

Delay Tolerant Applications for Low Bandwidth and Intermittently Connected Users: the aAQUA Experience

Saurabh Sahni
Indian Institute of Technology Bombay, India
saurabh@it.iitb.ac.in

Krithi Ramamritham
Indian Institute of Technology Bombay, India
krithi@cse.iitb.ac.in

ABSTRACT

With the explosive growth and spread of Internet, web access from mobile and rural users has become significant. But these users face problems of low bandwidth and intermittent Internet connectivity. To make the benefits of the Internet reach the common man in developing countries, accessibility and availability of the information has to be improved. aAQUA is an online multilingual, multimedia agricultural portal for disseminating information from and to rural communities. Considering resource constrained rural environments, we have designed and implemented an offline solution which provides an online experience to users in disconnected mode. Our solution is based on heterogeneous database synchronization which involves only a small synchronization payload ensuring an efficient use of available bandwidth. Offline aAQUA has been deployed in the field and systematic studies of our solution show that user experience has improved tremendously not only in disconnected mode but also in connected mode.

Categories and Subject Descriptors

H.3.5 [Online Information Services]: Web-based services

General Terms

Performance, Design, Human Factors

Keywords

Offline Access, Caching, Low-bandwidth Application, Information and Communication Technologies for Development, Heterogeneous database synchronization, Resource constrained low end PCs

1. INTRODUCTION

Information is key to development. Collecting and disseminating information relevant to rural folk is a daunting task for several reasons. Large sections of the society do not have access to the huge knowledge base acquired by scientific development over the centuries. Keeping the factors and the needs of Indian rural users in mind, a database-backed agri-portal has been developed as part of the research activities at the Developmental Informatics Lab, IIT Bombay. aAQUA [2], which stands for *almost All QUestions Answered*, provides a unique solution to this problem by allowing content to be built ground-up in a locally meaningful

and contextually rich manner. Along with content, another challenge faced in disseminating information to rural population is connectivity - it is intermittent, and even when available, the effective bandwidth is unpredictable.

Thus, when a farmer comes to a kiosk either he has to be told to come back later since Internet connectivity is not available or he gets connected to aAQUA but faces inordinate delays while interacting with the aAQUA server. What will be ideal is if the farmer is able to interact with aAQUA as though he is connected via a high speed link. This is precisely what offline aAQUA is designed for. aAQUA portal (www.aaqua.org) includes a web-based forum which is asynchronous and delay tolerant. The users do not expect an immediate response to their queries but response must be received within a reasonable delay, currently 24 hours. Exploiting this fact, we design offline aAQUA to provide better responsiveness to users.

2. OFFLINE BROWSING ALTERNATIVES

Most of the proposed and implemented techniques for offline browsing are based on caching pages by fish search [1]. For caching a website, starting page is downloaded first and then links from the downloaded page are recursively retrieved. The key drawback of this approach is the repetitive download of common fragments of different pages that leads to wastage of network resources and disk space. Further, a single thread addition updates several pages in aAQUA [4], making fish search based caching unsuitable.

Homogeneous database replication [4] is another model for offline access where a mirror of the website and the database is ported on a machine nearer to end users which are synchronized whenever connectivity is available. It improves the response time but deployment of such a system will require dedicated high performance systems.

Rhea et. al. proposed a technique called *value-based caching* [3] that eliminates redundant data transfer by breaking contents of a page into blocks and repetitive requests for the same blocks are served by sending the hash of those blocks. This technique improves response time but requires a proxy to store soft state for each client.

A forum can be made offline by using a RSS or a news reader. RSS will require all common fragments to be sent with every feed, while a news reader will be able to receive only new forum posts. However, we require a solution which allows all pages at aAQUA portal including polls, aAQUA keyword browser, etc. to be available offline efficiently.

Key approaches listed above are compared in Table 1 and properties which make a given approach unsuitable for offline aAQUA are italicized.

| Requirement | DB Replication | Fish search based caching | Value based caching | RSS/news-group | Heterogeneous db sync |
|---------------------------------|----------------|---------------------------|---------------------|----------------|-----------------------|
| Offline client system resources | High | Medium | Low | Low | Low |
| Security | Low | High | High | High | High |
| Source site modification | Medium | Nil | Nil | Low | Low |
| Size of offline repository | Small | Huge | Medium | Medium | Medium |
| Synchronization payload | V.Low | High | Medium | Medium | Low |
| Stateless | Yes | Yes | No | Yes | Yes |

Table 1: Comparing caching approaches

3. OFFLINE AAQUA'S ARCHITECTURE AND PERFORMANCE

Offline aAQUA allows users to navigate and search through all aAQUA threads, forums and other pages without any connectivity. The searching and browsing experience is very fast as the content is stored on user's computer itself, thereby suffering no network delays. The local cache or repository of offline aAQUA can be updated whenever Internet connectivity is present. The updates sent to the users are incremental, only "deltas" are transferred between the clients and the server. Users can also post an update in aAQUA forum in offline mode. The update is saved on the user's machine and sent to aAQUA server whenever the client node connects to it.

The offline aAQUA architecture (Figure 1) is based on what could be termed as **heterogeneous database synchronization**. A subset of server's database is stored on the offline client. On the server, the database stores complete post, metadata and all other required information for every thread. On the other hand, client stores only metadata of the thread in a light weight database where metadata includes attributes like thread id, thread subject, author name, etc. Storing only metadata allows use of a small size database which can be operated efficiently on a low end machine. Complete post is stored separately in a repository (local cache), thus reducing the load on the database.

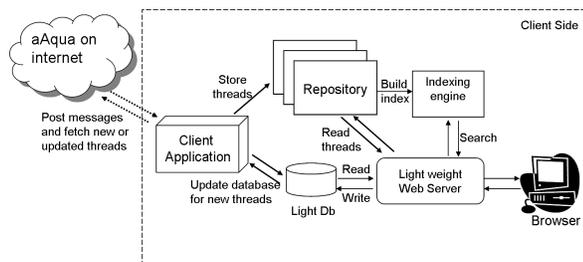


Figure 1: aAQUA offline architecture

To minimize the data transfer overhead, offline aAQUA installation software is shipped with data which will not change with further additions of threads including common fragments, static pages and files, *php* source code for offline web-interface, etc. These *php* pages fetch data from the local database which is updated during synchronization and produce required dynamic content when requested. So the "delta" referred to earlier only consists of metadata about the thread and content of posts excluding common

fragments and files. During synchronization, the server efficiently resolves update and delete conflicts. Using a database in offline aAQUA at the client side avoids the need to send common fragments of the dynamic pages more than once, thereby ensuring that the payload of an update is small. We found the heterogeneous database synchronization approach to be the best for an offline browsing solution for aAQUA (see Table 1), thus chose it for detailed design and implementation.

We analysed 594 *visits* to *www.aaqua.org* over a period of 5 days using logs at server's and client's end to obtain the following results. Table 2 shows the comparison of online and offline aAQUA data transfer and required connectivity duration (assuming data transfer speed of 2.89 KBps) for different tasks. Apart from time spent in data transfer, significant amount of time is spent by end users in viewing the page or typing a post [4] which is included in required network connectivity duration. Other studies [4] indicate that offline aAQUA clearly outperforms all other approaches including value based caching, RSS and newsgroup. After deployment of offline aAQUA at kiosks, Internet connectivity duration required has reduced drastically from about 1-2 hours to just 2-3 minutes per day.

| Task | Total bytes transferred | | Estimated data transfer duration (secs) | | Network connectivity duration (secs) | |
|----------------------|-------------------------|---------|---|---------|--------------------------------------|---------|
| | Online | Offline | Online | Offline | Online | Offline |
| Browsing * | 359855 | 0 | 121.6 | 0 | 335.63 | 0 |
| Creating a new post | 512255 | 12938 | 173.1 | 4.37 | 555 | 4.37 |
| Downloading a thread | 136841 | 14618 | 46.24 | 4.94 | 106 | 4.94 |

* In an average aAQUA session

Table 2: Comparing online and offline aAQUA

4. CONCLUSION

In this paper, we have compared different approaches for users to access the websites in disconnected mode. Offline aAQUA gives them access to aAQUA's website independent of connectivity and lets them synchronize whenever connectivity is present. It has not only increased the availability and reduced the Internet costs but also has improved the browsing experience tremendously.

5. ACKNOWLEDGMENTS

Authors would like to thank Anil, Subhasri, Chaitra, Mansi, Sujith and other DIL members for their support in offline aAQUA development.

6. REFERENCES

- [1] R. Post P. De Bra. Searching for arbitrary information in the WWW: the fish-search for Mosaic. In *Proceedings of the WWW conference*, 1994.
- [2] K. Ramamritham, A. Bahuman, and S. Dutttagupta. aAQUA: A database-backed multilingual, multimedia community forum. In *ACM SIGMOD International Conference on Management of Data*, 2006.
- [3] S. C. Rhea, K. Liang, and E. Brewer. Value-based web caching. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 619-628, New York, NY, USA, 2003. ACM Press.
- [4] S. Sahni and K. Ramamritham. Designing delay tolerant applications for low bandwidth and intermittently connected users: the aAQUA experience. Technical report, IIT Bombay, 2007.
<http://www.it.iitb.ac.in/research/techreport/reports/43.pdf>.