US006351775B1

(12) **United States Patent** (10) Patent No.: **US 6,351,775 B1**

Yu (45) **Date of Patent:** **Feb. 26, 2002**

(54) **LOADING BALANCING ACROSS SERVERS IN A COMPUTER NETWORK**

(75) Inventor: **Philip Shi-Lung Yu**, Chappaqua, NY (US)

(73) Assignee: **International Business Machines Corporation**, Armonk, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **08/866,461**

(22) Filed: **May 30, 1997**

(51) **Int. Cl.**[7] **G06F 15/173**
(52) **U.S. Cl.** **709/238**; 709/239; 709/240; 709/241; 709/242; 370/237; 370/400
(58) **Field of Search** 709/242, 239, 709/240, 241, 238; 370/237, 400

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,371,852 A		12/1994	Attanasio et al.
5,517,620 A	*	5/1996	Hashimoto et al. 709/242
5,526,414 A	*	6/1996	Bedard et al. 379/221
5,544,313 A	*	8/1996	Shachnai et al. 709/219
5,828,847 A	*	10/1998	Gehr et al. 709/239
5,864,535 A	*	1/1999	Basilico 370/231
5,930,348 A	*	7/1999	Regnier et al. 379/221
6,078,943 A	*	6/2000	Yu 709/105
6,091,720 A	*	7/2000	Bedard et al. 370/351

FOREIGN PATENT DOCUMENTS

JP   8-214063   8/1996

OTHER PUBLICATIONS

M. Colajanni et al., "Scheduling Algorithms for Distributed Web Servers", RC 20680 (91683) Jan. 6, 1997, Computer Science/Mathematics, Research Report, 29 pages.

T. Brisco, "DNS Support for Load Balancing", Apr. 1995, 6 pages, Network Working Group, Rutgers University.
Daniel M. Dias et al., "A Scalable and Highly Available Web Server", (not dated), 8 pages, IBM Research Division, T. J. Watson Research Center, Yorktown Heights, N. Y. 10598.
Eri D. Katz et al., "A scalable HTTP server: The NCSA prototype", 1994, pp. 155–164, vol. 27, Computer Networks and ISDN Systems.
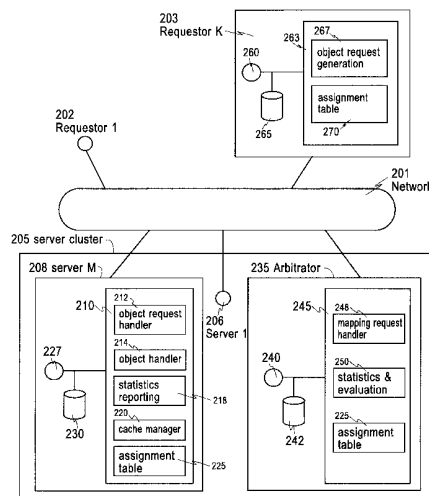
* cited by examiner

Primary Examiner—Krisna Lim
(74) *Attorney, Agent, or Firm*—F. Chau & Associates, LLP

(57) **ABSTRACT**

A dynamic routing of object requests among a collection or cluster of servers factors the caching efficiency of the servers and the load balance or just the load balance. The routing information on server location can be dynamically updated by piggybacking meta information with the request response. To improve the cache hit at the server, the server selection factors the identifier (e.g. URL) of the object requested. A partitioning method can map object identifiers into classes; and requester nodes maintain a server assignment table to map each class into a server selection. The class-to-server assignment table can change dynamically as the workload varies and also factors the server capacity. The requester node need only be informed on an "on-demand" basis on the dynamic change of the class-to-server assignment (and thus reduce communication traffic). In the Internet, the collection of servers can be either a proxy or Web server cluster and can include a DNS and/or TCP-router. The PICS protocol can be used by the server to provide the meta information on the "new" class-to-server mapping when a request is directed to a server based on an invalid or obsolete class-to-server mapping. DNS based routing for load balancing of a server cluster can also benefit. By piggybacking meta data with the returned object to reassign the requester to another server for future requests, adverse effects of the TTL on the load balance are overcome without increasing traffic.
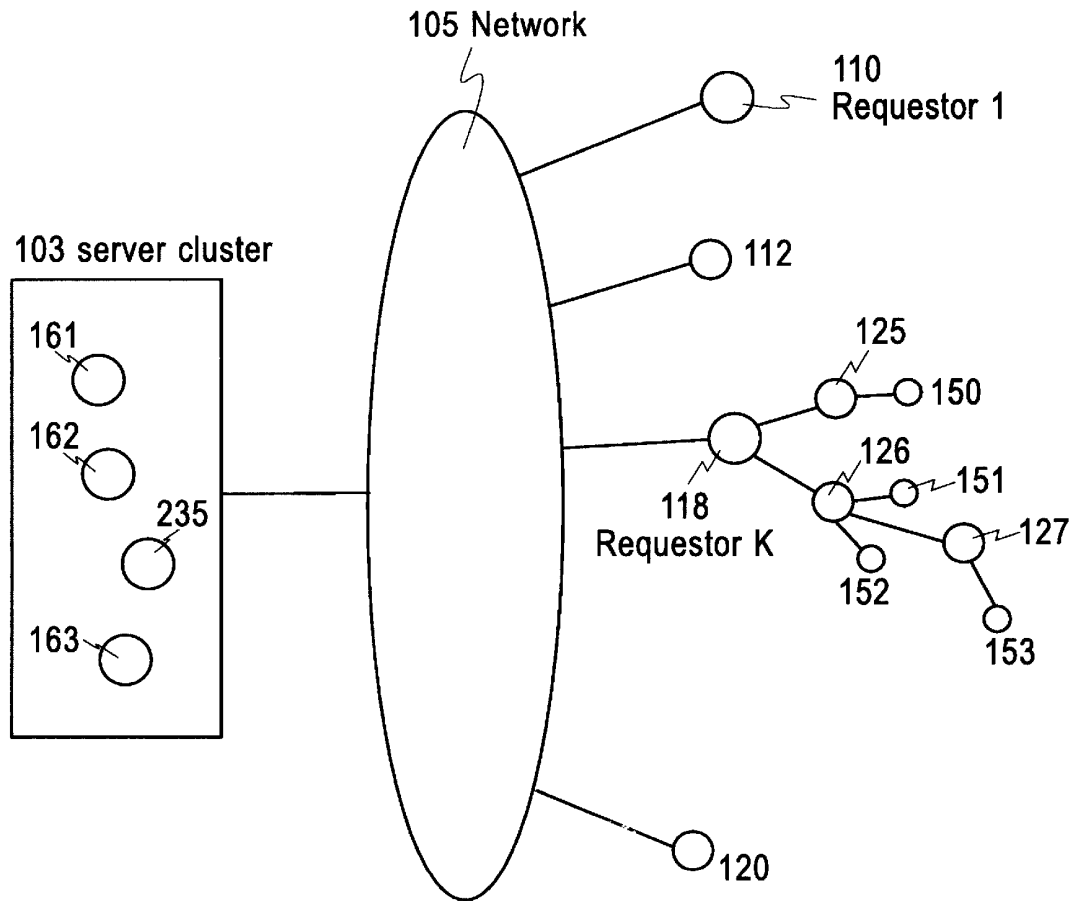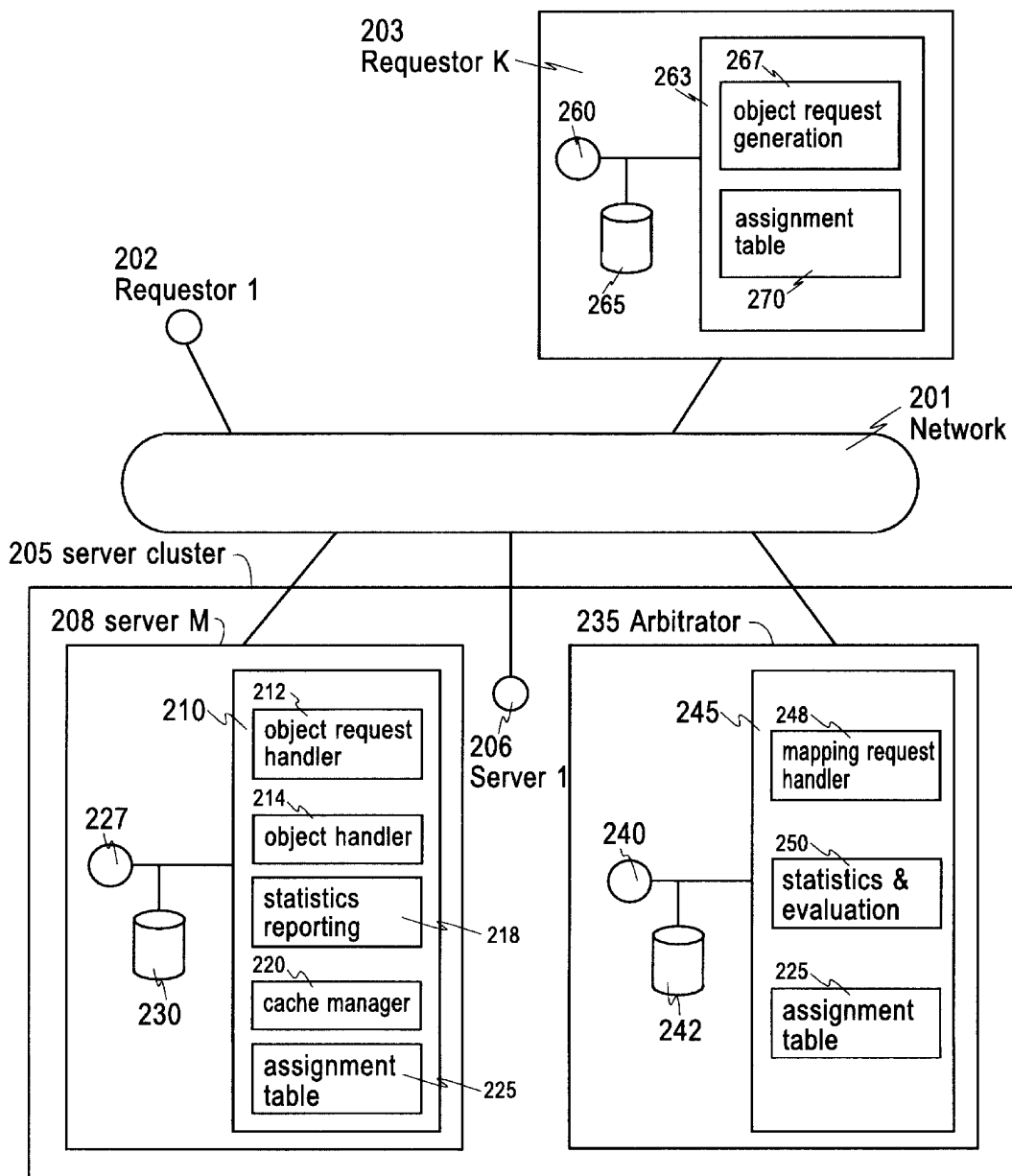
75 Claims, 15 Drawing Sheets

105 Network

110 Requestor 1

112

103 server cluster

161

162

235

163

125

150

126

151

118 Requestor K

127

152

153

120

Fig.1

Fig.2

Assignment Table
N=16,   M=3

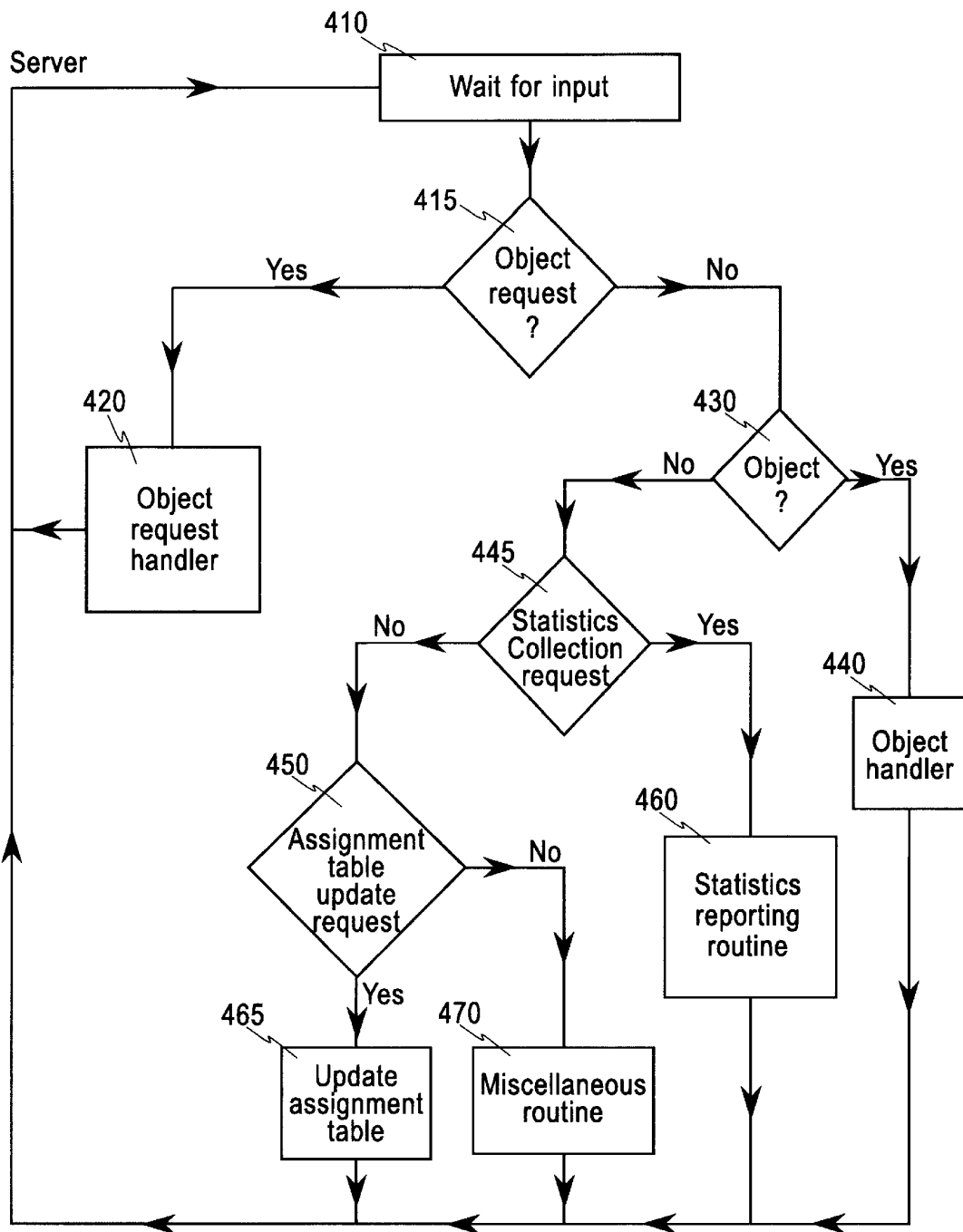| Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Server	2	3	2	2	2	2	2	3	1	3	1	2	2	3	2	3

Fig. 3

Fig. 4

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.

fastcase®
Smarter legal research.