Animated Space Architecture for Multimedia Experience - ASkME

Niranjan, Rakesh Kothari, and Aura Ganz

***Abstract*— In "Animated Spaces" real-world objects can communicate with users in order to convey their purpose, function, and history. Achieving the vision of animated spaces will require adaptive streaming and delivery of video to mobile users. In this paper, we specify the technical requirements of such an animated space and use RFID technology to develop smart objects. We present an architecture for adaptive delivery of stored MPEG-4 encoded audio-visual information with varying wireless network conditions and client device capabilities. The proposed architecture is developed under constraints imposed by user preferences and multimedia content, to ensure smooth delivery of multimedia contents to mobile users. We implement a pilot implementation of this concept using RFID technology and evaluate it over a real network testbed.**

***Index Terms*—Adaptive Multimedia, Information Systems, MPEG-4, RFID.**

I. Introduction

WITH the rise of unprecedented new technologies (e.g., smart homes, shop-bots, pedagogical agents, wearable computers, personal robots, multi-agent systems, sensors, grids, knowledge environments) and their increasing ubiquity in our social and economic lives, ubiquitous computing offers a means to create "Animated Spaces" for users through their mobile devices (laptops, PDAs, mobile phones, etc.). In Animated Spaces, real-world objects can communicate with users in order to convey their purpose, function, and history. It can be used to provide location-aware information services, to support learning through design experiences [1], [2], [3] and cultivate interests and emergent community [4]. The vision of putting information in places has been a key goal of researchers developing augmented reality systems [5], [6], [7]. Information exchange through streaming video is a vital part of animated space architecture envisioned in this paper.

Manuscript received June25, 2004. This project was supported in part by the following grants: NSF-ANI-0319871, NSF-ANI-0230812, NSF-EIA-0080119, and ARO-DAAD19-03-1-0195.

Niranjan is with Electrical and Computer Engineering Department, University of Massachusetts, Amherst, MA 01003 USA. He is member of Multimedia Networks Lab. (e-mail: nniranja@ecs.umass.edu).

Rakesh Kothari is with Electrical and Computer Engineering Department, University of Massachusetts, Amherst, MA 01003 USA. He is member of Multimedia Networks Lab. (e-mail: rkothari@ecs.umass.edu).

Aura Ganz is with Electrical and Computer Engineering Department, University of Massachusetts, Amherst, MA 01003 USA. She is director of Multimedia Networks Lab. (e-mail: ganz@ecs.umass.edu).

In this paper, we present ASkME, software architecture for creating "animated spaces" comprising of smart real-world objects tagged with RFIDs and associated with audio-visual information. Deploying Animated Spaces architecture for mobile users, is a challenging problem due to the wireless channel characteristics such as high latency, high bit-error rates, limited bandwidth and frequent disconnections. To meet these challenges, ASkME employs MPEG-4 compression standard [8], [9] with its high bitrate scalability, compression efficiency and superior quality. Moreover, MPEG-4 compression standard is divided into a number of profiles and levels that makes it operable for a wide range of applications, end device and network conditions. ASkME architecture provides support for cataloguing the information in the form of video encoded in MPEG-4 at multiple bit-rates ranging from high to coarse quality and delivering the appropriate version of it depending on the network traffic. We also present the decision rules designed for intelligent mapping of device capabilities and network conditions to decide the quality of video to be delivered to the mobile user.

There are a variety of existing projects working to increase interaction between the physical and digital worlds [10], [11], [12], [13], [14]. None of these projects include the use of RFID for object identification and MPEG-4 encoded multimedia data for information delivery.

The paper is organized as follows: Section II presents the animated space architecture and its key components. In Section III, we introduce the implementation details and the testbed description used for evaluating the animated space architecture. Section IV concludes the paper.

II. Architecture

We developed a comprehensive system, composed of a test bed and framework for web intermediaries, aligned with existing related standards, for animated spaces where smart objects can dialogue with users about their function and purpose.

Some of the key components of our architecture are:

- RFID technology
- Proxy-based Information Exchange
- Multimedia Delivery over Wireless Networks
- Server Component Architecture

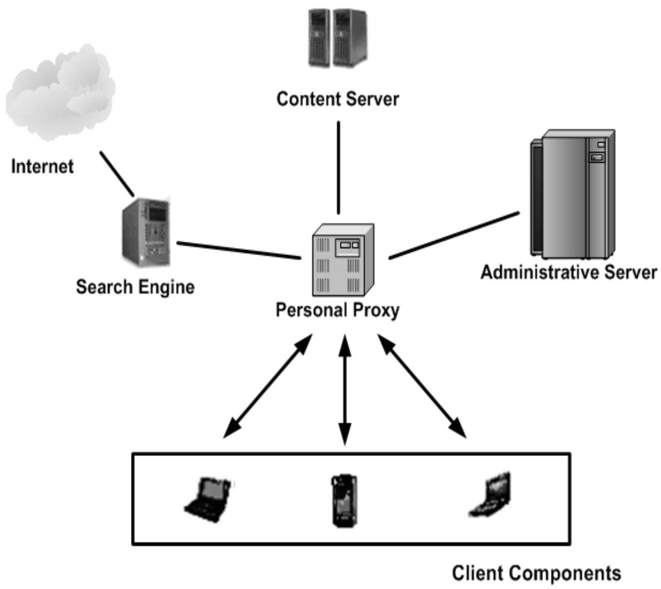

Fig. 1. ASkME Component Architecture

A. RFID Technology

All the existing projects related to smart room or augmented realities uses cameras, microphones, touched sensitive screens or GPS for user or object identification and location information. We use RFID technology for developing smart objects. RFID (Radio Frequency Identification) allows contact-less identification of objects using RF [15], [16], [17]. RFID system consists of a tag (transponder) and a reader (interrogator). A RFID tag is a small and inexpensive microchip that emits an identifier in response to a query from a RFID reader. RFID tags are used for virtual genesis of smart objects and RFID readers are attached to user's wireless handheld devices. Some of the advantages of using RFID for object identification over other technologies are:

1) Read-Write: The data stored in RFID tag can be updated which is useful for object classification and its indexing to provide low delay on user interface.

2) Non line of sight: The positioning of RFID tags on physical world objects are not critical as it doesn't require line of sight to read the RF identification tags.

3) Data capacity: It can store a larger amount of data which will reduce connection and data transmission on wireless medium.

4) Re-usability and Durability

5) Multiple read: Many tags can be read at same time in the field of view of the reader.

6) Security: Security (e.g. password authentication) can be implemented to protect private objects.

B. Proxy-based Information Exchange

The proxy architecture which includes a third entity between the server(s) and the client(s) represents a good approach to address the heterogeneity of clients and servers in the ASkME architecture. Adapting on the proxy means that there is no need to change existing clients and servers, and it achieves economy of scale. Proxy tasks are designed to behave transparently to clients and content servers.

C. Multimedia Delivery over Wireless Networks

Contextual learning is provided through information associated with the objects in the form of text or compressed audio and video. Audio-Visual information has significant bandwidth and latency requirements. We use MPEG-4 compression standard for video compression. MPEG-4 video compression specification has been developed as an open standard to encourage interoperability and widespread use. MPEG-4 has enjoyed wide acceptance in the research community as well as in commercial development owing to its high bitrate scalability and compression efficiency. Other features of MPEG-4 that makes it suitable for our architecture are:

1) Open source encoders for MPEG-4 are widely available and hence the architecture could be easy to develop, deploy and modify.

2) MPEG-4 standard is divided into number of profiles and levels that makes it operable for wide range of applications, end device and network conditions.

3) MPEG-4 video encoder is capable of compressing video from 5 Kbps to 6 Mbps and hence it's highly scalable.

4) MPEG-4 supports the encoding of arbitrarily shaped objects. Individual objects in an audio-visual stream provide better understanding and enhances overall learning experience.

Each object in animated space is associated with static visual information in the form of MPEG-4 video compressed at various bit-rates. Video encoded with higher bit-rate is of higher quality then the same video encoded at low bit-rate. On receiving request from user for fetching visual information about a particular object in animated space, network is actively probed for available bandwidth. The decision about the quality to video to be delivered to the user is taken on the basis of available bandwidth in the network. In this way network bandwidth can be efficiently utilized and at the same the high quality video can be delivered to the end user.

Streaming video has soft real-time delay constraints and at the same time accuracy of end-to-end available bandwidth estimation is constrained by level of CPU availability and large number of context switches at either network end points [18]. Hence in our architecture, available bandwidth estimation in done at the proxy while streaming server is responsible for video streaming. This approach helps in preventing unnecessary overloading of streaming server.

D. Server Component Architecture

Figure 1 shows the ASkME component architecture and Figure 2 shows the modules architecture of ASkME. Looking at the networking middleware from a system architecture perspective as shown above, we can identify the following basic components in ASkME architecture:

1) Client components: These are various heterogeneous

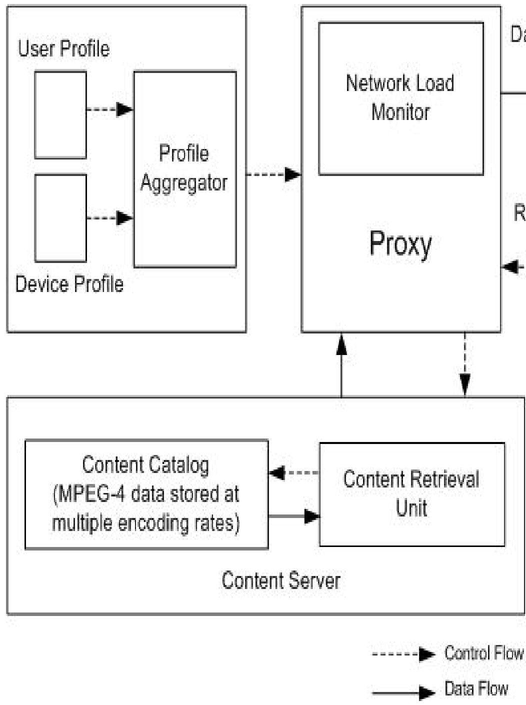

Fig. 2. ASkME Module Architecture

devices (Laptops, PDAs, Tablet PC …) used by users through web services to interact with the smart objects in an animated space environment. All these devices have a RFID reader to sense the smart objects (marked with RFID tags) and a networking card to connect to the ASkME server. Client program does the initial filtering of objects identifier read by RFID reader based on the User Interest Code (UIC). UIC is generated by ASkME server depending on user interest profile.

2) Search Engine: The architecture also contains a search engine to provide more information about any animated objects. The users may get highly immersed in some objects and would like to know more about them. The search engine gets more information about the internet through APIs provided by various search engine (e.g. Google) and makes it available to the interested users.

3) Administrative server: This server manages Authorization, Authentication, and Accounting details. It also deals with the maintenance of user preferences and device profile. The user-preferences and device profile stored in this server will allow users the flexibility to modify the contents of the data being delivered as per their preferences and their device capability.

4) The Proxy server: This is the intermediary acting on behalf of the user. All the data traffic sent from and to the client devices flow through the Proxy. The proxy uses object-oriented component model [19], [20], [21] to provide information to heterogeneous devices depending on their interest profile and device profile to avoid retrieving and storing same content many times. The idea is based on the well known model-view concept to maintain different views of the same content. It also maintains the session-state

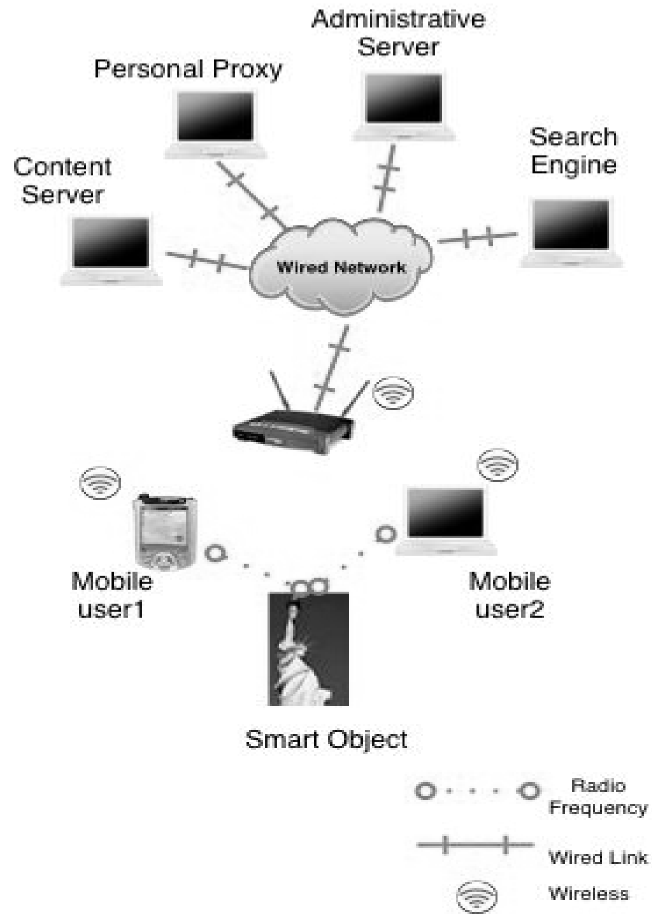

Fig. 3. ASkME Testbed

information of various communication sessions. Network Load Monitor inside the proxy (see Fig. 3) periodically collects the network load information by sniffing aggregate data flowing through the network. Depending on the network load, device and user profile information Proxy makes an informed decision about the encoding rate at which the audio-visual information should be streamed to the mobile users [22]. For e.g.: if there is congestion, the server will reduce its sending rate and the video encoding rate to a level the network can accommodate. Although a decrease in the video bitrate produces image of coarser resolution, it is not nearly as detrimental to the perceived video quality as inconsistent, start-stop play out.

5) Content Server: This is source of data for animated space. The information about smart objects is available in various formats and learning levels. Content Server is made up of Content Retrieval Unit (CRU) and Content Catalog. Content Catalog stores the same video sequence encoded in MPEG-4 at several different compression levels. CRU is responsible for sending multimedia content of specified bitrate from content catalog to client device through proxy.

TABLE 1
Decision Rule Set for Adaptive Content Delivery

	Device and Network Profile (input)			**Multimedia Attributes (output)**			
	End-device processing capability	Device resolution	Network Utilization	Video encoding rate (bit-rate)	Video frame rate	MPEG-4 complexity level (profile)	Video resolution
Case 1	Low (<400 MHz)	320x240	Low (<10%)	High (>300 Kbps)	Low (< 15 fps)	Simple profile, Level 0	176x144
Case 2	Medium (<800 MHz)	640x480	Low (<10%)	High (>300 Kbps)	High (>22.69 fps)	Simple profile, Level 1	320x240
Case 3	High (>800 MHz)	1024x768	Medium (<50%)	Medium (>100Kbps and <300 Kbps)	Medium (>15 fps, <22.69 fps)	Advanced simple profile, Level 1	320x240
Case 4	High (>800 MHz)	1024x768	High (>50%)	Low (<100 Kbps)	High (>22.69 fps)	Advanced simple profile, Level 1	320x240

III. IMPLEMENTATION AND TESTBED

The framework described above, is developed on desktops/laptops running the Linux operating system, and is coded in Java. The popularity of Java as a programming language for mobile devices [23] encourages its adoption as the language for development of this test bed. The schematic layout of the test bed is shown in Figure 3.

All real-world objects are marked with RFID tags with a unique identifier on it. The address space of RFID tags are classified into different groups on the basis of various topics. For e.g.: the first 4 bits for all the monuments are 1001 and the monuments older than 5 years have the identifier starting with 100101. This allows fast filtering of RFID tags depending on user profile and also helps in indexing the object information. There is a database server which provides the mapping from RFID tags to real world objects and stores information about the objects.

User's mobile devices are equipped with IEEE802.11a WLAN cards to connect to ASkME Proxy Server and have a RFID reader to read the smart objects in the real world. Initially, each user creates a login and his user and device profile with the proxy server through the web services running on it. Depending on user profile and device profile, the proxy rule parser generates a User Interest Code (UIC) and rule set for the multimedia content (audio and video) corresponding to the user. The user interest code (UIC) and multimedia rules are stored at administrative server for each user. This information are also stored on user's devices as cookies in which case users are not required to login every time it connects to ASkME server. Cookies serve as a facility for servers to send information to a client. This information is then housed on the client, from which the server can later retrieve the information. All the rules are defined in a XML [24] format with set of attributes and values pair. After profiling is done, there is ASkME client running on user mobile device which reads the tag identifiers read by the attached RFID reader and does the initial filtering of objects IDs using the UIC. The ASkME client shows the smart objects of user interest in a web browser with which user can communicate.

The ASkME server uses Apache Tomcat (Jakarta-tomcat-4.1.30) and Axis Engine (Apache Axis 1.1) for hosting web services. The Apache Xerces parser (Xerces 2.6.2) is used for parsing XML content. When a user connects to ASPEN proxy server, the HttpSession variable is initialized with the user's UIC and multimedia content rules. Network monitoring has been an extensively researched area with many open source monitoring tools widely available [25], [26], [27]. We are using nettimer [25] inside Proxy for probing for the end-to-end bottleneck link bandwidth upon request from the client for delivering video. Audio-Visual information is compressed in MPEG-4 at various bit rates using Xvid [28] and is stored in the Content Server. MPEG-4 content is streamed on demand through Darwin streaming server [29] installed at the content server. Darwin streaming server uses RTSP [30] for real-time streaming of hinted MPEG-4 video stream. MPEG-4 video is rendered in the web-browser of the client through QuickTime plug-in. The different server components and their functions are explained in details in Section IV.

Table 1 enlists typical cases showing the ways in which variety of video encoding parameters is related to various network and device constraints. It shows that for thin-clients like PDAs whose processing speed is low (case 1), MPEG-4 simple profile with level 0 is chosen whose encoding and decoding complexity is less. Also resolution of the video depends on device characteristics. Videos with high bit-rates are streamed if enough bandwidth is available (case 1 and 2).

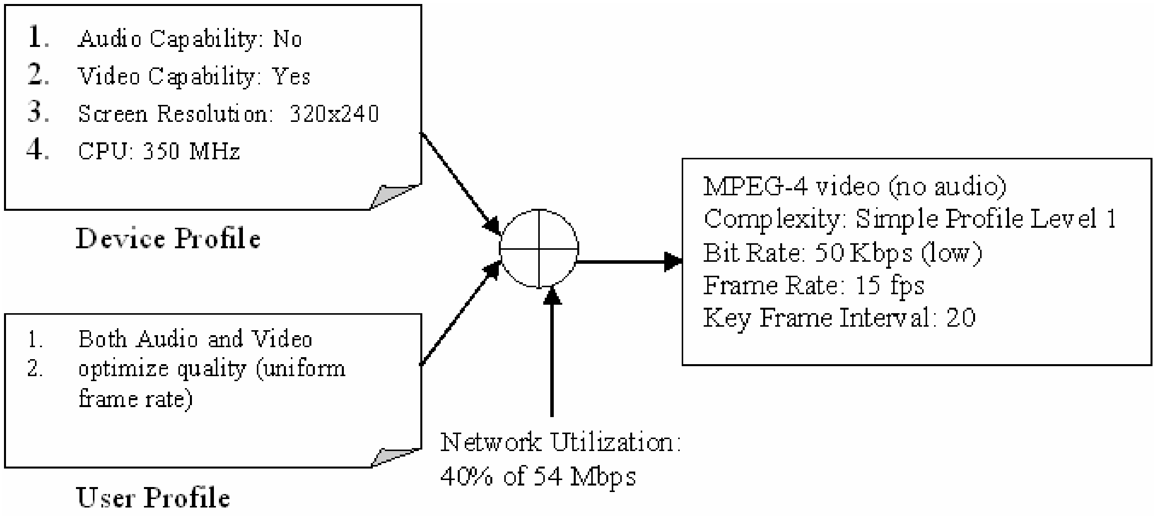

Figure 5 Typical Rule Set for a user

The figure 5 shows the typical rule-set which translates device profile, user profile and network feedback into final MPEG-4 video properties. These rule-sets are defined inside the content server, which receives network traffic report from proxy server and makes decision about the appropriate version of the video to stream on the basis of these rule-sets.

IV. Conclusions

In this paper, we presented a new form of space, *Animated Space,* in which real world objects can communicate with users about their purpose, function and history. We designed and implemented an architecture for adaptive audio-video delivery depending on user profile and network characteristics using RFID technology and MPEG-4. We also developed a testbed for evaluating the feasibility and usability of the animated space.

In the future, we plan to integrate other type of multimedia data (text and image) and design unified real-time content adaptation architecture using an object-oriented approach for multimedia content delivery to heterogeneous devices in wireless networks. We also plan to include the ability to maintain discussion groups among users interested in the same types of information and enable the storage of any sights that a user might contribute to a shared discussion.

References

[1] A. Ferscha, “Awareness in Mobile Learning Teams”, Universität Linz, Institut für Praktische Informatik.

[2] A. Ferscha, “Wireless Learning Networks”, E-learning, 7.11.2002, Wien.

[3] S. Hsi, M. Spasojevic, “Learning in Informal Settings using Nomadic Inquiry”, The Exploratorium, DIMI Workshop, January 17, 2002.

[4] M. Resnick, and N. Rusk, “The Computer Clubhouse: Preparing for life in a digital world”, IBM Systems Journal, 1996

[5] M. Spasojevic, and T. Kindberg, “A Study of an Augmented Museum Experience”, Hewlett-Packard Laboratories.

[6] C. Dede, Vignettes about the Future of Learning Technologies. 2020 Visions: Transforming Education and Training through Advanced Technologies. Washington, DC: U.S. Department of Commerce, 2002, Available: http://www.ta.doc.gov/reports/TechPolicy/2020Visions.pdf

[7] C. Dede, “Creating Research Centers to Enhance the Effective Use of Learning Technologies. (Testimony to the Research Subcommittee”, Science Committee, U.S. House of Representatives, May 10th, 2001). Available: http://www.house.gov/science/research/reshearings.htm

[8] International Organization for Standardization. Overview of the MPEG-4 Standard, Dec. 1999

[9] The MPEG home page. Available: http://mpeg.telecomitalialab.com/

[10] Rodney A. Brooks, “The Intelligent Room Project”, *In Proceedings of the Second International Cognitive Technology Conference (CT'97)*, Aizu, Japan, August 1997.

[11] Steven Feiner, Blair MacIntyre, and Tobias Hollerer, “A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment.”, In *Proc ISWC ‘97 (Int. Symp. on Wearable Computing), Cambridge, MA, October 13–14, 1997, pages 74–81.*

[12] Steven Feiner, and Tobias Hollerer, “Situated Documentaries: Embedding Multimedia Presentations in the Real World.”, *1999 IEEE Proceedings of ISWC ’99 (International Symposium on Wearable Computers),* San Francisco, CA, October 18–19, 1999, pp. 79–86

[13] Brad Johanson, Armando Fox, and Terry Winograd. “The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms”, *IEEE Pervasive Computing*, Volume 1, Issue 2 (April 2002).

[14] S. Shafer, "The New EasyLiving Project at Microsoft Research," Joint DARPA/NIST Smart Spaces Wksp., July 30--31, 1998, Gaithersburg, MD, USA

[15] RFID Technology. Available: http://www.rfidjournal.com/

[16] EAN.UCC White Paper on Radio Frequency Identification – Draft for Approval – November 1999

[17] P/ Bahl, and V. N. Padmanabhan,, “RADAR: An In-Building RF based User Location and Tracking System”, *Proceedings of IEEE INFOCOM 2000*, Tel-Aviv, Israel, March 2000

[18] M. Jain and C. Dovrolis, "End-to-end available bandwidth: measurement methodology, dynamics, and relation with TCP throughput," Tech. Rep., University of Delaware, Feb. 2002

[19] M. Gaedke, C. Segor, and H.-W. Gellersen, “WCML: Paving the Way for Reuse in Object-Oriented Web Engineering”, *2000 ACM Symposium on Applied Computing (SAC 2000)*, Villa Olmo, Como, Italy, March 19-21, 2000.

[20] M. Gaedke, “Web Content Delivery to Heterogeneous Mobile Platforms”, University of Karlsruhe, Germany, 1998.

[21] D. Schneider, “Next Generation Content Management & Delivery: Architecture of the Coremedia & ATG Integration”, ATG Partner Update, 2001
[22] A. Balk, D. Maggiorini, M. Gerla, M. Y. Sanadid, “Adaptive MPEG-4 Video Streaming with Bandwidth Estimation”, *QoS-IP 2003*, Milano, Italy, Feb 2003.
[23] S. Marti, Master's thesis, “Active Messenger: Email Filtering and Mobile Delivery”, MIT Media Laboratory, Speech Interface Group.
[24] World-Wide Web Consortium. XML: eXtensible Markup Language. Available: http://www.w3c.org/XML/
[25] K. Lai and M. Baker, "Nettimer: A Tool for Measuring Bottleneck Link Bandwidth", *Proceedings of the USENIX Symposium on Internet Technologies and Systems*, March 2001.
[26] Tcpdump: Network Traffic Analyzer. Available: http://www.tcpdump.org
[27] Network Performance and benchmarking. Available: http://www.netperf.org
[28] Home of the XviD codec. Available: http://www.xvid.org
[29] Darwin Streaming Server. Available: http://www.opensource.apple.com/projects/streaming/
[30] RFC 2326. IETF: Available: http://www.ietf.org