

## ContentKeeper Closes Loopholes in Collaborative Web Filtering

March 4, 2002  
By Lee Badman

Sometimes a fast Internet connection at work is just too good for your users to resist: news, sports, porn, job hunting--the sky's the limit! The need to control Internet activities within the workplace has spawned the rapid growth of the content-filtering industry. With its new ContentKeeper filtering service and appliance, ContentKeeper Technologies has introduced a collaborative filtering system that allows for on-the-fly acceptable-use policy definitions.



Various statistics show that employees spend a significant fraction of their workdays online, indulging in non-business-related matters. These activities also waste bandwidth, reducing company profits and connectivity efficiency. Illicit surfing also can bring harm to the enterprise in the form of leaked corporate data and acquired viruses, such as the Nimda worm. The burgeoning peer-to-peer connections and streaming media that many connected employees have grown fond of come at a very real cost. Those seeking to put the breaks on "bad" traffic should be pleased with ContentKeeper's fresh bag of tricks.

### Global Collaboration

Breaking the mold of filtering processes used by most of its competitors--including Secure Computing Corp.'s SmartFilter and SurfControl's SuperScout Web Filter and CyberPatrol Web Filter--ContentKeeper uses the company's patented Closed-Loop Collaborative Filtering technology. Typical plug-in appliances often play catch-up with proxy server software and operating systems as they evolve--letting an obsolete filtering plug-in disrupt the network, for example. Because ContentKeeper runs on servers configured as adaptive transparent Ethernet bridges, it functions autonomously with no concern for firewalls and router configurations, and it in no way relies on proxy servers. HTTP-only communication with data centers ensures that no protocol squabbles or similar ill dealings will occur between ContentKeeper and customer firewall/proxy server setups, no matter how they may be configured.

Whether purchased as a turnkey system or built with downloadable code on your own Red Hat Linux server, the local ContentKeeper appliance becomes part of a distributed and dynamic system that categorizes and updates both URL and Web page content for millions of pages on an hourly basis. Based in Australia, ContentKeeper Technologies has built a global network of data centers, each servicing a specific geographic area. From each data center, customers' ContentKeepers coordinate to update the data center's enormous database of URLs from which browsing decisions can be made. Current data-center facilities are in Canberra, Australia; London; and Palo Alto, Calif. The company says it plans to add facilities in Frankfurt, Germany, and Hong Kong.

This dispersion of data centers provides not only thorough location-based coverage but also redundancy; connectivity is shipped automatically to the other data centers should any one of

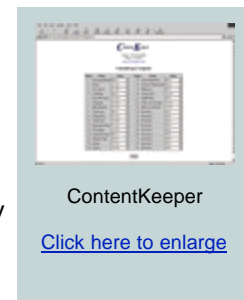
them go down. The data centers update each other and connected clients hourly.

## Packet Dissection

ContentKeeper goes beyond run-of-the-mill URL look-up services, starting with the process whereby each HTTP packet passing through it is checked and dissected to ensure proper classification into one of 32 categories. Real-time packet manipulation in the device works in tandem with the data center, which forms the rest of the collaboration loop.

When a not-yet-classified address is visited, the contents of the page are analyzed by the on-board real-time analysis engine as the traffic goes through the bridge on its way back to the browser. A decision is made within 60 seconds as to which category applies to the previously unknown URL, and the URL is then added to the blocking database. Within the next 60 minutes, the newly classified URL information is sent to the data center for deeper analysis and distribution to the entire ContentKeeper user population.

At the data center, each new URL received from deployed ContentKeeper devices is examined as a complete entity, including page content and all linked pages, by neural-network engines for agreement with the ContentKeepers that did the initial categorization. Should the analysis engines disagree with the real-time findings from the field, data-center experts will review the URL. After the entire process plays itself out (usually in less than an hour), the site information is pushed to the individual ContentKeepers as they synchronize with the database.



## Taking a Spin

From the user's perspective, browsing through ContentKeeper is no different from any other browsing session, at least until the administrative acceptable-use policy is called into play. Based on out-of-the-box and locally configured rules governing 32 site categories (such as politics, porn, shopping, job searching and news), ContentKeeper completely blocks users from a Web site, "coaches" them that they are about to go somewhere inappropriate or allows them to authenticate against the its internal user database or the network's back-end database to proceed to the site.

Exceptions to a given category are simple to build. Each authentication is logged in the detailed usage logs, which become part of the information available in the service's extensive reporting capability. The pages that block, coach or request authentication can be customized to allow for customer-developed pages to go in place of ContentKeeper's stark red, green and blue user-management pages.

The process of setting up and administering the ContentKeeper service is both intuitive and quick. Whether downloading the fully functional evaluation software for your own machine (minimum Intel Corp. 1-GHz CPU, three NICs and 512 MB of RAM with a Red Hat Linux 7.1 OS) or purchasing the whole solution from an Open Systems sales partner, getting the device to run is simple.

During testing in my 10/100-Mbps environment (load-balancing/gigabit configurations are also available), I had my demo appliance running and connected to the Canberra data center in minutes. I also was provided with an optional Shore Microsystems' bypass switch, which installs in parallel with ContentKeeper to keep it from becoming a single point of failure: Should the Ethernet bridge fail, unfiltered Internet connectivity is maintained.

After getting to ContentKeeper's simple-but-effective administrative GUI screen, I built the first of many custom policies for various sites and file types, and dug in on the browsing. I was satisfied with the results; the sites I attempted to access to test my policies were blocked, coached or allowed as per my expectations. After testing multiple policy definitions through Microsoft Internet Explorer 5.5 and 6.0, along with several versions of Netscape (including a quick glimpse through Red Hat Linux) over several days, I became a believer.

The few surprises I found during my testing were more curiosities than problems. For example, when I chose to administratively block all search sites, ContentKeeper barred entrance to all with the exception of [www.google.com](http://www.google.com), which is one of the more heavily used search engines. Despite this oddity and others too inconsequential to mention, the filtering process, including the collaborative filtering, worked well; not only I but other users were blocked from accessing the particular sites.

ContentKeeper will be available through channel partners, including FreeStone Software, Inc. in the United States and ContentKeeper UK. Along with Ethernet/Fast Ethernet and Gigabit Ethernet compatible server appliances, the channel partners will provide one-to-three-year ContentKeeper subscriptions with hourly database updates.

*Lee Badman is an IT analyst and project manager at Syracuse University. Prior to his current position, Lee had a distinguished career with the U.S. Air Force, both maintaining and teaching maintenance of a variety of systems. Send your comments on this article to him at [lhbadman@syr.edu](mailto:lhbadman@syr.edu).*

### Vendor Information

ContentKeeper, \$4,800 for 250-user license. Available: Currently. FreeStone Software, Inc.; 303-398-7016  
[www.freestonesoftware.com](http://www.freestonesoftware.com)