Adjustment Mechanisms are supported by the Active Grid nodes, which control the dynamic buffering.

In conclusion, the problem to deal is the video streaming retrieval. This recovery must be reliable, efficient, scalable, economic and adaptive. Dynamic and restricted environments like Internet are considered as part of the problem. The system constraints are: 1) dynamic behaviour and cooperative level of the V.O. nodes; 2) heterogeneity of nodes; 3) changes in the network conditions; 4) video application non tolerant to the delay; and 5) reception constrains in the receiver peer side.

3. ACTIVE P2P GRID MODEL

A. Principles of the proposed model

1) Collaborative work and Load Balancing: To apply the foundations of collaborative work means to obtain the results of a process with a smaller cost. Optimizing the use of the available resources and achieving an efficient distribution of work between the nodes of a network. The Active Grid network facilitates the application of this concept, using its global knowledge of the system. An initial load balance is performed for the active grid service. The resources of each candidate node are considered in the streaming distribution process. A load balancing efficient mechanism will help to reach suitable results derived from a distributive fairness. A initial media content distribution is performed by the Grid system, based on the popularity of the video and the hierarchical organization of the nodes [7]. Likewise, Video sequences are distributed over the grid network taking into account the CPU, Memory, disk I/O workload, and disk capacity of the active grid nodes.

2) Multi-source recovery of Video Streaming: The global behaviour of the system depends on the individual performance of each one of its components. In the case of video streaming services will be much more efficient in time and cost when a recovery from multiple sources is carried out instead of a single recovery. Efficient exploitation of resources, scalability for media services and

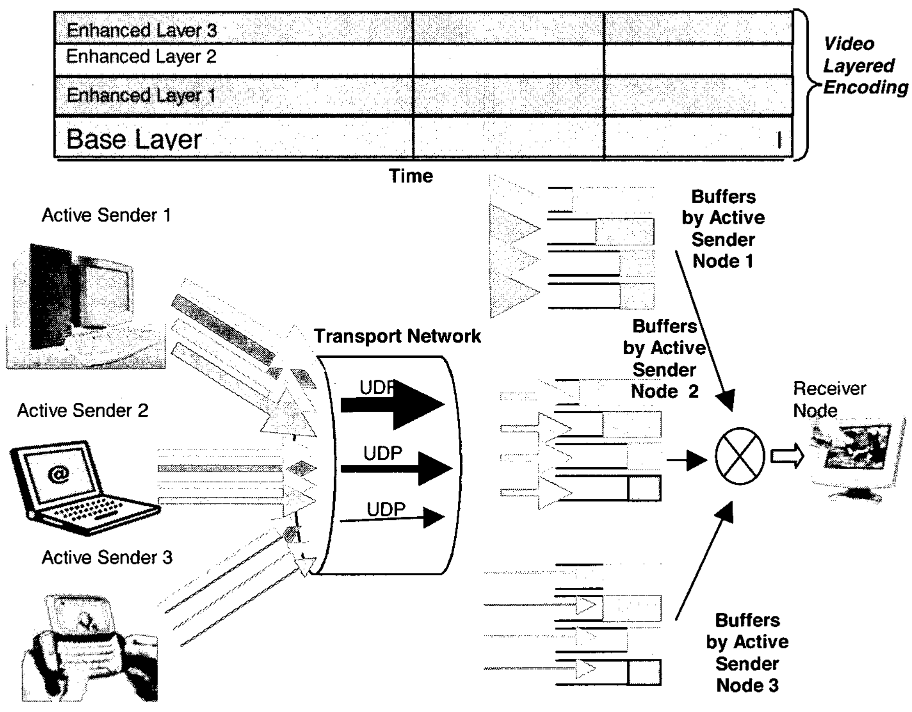


Fig. 2: Buffer Distribution by peer and by layer.

protection to error-prone wireless networks are achieved by this multi-source recovery scheme (see Fig. 1).

3) Adaptive Quality and Buffering Techniques: The dynamism of the P2P Grid network is motivated by the variability of the peer connection time. On the other hand, the conditions of the network change too. This variation is produced by the congestion, resources reservation, management policies, etc. Hence, it is required the implementation of a Grid Computing QoS framework, where the video layered encoding joined to TCP-friendly congestion control protocols are used. Moreover, the use of buffers in the receiver side is requested. The buffers smooth the variations of the arrival rate and allow to maintain a constant presentation rate. A control mechanism of the buffers state in the receiver has high-priority in our proposal. Multiples buffers are employed, one buffer per peer and per layer. This scalable model of multiples buffers is more efficient and more reliable during a system failure and network congestion situations.

B. Components of the proposed model

1) Receiver Node: This node will be the main participant in the video recovery system. It will initiate the request of a multimedia content and interacts with Quality Adjustment Mechanism of the Active Grid in order to redistribute the work between the active nodes, and acts over the active networks nodes to maintain the QoS. This work will be allocated in a fair and weighted sense, considering the constraints of sender nodes. The receiver node will demand as minimum the fulfilment of the service level specification to obtain a playout quality.

2) Active Sender Node: It will be one of the content suppliers, where the contribution is according to its transmission capabilities and its processing availability. This node will receive an allocation packet from the receiver node. Furthermore, it will establish a congestion control TCP-friendly mechanism. Also, each Active Sender Node will communicate with Passive Sender Nodes providing control information to all participants.

3) Passive Sender Node (backup): This peer will perform backup functions when an active sender node fails or its behaviour degrades the system quality. This node acts as a Pro-active listener and agent. It attempts to cover immediately any incidence in the subgroup of active sender nodes. The backup node has information of the whole system, since it exchanges control information with receiver node as well as all active sender nodes. Receiver node could pre-assign (using initial distribution algorithms) the backup role with respect to a possible failure of a determined active sender node. This pre-assignation will be based on active test algorithms of possible disconnections or failures. The target is to reduce the reaction time of the backup nodes, and hence, minimize the time to restore the system.

4) Video Layered Encoding: Our objective is to implement the video codification by multiple layers. This implementation will improve gradually the playout quality, according to the availability of system resources. There are two mechanisms to transmit the video in a flexible way. The first option is the Video Layered Encoding which provides high scalability. It divides stream into multiple video layers, being the base layer the fundamental one. The other layers, called enhanced layers, are used to improve the image quality. These enhanced layers require the base layer to be decoded in the receiver side. The second option is the Multiple Description Coding (MDC). This technique is a scalable solution, where each layer can be decoded separately. However, a reduction in the efficiency of the video compression is introduced; in consequence the transmission needs more bandwidth. MDC is more suitable in scenarios where the error-prone channels are considered. Additionally, when the video format is not the same in both sender and receiver sides, mechanisms for video format adaptation are required. Transcoding techniques could be used to adapt the video format (i.e. Format change of MPEG to H.263). In this case, a third element within the Active Grid Network is needed to implement the transcoding function. The new transcoding nodes contribute with their CPU and bandwidth resources.

5) Number and capacity of receiver buffers: The receiver node has a buffer distribution mechanism by sender node and by layer (see Fig.

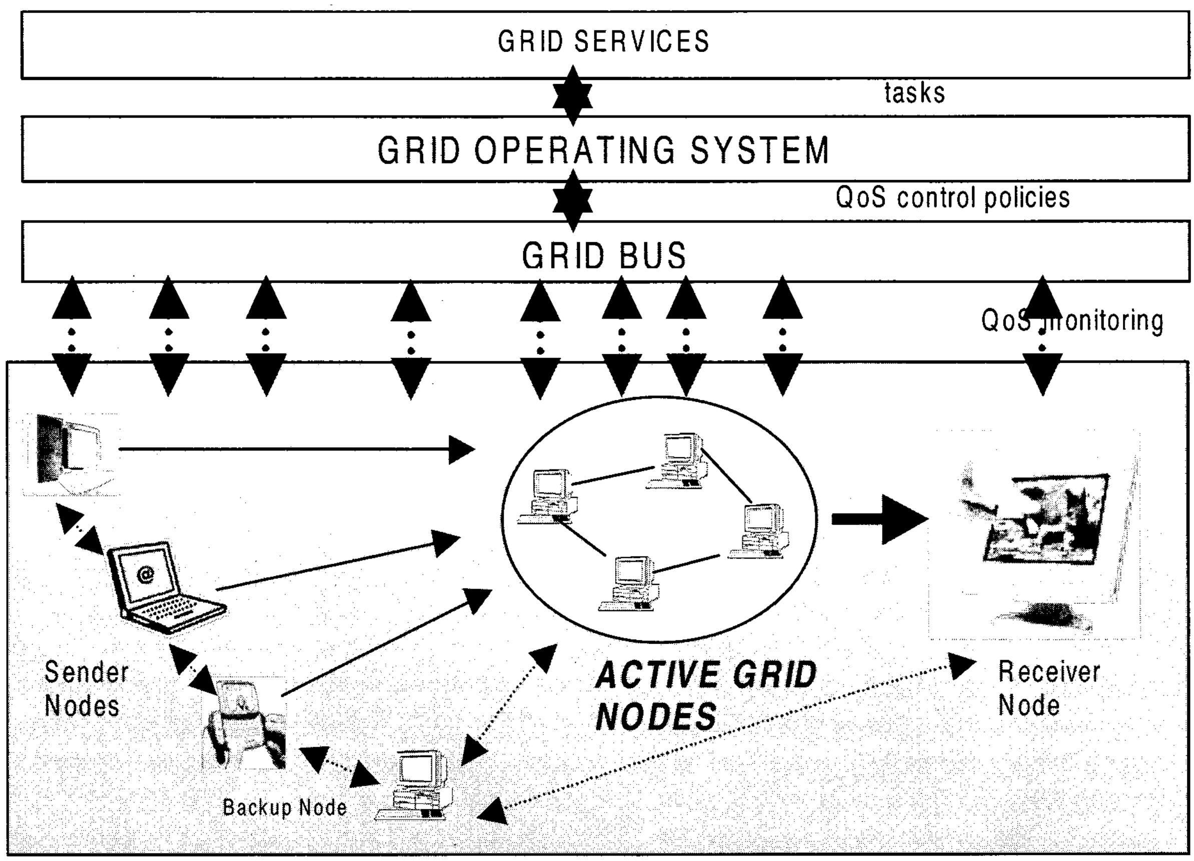


Fig. 3: Active P2P Grid Transmission Architecture

2). The receiver node will request a percentage of base layer and one or more enhanced layers to each active sender nodes. Hence, a control by peer and by layer will be applied. Our model prioritizes the control of buffers corresponding to the base layer. This priority is required to mitigate the effects of congestion, loss or delay in the network; hence, the first buffers that will be withdrawn will be the buffers associated to the enhanced layers. To find a function that helps to define the optimal state of each buffer is one of our challenges. Finally, a control function is implemented to regulate the incoming flow rate of each buffer in order to control their fullness. In an extreme situation, the system must react to fitting the video quality.

6) Active Grid Computing QoS framework: The quality adjustment could be seen from two points of views. The first viewpoint is an adjustment of coarse granularity. If a degradation of the system occurs and there is not enough information in buffers, the receiver must eliminate the highest video layer. The second viewpoint is an adjustment of fine granularity, which implies the reduction of intra-layer transmission rate. When a temporal congestion occurs, the receiver request to active nodes to reduce the transmission rate per layer. Major changes will be managed by the Active Grid. Each active grid node is able to process additional complex functions, and not only routing traffic. These functions could be the video transcoding, transmission forwarding decisions, terminal dependence forwarding, QoS monitoring, among others.

In conclusion, in this section the basis of the proposed model (Fig. 3) have been presented. The model consists in the following relevant principles: collaborative work, load balancing, multi-source recovery, adaptive quality and buffering techniques. All of these principles are implemented by the above mentioned elements.

4. CONCLUSION AND FUTURE WORK

In this work, an exhaustive analysis of the topics that involve the Video-on-Demand streaming services using Active P2P Grid networks has been addressed. The emergence of these services and the heterogeneity of resources in VOs make this topic an important and vast field of latest research work. This work presents a novel Active Grid Video Streaming Recovery architecture. The concepts of Collaborative Computing, Fairness Distribution and Quality Adaptation are the basis of this proposal.

This architecture has been designed using two roles of sender nodes. These nodes are differentiated by their capabilities and their contribution to the global system performance. The presented model divides the nodes set in two subclasses. The first subclass is the active sender nodes which send the video to the receiver node in a collaborative manner. This collaboration mechanism involves a multi-source retrieval operation. The other subclass is the passive sender nodes, called backup nodes. These backup nodes are prepared to react in whatever irregular situation or whenever the Active Grid operating system decides. The collaboration between active and passive sender nodes improves the behaviour of the system and hence, the video quality perceived by the users.

Our model uses an Active Grid QoS framework to adjust the performance of the system. The Grid system applies QoS policies; meanwhile the active grid nodes will be monitored by processes to accomplish the quality requirements. Video Layered Encoding is used to exploit the implementation of multiples buffers. Buffering techniques are implemented to control the state of each buffer. These techniques ensure available buffers for the base layer mainly, and others buffers for the enhanced buffers. Security Policies are also integrated to guarantee the authenticity and integrity of the received video information. VoD services are provided based on Grid Security Infrastructure (GSI) protocols.

Finally, we are convinced that the present proposal can be extended in several directions, not without before making an analysis process of results and simulations. We are considering to use NS2 and Globus Toolkit. The future work could include the following subjects: Design of a prototype and extend to real applications.

REFERENCES

[1] I.Foster, C.Kesselman and S.Tuecke. «The Anatomy of the Grid: Enabling Scalable Virtual Organizations». Int. J. Supercomputer Apps., Vol. 15, No. 3, 2001

[2] Napster: http://www.napster.com

[3] L.Fan, P.Cao, J.Almeida, Z.Broker, «Summary cache: A Scalable Wide-Area Web Cache Sharing Protocol». IEEE Trans. Networking vol. 8, pp. 281-293, June 2000.

[4] G.Peng,»Content Distribution Network» www.ecsl.cs.sunysb.edu/~gpeng/ June 2003.

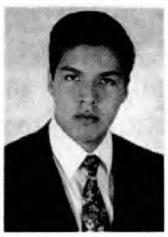
[5] Y.Chu, S.G.Rao, H.Zhang, «A Case For End System Multicast», Proc. of ACM SIGMETRICS, Santa Clara, CA, June 2000, pp 1-12.

[6] S.Banerjee, B.Bhattacharjee, C.Kommareddy, G.Varguese. «Scalable Application Layer Multicast». In Proc. of ACM SIGCOMM'02, 205-220, USA, August 2002

[7] X.He, X.Tang, J.You. «Supplying Instantaneous Video-on-Demand Services Based on Grid Computing». GCC2003: The Second International Workshop on Grid and Cooperative Computing, Shanghai, China, December 2003.

AUTHORS

Jorge Mata Díaz, es Doctor Ingeniero en Telecomunicación y profesor del Departamento de Ingeniería Telemática de la Universidad Politécnica de Cataluña. Su trabajo de investigación se enmarca dentro del grupo «Servicios Telemáticos» del mismo departamento y su actividad docente se desarrolla en la ETS de Ingeniería de Telecomunicación de Barcelona. Entre sus campos de investigación se encuentra el desarrollo y propuesta de servicios telemáticos de redes de comunicación terrestres, sin hilos y VSAT: servicios audiovisuales, sistemas distribuidos multimedia, servicios corporativos en entornos intranet e Internet, y servicios de seguridad y comercio electrónico. Los resultados de sus trabajos se han publicado en numerosas revistas y congresos tanto nacionales como internacionales (http://sertel.upc.es)



Juan Carlos Peláez López, es estudiante de Doctorado de la Universidad Politécnica de Cataluña, en el Departamento de Ingeniería Telemática, donde se encuentra vinculado al grupo de investigación de Servicios Telemáticos desde 2001. Recibió el grado de Ingeniero en Sistemas Computacionales de la Universidad Católica de Santiago de Guayaquil, Ecuador, en 1998. Entre sus campos de investigación se encuentran el desarrollo y propuesta de servicios distribuidos multimedia, tecnología peer-to-peer, Web Services, Grid Computing, y comercio electrónico. Actualmente se encuentra trabajando en su proyecto de tesis sobre «Video Streaming Peer-to-Peer with QoS for the recovery and delivery».