Etel: A Newspaper-Based Distributed Information System

Michel Banâtre, Valérie Issarny, Frédéric Leleu
Inria / Irisa
{banatre, issarny, leleu}@irisa.fr
Campus de Beaulieu, 35042 Rennes Cédex, France

Position paper

1 Introduction

The ever-growing number of on-line multimedia services relating to application fields as diverse as commercial, entertainment, and educational fields calls for distributed information systems that help the end-user accessing available services. Most popular information systems, also qualified as resource discovery services, are those provided in the Internet such as the World-Wide Web (Www), which is becoming the dominant Internet Resource [4]. Other systems include those based on the use of private resources, which are aimed at ensuring qualitative access to the information.

This paper discusses the design of a new distributed information system, called Etel (Edition TELématique), that combines the advantages of the above approaches, i.e., we want to provide a system that gives access to a large number of services without sacrificing the services' quality. This paper is structured as follows. The next section outlines approaches to distributed information systems and is followed by an overview of the Etel system in section 3. Section 4 addresses our work towards ensuring access quality in the Etel system. Finally, we conclude in section 5, summarizing our work and giving the project status.

2 Approaches to Distributed Information Systems

As mentioned in the introduction, two main approaches to distributed information systems can be identified depending on whether they rely on the Internet or on private resources. These are evaluated in the next two paragraphs.

The Internet approach

The Www combined with an adequate navigator/browser tool, allows the end-user to access the resources he/she is looking for. However, the searching process is burdened by the ever-increasing popularity of the Internet, whose performance is questionable at peak hours, and the relevance of the information varies from one server to another. The latter aspect can be alleviated through the access to a WWW server (e.g., MecklerWeb [9]) that provides a selection of existing WWW servers based on moderation by full-time experts and subscriptions by the registered servers' providers [11]. On the other hand, the former aspect is still an open issue (e.g., see [13]); it relates to ensuring Quality of Service (QoS) including security, availability, and responsiveness, in the Internet.

Without contest, systems based on the Internet have the advantage of having a huge user base as well as providing a wide diversity of information services. However, this is at the expense of the quality of the offered services. At this time, there is no guarantee about security, responsiveness or relevance of the information.

The proprietary approach

In order to guarantee qualitative access to on-line services, proprietary information systems using private resources and specific networks such as cables or telephone lines, have been developed (e.g., CompuServe [7], America On Line [2], Prodigy [12]). Due to the offered QoS, these systems provide access to various commercial services such as banking and are accessible upon subscription.

Compared to the Internet-based systems, proprietary ones ensure a higher degree of quality at the expense of the size of the concerned population. Moreover, the volume of information available to users is negligible compared to the Internet. This has led proprietary systems to provide access to the Internet for which there is no guarantee of quality.

The ETEL approach

Based on the weaknesses of the two above approaches to distributed information systems, we have been designing a new distributed information system, called ETEL, in order to offer access to a large information base with a high degree of QoS. QoS criteria of information systems can be subdivided into two groups depending on whether they relate to the content's quality or to the access' quality. From the standpoint of the former aspect, our goal is to guarantee the relevance of the provided information, which is customized according to the user's profile. The access quality depends upon numerous factors including responsiveness, availability, and scalability. Our approach to ensure QoS (be it content- or access-related) of distributed information systems relies primarily on a newspaper-based solution and is further detailed in the next section.

3 Overview of the ETEL System

Newspaper editors have a long experience in providing qualitative information according to various criteria including ethic, correctness, and presentation. This has led us to base the ETEL information system on a newspaper-like approach. More precisely, the entry point of ETEL is an electronic newspaper service from which the user has access to the usual content of a newspaper as well as to a set of plugged-in interactive multimedia services that are either

Figure 1: The ETEL Architecture

The newspaper service

Our goal in the design of an electronic newspaper service that is developed in close collaboration with a local newspaper editor, is to provide a service with the following features:

- Coupled production of paper and electronic editions, i.e., production of the two versions from the same data,
- Presentation of the information that combines the advantages of both the paper version (e.g., layout) and the electronic support (e.g., interactivity), and
- Integrated view of newspaper-based information and links to various services.

From the standpoint of version production, our solution relies on the subdivision of data structures into two groups: (*i*) the logical data structure embedding the core information that is common to both electronic and paper editions, and (*ii*) the physical data structure that is specific to each type of editions and that describes how core information are laid out. Using these data structures, we are able to build automatically electronic editions from the data used for the construction of paper versions.

Electronic editions are built using the PDF format. More precisely, the construction of the electronic editions relies on the use of two software tools:

(1) XPress that is aimed at press editing, for the generation of the editions' pages in the PDF format, from logical and physical data structures describing the editions' contents, and

(2) Acrobat for enriching the editions' PDF-formatted pages with hyperlinks, and for the presentation of the so-built editions on the users' displays.

Figure 2: A Summary Page

Compared to the paper version, the electronic edition differs through the provision of edition customization with respect to the user's profile (e.g., see [3]). An electronic edition is organized as a set of pages where each page embeds either a summary or an article. A summary page corresponds to a page of the paper version except that every article is summarized into the article's title and photograph (if any), and a hyperlink to the corresponding article. Every summary page further integrates customized hyperlinks to thematic columns according to the user's profile (see figure 2).

Finally, the electronic support allows to revise the presentation of general information such as advertising, TV programs, or cinemas. Instead of having a static paper-like presentation, this information is produced through interactive, multimedia services.

Plugging-in interactive multimedia services

In addition to the presentation of articles, the electronic newspaper service gives access to a set of interactive multimedia services. As mentioned above, these include ETEL-provided services derived from the general information embedded in usual newspapers. The system further offers encyclopedia-like services that exploit the information base of the newspaper editor. In particular, such services allow end-users to query for customized thematic newspaper-like documents.

Another type of offered services results from the integration of a gateway to the Internet via the WWW. However, in order to guarantee the quality of the provided information, ETEL provides only access to a set of WWW servers, selected by the newspaper editor. The proposed selection of WWW servers raises a set of open issues due to the possible nesting of servers through hyperlinks. We have to ensure that any access to a WWW server originating from ETEL, is ETEL-guaranteed. We are currently investigating a solution towards that goal.

4 Towards Ensuring Access Quality in the ETEL System

Up to this point, we have discussed our approach for ensuring information quality in distributed information systems. Those systems have additional requirements relating to the access quality such as responsiveness, scalability and availability. Meeting such requirements calls for system mechanisms that together guarantee the availability of system resources for handling each of the user requests in time. Basically, these mechanisms deal with the management of replication and caching so as to balance the requests' load and with resource reservation so as to guarantee timeliness.

System support for ensuring access quality

The ETEL framework allows us to revise base mechanisms aimed at ensuring the access quality. First, user requests amount to read accesses, which simplifies replication and caching management since it is not burdened with coherency issues. Second, resource reservation leads to resource wastage due to the media discontinuity, i.e., the system's resources are not used while the user is reading the information on his/her display. Finally, from the standpoint of the newspaper service, the service's features allow us to devise service-specific mechanisms that exploit the access patterns of users. Based on these premises, we have been undertaking a new approach towards ensuring access quality. This relies on:

- The provision of caching policies that are based on the anticipation of user accesses with respect to the users' reading behaviors, and

- Implicit resource reservation at the network level through the use of ISDN.

In the following, we introduce the basics of anticipation-based cache management. This is discussed in the framework of the electronic newspaper service. The extension to plugged-in services is an area for future research.

Anticipating user accesses

The information of an ETEL electronic edition is organized according to an invariant structure that is defined with respect to thematic columns. Such a structure can then be represented as a directed graph where:

- Each node represents a thematic column, and
- Each edge corresponds to a hyperlink between two thematic columns within the document.

As editions are customized, the directed graph can be reduced when defining a user edition; a user's graph contains only accessed nodes. Furthermore, a user's graph can be enriched by weighting the edges with the frequency of access based on the selections of hyperlinks that were performed during the user's previous sessions. The weighted graph then allows to forecast the editions' pages that will be accessed next by the user, and hence to anticipate requests to the ETEL system accordingly. However, the direct use of the edges' weights to anticipate requests can be misleading when a user changes his/her reading behavior; the value of an access frequency may take some time prior to reflect the actual reading behavior. In order to alleviate this problem, we use statistical control charts and more specifically the EWMA (Exponentially Weighted Moving Average) scheme [8]. This scheme enables to detect rapidly when a behavior deviates from the one expected, and to modify the edges' weights so as reflect this change.

Anticipation-based cache management

Our approach for anticipating user requests allows us to optimize cache management both on the side of user terminals and on the side of the ETEL system.

At the level of the user cache, our goal is to guarantee the service responsiveness by bounding response time for every page access. In order to do so, a user's cache is managed with respect to the user's weighted graph. This allows to anticipate user requests for pages to the ETEL system, and hence to prefetch corresponding data.

From the standpoint of the ETEL system, our goal is to guarantee the system's scalability together with limitating the number of servers needed to handle the requests of thousands of users. In that framework, scalability is evaluated with respect to ensuring the service responsiveness while the number of users increases. The knowledge of both user profiles and user reading behaviors enables to group users that have similar access patterns. Such groups further characterize groups of data, which are distributed among the ETEL system's servers.

In that framework, a user issues his/her requests to the server that manages caching of the user group's data.

We are currently evaluating the proposed cache management with respect to the service responsiveness and scalability through simulation using the Qnap tool. Data used for characterizing users reading behaviors result from a reader poll that was performed by the local newspaper editor participating to the project. Corresponding mechanisms will be integrated in the Etel prototype. In addition to responsiveness and scalability criteria, another access quality criterion relates to the service availability in the presence of failure. In the framework of another project aimed at the construction of a video on demand server, we are designing fault-tolerance mechanisms dedicated to multimedia applications [5]. We are currently considering the use of such mechanisms for the Etel system.

5 Conclusion

Existing distributed information systems subdivide into two groups depending on whether they provide access to information over the Internet or they are proprietary. The former has the advantage of providing access to a huge information base at the expense of poor quality in terms of information value and responsiveness. On the other hand, the latter guarantees a certain quality of service to the end-users but reduces significantly the information base.

This paper has outlined the design of a new distributed information system, called Etel, that is aimed at combining the advantages of the two existing approaches. Our solution is newspaper-based: the system offers access to a set of Etel- and Www-provided services from a newspaper service presenting the users with a customized, integrated view of the newspaper information together with links to available services. Furthermore, the provided information, be it issued by the newspaper or by the connected services, is subject to moderation by the newspaper editor. Finally, the system responsiveness relies on the use of an Isdn network between end-users and the Etel system together with caching management based on the individual reading behaviors. Much work has been undertaken in the operating system community to address the issues of responsiveness, scalability, and availability in distributed information systems. Solutions are proposed in the framework of both proprietary services such as Interactive TV systems (e.g., see [10]) and Internet services (e.g., see [6, 1]). However, to our knowledge, the design of mechanisms aimed at ensuring quality of service that are based on individual access behaviors has not been examined. From our point of view, taking into account individual behaviors will allow to approximate more closely the resources that are actually needed to serve QoS-guaranteed user requests.

The design of the Etel system has begun in October 1995. A prototype demonstrating the automatic construction of electronic editions, and the presentation of the editions on the user's display is now operational. This prototype will be extended by integrating our caching mechanisms for ensuring quality of service. We are also examining solutions to guarantee quality of service when accessing Internet-provided services via Etel.

Acknowledgments: The authors would like to acknowledge Jean-Paul Routeau and Michel Le Nouy for their work on the implementation of the ETEL prototype. They also thank Boris Charpiot that participates to the ETEL design.

References

[1] R. Carter A. Bestavros and M. Crovella. Application-Level Document Caching in the Internet. Technical Report BU-CS-95-002, Boston University, College of Liberal Arts, Computer Science Department, Boston, MA, USA, 1995.

[2] America 0n Line. http://www.aol.com.

[3] W. Bender, H. Lie, J. Orwant, L. Teodosio, and N. Abramson. Newspace: Mass media and personal computing. In *Proceedings of the Usenix 1991 summer Conference*, pages 329–348, 1991.

[4] H. Berghel. The client's side of the World-Wide Web. *Communications of the ACM*, 39(1):30–40, 1996.

[5] M. Billot, V. Issarny, I. Puaut, and M. Banâtre. A proposal for ensuring high availability of distributed multimedia applications. In *Proceedings of the Fifteenth IEEE Symposium on Reliable Distributed Systems*, 1996.

[6] A. Chankhunthod and M. J. Schwartz. A hierarchical Internet object cache. In *Proceedings of the ACM* SIGCOMM *'94 Conference*, 1994.

[7] Compuserve. http://www.compuserve.com.

[8] J. M. Lucas and M. S. Saccucci. Exponentially weighted moving average control schemes: Properties and enhancement. *Technometrics*, 32(1):1–30, 1990.

[9] Mecklermedia Corp. http://www.mecklerweb.com/demo.html.

[10] M. N. Nelson, M. Linton, and S. Owicki. A highly available scalable ITV system. In *Proceedings of the Fifteenth ACM Symposium on Operating Systems Principles*, pages 54–67, 1995.

[11] L. Press. Commercialization of the Internet. *Communications of the ACM*, 37(11):17–21, 1994.

[12] Prodigy. http://www.prodigy.com/.

[13] W3 Consortium. http://www.netins.net/mega/ww-menu.html.