BU-Web-Client

Description

These traces contain records of the HTTP requests and user behavior of a set of Mosaic clients running in the Boston University Computer Science Department, spanning the timeframe of 21 November 1994 through 8 May 1995.

During the data collection period a total of 9,633 Mosaic sessions were traced, representing a population of 762 different users, and resulting in 1,143,839 requests for data transfer.

Format

Trace logfiles contain the sequence of WWW object requests (whether the object was served from the local cache or from the network). Each log file name contains a *user id number*, converted from Unix UIDs via a one-way function that allows user IDs to be compared for equality but not to be easily traced back to particular users. The file name also gives the machine on which the session took place, and the Unix timestamp when the session started. Boston University is located in the United States Eastern Time Zone. For example, a file named *con1.cs20.785526125* is a log of a session from user **1**, on machine **cs20**, starting at time **785526125** (12:42:05 EST, Tuesday, November 22, 1994).

Each line in a log corresponds to a single URL requested by the user; it contains the machine name, the timestamp when the request was made, the user id number, the URL, the size of the document (including the overhead of the protocol) and the object retrieval time in seconds (reflecting only actual communication time, and not including the intermediate processing performed by Mosaic in a multi-connection transfer). An example of a line from a condensed log is:
cs20 785526142 920156 "http://cs-www.bu.edu/lib/pics/bu-logo.gif" 1804 0.484092
Lines with the number of bytes equal to 0 and retrieval delay equal to 0.0 mean that the request was satisfied by Mosaic's internal cache.

Measurement

To collect this data we installed an instrumented version of Mosaic in the general computing environment at Boston University's Computer Science Department. This environment consists principally of 37 SparcStation 2 workstations connected in a local network, which is divided in 2 subnets. Each workstation has its own local disk; logs were written to the local disk and subsequently transferred to a central repository.

We began by collecting data on a subset of the workstations only, while testing our data collection process. This period lasted from 21 November 1994 until 17 January 1995. When we were statisfied that data collection was occurring correctly, we extended the data collection process to include all workstations; data collection then took place until 8 May 1995. Since Mosaic ceased to be the dominant browser in use by early March 1995, the most representative portion of the traces are those covering the period 21 November 1995 through 28 February 1995.

Privacy

The user IDs in these logs have been renumbered to protect privacy.

Acknowledgements

These logs were collected by the members of the [Oceans research group](#) at Boston University. Mosaic was instrumented by Carlos Cunha (*carro@cs.bu.edu*). When referring to the use of these traces in published work, please cite [Characteristics of WWW Client Traces](#), Carlos A. Cunha, Azer Bestavros and Mark E. Crovella, Boston University Department of Computer Science, Technical Report TR-95-010, April 1995.

Restrictions

The traces may be freely redistributed.

Distribution

Available from the Archive as a [gzip compressed tar](#) file (13.8 MB; 101 MB uncompressed).

Up to [Traces In The Internet Traffic Archive](#).