China's eye on the Internet

September 12 2007

The "Great Firewall of China," used by the government of the People's Republic of China to block users from reaching content it finds objectionable, is actually a "panopticon" that encourages self-censorship through the perception that users are being watched, rather than a true firewall, according to researchers at UC Davis and the University of New Mexico.

The researchers are developing an automated tool, called ConceptDoppler, to act as a weather report on changes in Internet censorship in China. ConceptDoppler uses mathematical techniques to cluster words by meaning and identify keywords that are likely to be blacklisted.

Many countries carry out some form of Internet censorship. Most rely on systems that block specific Web sites or Web addresses, said Earl Barr, a graduate student in computer science at UC Davis who is an author on the paper. China takes a different approach by filtering Web content for specific keywords and selectively blocking Web pages.

In 2006, a team at the University of Cambridge, England, discovered that when the Chinese system detects a banned word in data traveling across the network, it sends a series of three "reset" commands to both the source and the destination. These "resets" effectively break the connection. But they also allow researchers to test words and see which ones are censored.

Barr, along with Jed Crandall, a recent UC Davis graduate who is now an

assistant professor of computer science at the School of Engineering, University of New Mexico; UC Davis graduate students Daniel Zinn and Michael Byrd; and independent researcher Rich East sent messages to Internet addresses within China containing a variety of different words that might be subject to censorship.

If China's censorship system were a true firewall, most blocking would take place at the border with the rest of the Internet, Barr said. But the researchers found that some messages passed through several routers before being blocked.

A firewall would also block all mentions of a banned word or phrase, but banned words reached their destinations on about 28 percent of the tested paths, Byrd said. Filtering was particularly erratic at times of heavy Internet use.

The words used to probe the Chinese Internet were not selected at random.

"If we simply bombarded the Great Firewall with random words, we would waste resources and time," Zinn said.

The researchers took the Chinese version of Wikipedia, extracted individual words and used a mathematical technique called latent semantic analysis to work out the relationships between different words. If one of the words was censored within China, they could look up which other closely related words are likely to be blocked as well.

Examples of words tested by the researchers and found to be banned included references to the Falun Gong movement and the protest movements of 1989; Nazi Germany and other historical events; and general concepts related to democracy and political protest.

"Imagine you want to remove the history of the Wounded Knee massacre from the Library of Congress," Crandall said. "You could remove 'Bury My Heart at Wounded Knee' and a few other selected books, or you could remove every book in the entire library that contains the word 'massacre.'"

By analogy, Chinese Internet censorship based on keyword filtering is the equivalent of the latter -- and indeed, the keyword "massacre" (in Chinese) is on the blacklist.

Because it filters ideas rather than specific Web sites, keyword filtering stops people from using proxy servers or "mirror" Web sites to evade censorship. But because it is not completely effective all the time, it probably acts partly by encouraging self-censorship, Barr said. When users within China see that certain words, ideas and concepts are blocked most of the time, they might assume that they should avoid those topics.

The original panopticon was a prison design developed by the English philosopher Jeremy Bentham in the 18th century. Bentham proposed that a central observer would be able to watch all the prisoners, while the prisoners would not know when they were being watched.

The work will be presented at the Association for Computing Machinery Computer and Communications Security Conference in Alexandria, Va., Oct. 29-Nov. 2, 2007.

Source: University of California - Davis