Web Acceleration for E-Commerce Web Applications

Wen-Syan Li   K. Selçuk Candan   Wang-Pin Hsiung   Divyakant Agrawal
Qiong Luo   Wei-Kuang Wayne Huang   Yusuf Akça   Cemal Yılmaz

C&C Research Laboratories, NEC USA, Inc.
110 Rio Robles, M/S SJ100, San Jose, CA 95134, USA
Email:wen@ccrl.sj.nec.com   Tel:408-943-3008   Fax:408-943-3099

ABSTRACT

Response time is one key point of differentiation among electronic commerce (e-commerce) Web sites. For many e-commerce sites, Web pages are created dynamically based on the current state of a business stored in database systems. Snafu and slow-downs during special events or peak times demonstrate the challenges to engineer high performance database-driven e-commerce Web sites. One way to achieve this goal is through dynamic content caching. To ensure the freshness of dynamic content in the caches, it requires integration of caches, Web servers, application servers, and back-end database systems. `CachePortal` is a technology developed at NEC for accelerating database-driven e-commerce Web sites. The technology enables dynamic content caching by (1) deriving the relationships between cached pages and database contents via a *sniffer*; and (2) intelligently monitoring database changes to "eject/delete" related pages from caches via an *invalidator*. Being evaluated and compared with many alternative solutions, `CachePortal` has shown that it can provide many e-commerce applications 3 to 25 times improvement for user response time.

Keywords

dynamic content caching, Web acceleration, invalidation

1. INTRODUCTION

Response time is one key point of differentiation among e-commerce Web sites. Snafu and slow-downs at major Web sites during special events or peak times demonstrate the difficulty of scaling up e-commerce sites. Such slow response times and down times can be devastating for e-commerce sites as indicated in a recent study by Zona Research[1] on the relationship between Web page download time and user abandonment rate. The study shows that only 2% of users will leave a Web site (i.e. abandonment rate) if the download time is less than 7 seconds. However, the abandonment rate jumps to 30% if the download time is around 8 seconds. The abandonment rate goes up to 70% when the download time is around 12 seconds. This study clearly establishes the importance of fast response times to an e-commerce Web site to retain its customers.

One possible solution to this problem is to deploy network-wide caches so that a large fraction of requests can be served remotely rather than all of them being served from the origin Web site. However, since freshness of Web pages is essential to e-commerce practices, for many e-commerce applications, HTML pages are generated dynamically based on the current state of a business, such as product prices and inventory, typically stored in a database.
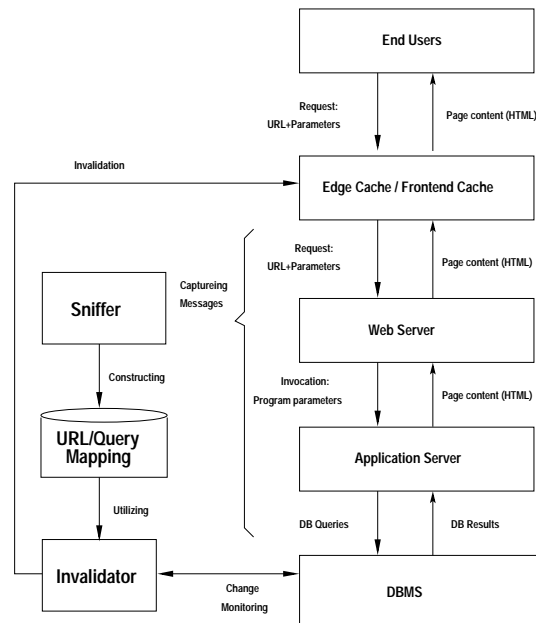

Figure 1: Architecture of a Database-driven E-Commerce Site with `CachePortal` Technology

A typical database-driven Web site (illustrated on the right of Figure 1) consists of a database management system (DBMS), an application server (AS), and a Web server (WS). When a user accesses the Web site, the request and its associated parameters, such as the product name and model number, are passed to an application server. The application server performs necessary computation to identify what kind of data it needs from the database. Then the application server sends appropriate queries to the database or other sources. After the database returns the query results to the application server, the application server uses these to prepare a Web page and passes it to the Web server, which then sends it to the user.

2. PROBLEM STATEMENT

The problem of dynamic content caching has received significant attention recently. Since cache servers, WS, AS, and DBMS are independent components, currently there is no coordination mechanism to ensure that database content changes are reflected to the caches. As a result, most e-commerce sites have to specify dynamic contents as non-cacheable. Consequently, each request to an e-commerce site results in both network delay and server delays (i.e. WS delay, AS delay, and DBMS delay) since the request must be processed each time at the Web site.

Dynamai [2] from Persistence Software is one of the first dynamic caching solution that is available as a product. However, Dynamai relies on proprietary software for both database and application server components. Thus it can not be easily incorporated in existing e-commerce framework. Challenger et al. [3] at IBM Research have developed a scalable and highly available system for serving dynamic data over the Web. The IBM system was used at Olympics 2000 to post sport event results on the Web in timely manner. This system utilizes database triggers for generating update events as well as intimately relies on the semantics of the application to map database update events to appropriate Web pages. Oracle i-cache [4] provides database content caching for application server to generate dynamic pages at remote sites close to users. However, it has two major shortcomings. First, it still requires redundant computation when users make duplicate requests. Second, it is generally very costly to keep multiple database synchronized and up-to-date in the presence of update activity at the databases [5].

3. PROPOSED SOLUTION

CachePortal [6] is a technology developed at NEC for accelerating database-driven e-commerce Web sites. The technology enables dynamic content caching by (1) automatically deriving the relationships between cached pages and database contents (i.e. *URL/Query Mapping*) via a *sniffer*; and (2) intelligently monitoring database changes to "eject" (i.e. delete) related pages from caches via an *invalidator*. The architecture of e-commerce Web sites deploying CachePortal is illustrated on the right of Figure 1. Compared with most existing work, where the relationships between Web pages and underlying data are specified manually, CachePortal features a *sniffer* to automatically generate the query/URL mapping. Compared with existing work where the invalidation checking is implemented using database trigger functions, *invalidator* in CachePortal is implemented as an external software component based on incremental view maintenance techniques, which does not add substantial load to the underlying DBMS. Being evaluated and compared with many alternative solutions, CachePortal has shown that it can provide many e-commerce applications substantial improvement for user response time. We summarize one experimental result as follows.

We tested our technology on an e-commerce Web site which deploys Oracle 8i as the DBMS, BEA WebLogic as the application server and Web servers, and a modified Apache Web server as the cache server. The database used in this experiment contains 7 tables and 1M records total. The update rate is 1 per second and each dynamic content page request results in a query with one join operation to the database. All machines are Pentium III 700 Mhz CPUs running Redhat Linux 6.2 with 1 GByte of main memory.

We recorded both the user response time (i.e. round trip time observed by the users). Our results indicate that the response time for the Web site without CachePortal technology increases sharply when the number of user requests increases. On the other hand, the response time for other configurations with CachePortal remains relatively flat for the settings where the hit ratios are 60%, 80%, and 100% respectively. As a matter of fact, a more meaningful measure is the percentage of the requests that are served under 7 seconds, where the user abandonment rate is less
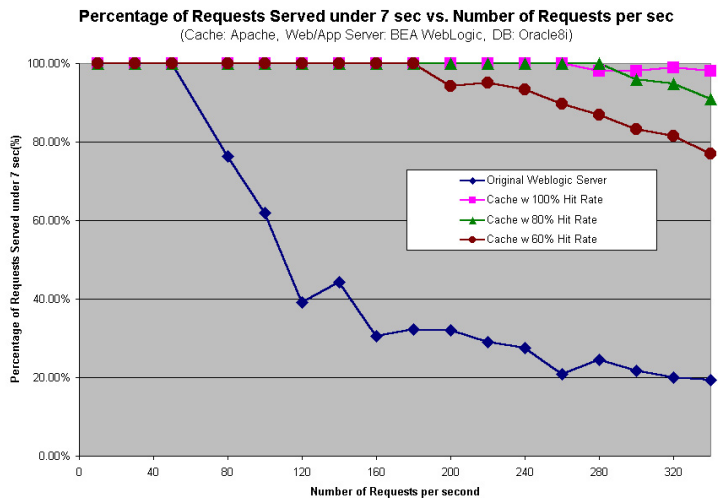

Percentage of Requests Served under 7 sec vs. Number of Requests per sec
(Cache: Apache, Web/App Server: BEA WebLogic, DB: Oracle8i)

Figure 2: Evaluation on Percentage of Response Time under 7 Seconds

than 2 percent[1]. In Figure 2, we see that the system architectures (upper three plots) with CachePortal technology provide good response time most of time whereas the system architecture without CachePortal technology (the lower plot) have poor performance.

4. CONCLUDING REMARKS

It is essential but difficult for e-commerce Web sites to handle the traffic arise at special events or peak times. In this paper, we describe CachePortal technology, which can be loosely coupled with e-commerce site for dynamic content delivery acceleration and cached content invalidation. We are currently incorporating the technology into commercial application servers and are testing the scalability of CachePortal technology on several e-commerce Web sites.

5. REFERENCES

[1] Zona Research, http://www.zonaresearch.com/

[2] Persistence Software,
http://www.persistence.com/dynamai/

[3] Jim Challenger, Arun Iyengar, and Paul Dantzig. *A Scalable System for Consistently Caching Dynamic Web Data*, in Proceedings of IEEE INFOCOM'99, New York, New York, March 1999.

[4] Oracle Corporation, http://www.oracle.com/

[5] Jim Gray, Pat Helland, Patrick E. O'Neil, and Dennis Shasha. The dangers of replication and a solution. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, Montreal, Quebec, Canada, June 4-6, 1996*, pages 173–182. ACM Press, 1996.

[6] K. Seluk Candan, Wen-Syan Li, Qiong Luo, Wang-Pin Hsiung, and Divyakant Agrawal. Enabling Dynamic Content Caching for Database-Driven Web Sites. In *Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data, Santa Barbara, CA, USA*, ACM Press, 2001.