Web Filtering and Censoring

Thomas M. Chen and Victoria Wang,
Swansea University, Wales

Information on the Web is not as uncontrolled as it may appear.

In May 2009, the Chinese government mandated installation of the Green Dam Youth Escort Web filter on all PCs sold in China starting in July. The filter's official purpose was to protect children from pornographic, violent, and other "unhealthy" websites, but it immediately aroused opposition from several quarters.

International security researchers argued that the filter was inaccurate and contained vulnerabilities potentially exposing PCs to security threats. Free speech advocates expressed concern that the government could use the filter to monitor users' online activities and block politically sensitive websites. And the US government urged the Chinese Ministry of Industry and Information Technology and Ministry of Commerce to revoke the Green Dam requirement on the basis of free trade.

In response to the controversy, the Chinese government indefinitely "delayed" the requirement except for PCs used in schools, cyber cafes, and other public access locations.

Around the same time, following the controversial reelection of President Mahmoud Ahmadinejad in Iran, critics accused the regime of blocking certain websites such as Facebook and YouTube—which had been used to post confrontations with the police—as well as sites affiliated with the opposition leader. The Iranian government was also suspected of monitoring Internet usage to track down election protesters.

In January 2010, public attention was again drawn to the issue of Web censoring when Google.cn decided to stop complying with Chinese government requirements to censor search results related to politically and socially sensitive issues. Google reached its decision, considered long overdue by some, in response to attacks by Chinese hackers on Gmail accounts of Chinese human rights activists during the previous month. US Secretary of State Hillary Clinton publicly praised Google and called for a global end to Web censoring, prompting a critical response from the Chinese government.

As a mass medium, the Web is unique. Traditional print media such as books and newspapers require time to publish and distribute physically, while electronic mass media, namely radio and television, typically depend on official licenses for operation and are either state-regulated or self-censored in exchange for fewer regulations. In contrast, the Web lets individuals easily publish a rich range of content—such as multimedia, hyperlinks, and JavaScript—globally in real time to PCs and various mobile devices.

The public perceives the Web as a wildly uncensored medium, the embodiment of complete freedom of expression (for good or bad). In reality, information on the Web is not as uncontrolled as it may appear. According to the OpenNet Initiative (www.opennet.net), more than 40 countries actively filter Web access—selectively blocking content to a billion Internet users worldwide—and many others are considering doing so. These countries use various methods to apply different levels of filtering.

HOW WEB FILTERING WORKS

Web filters differ in complexity, granularity, accuracy, location, and transparency. Generally, simpler Web filters are easy to implement but work at a coarser granularity. Just as importantly, motivated individuals tend to discover and circumvent them more readily.

IP and URL blacklists

The simplest Web filters depend on blacklists of IP addresses. The main advantage of blacklists is speed—essentially a fast table lookup. Speed allows Web filtering at "choke points" in the network where traffic is aggregated such as gateways between neighboring national networks. However, the simplicity of blacklists has two major drawbacks. First, continual blacklist updating requires enormous effort and resources. Second, IP blacklists work at a coarse granularity—the filter either blocks or allows all Web content from an IP address.

Published by the IEEE Computer Society

URL blacklists offer a finer level of granularity. These are often implemented at DNS servers that resolve domain names to IP addresses. When a DNS server receives a resolution request, it checks the URL against the blacklist. If the requested URL is on the list, the DNS server will return an incorrect or default IP address. URL blacklists suffer the same drawback as IP blacklists—maintaining and updating them requires considerable effort.

IP and URL blacklists can be deployed at proxy-based filters. Organizations commonly use proxy servers to cache Web content locally: The proxy cache keeps recently requested content, and if users request the same content again, it is served from the proxy instead of the origin server. Because all Web content goes through the proxy server, it is an attractive location for filtering.

A proxy-based filter checks the IP addresses or URLs in all Web requests against a blacklist. If it detects a blacklisted IP address or URL, the proxy filter can return a "blockpage" with an error message or explanation that the content was blocked. Some nations such as China and Iran go further and block URLs containing prohibited keywords in the URL path. However, keywords in the URL do not necessarily reflect the webpage contents accurately. For example, the word "sex" in a URL would include many other types of websites besides pornographic ones and blocking URLs containing the word would not catch all pornographic sites.

(a)

(b)

Figure 1. Green Dam Web filter. (a) Users or network administrators can choose to block or admit access to five categories of Web content. Source: University of Michigan. (b) Green Dam's image scanner relies on skin color and face recognition to block access to pornographic websites.

Content filtering

Real-time content filtering at a proxy or the Web client offers two important advantages over IP or URL blacklists. First, the filter examines Web contents when a page is requested, without the need to preestablish a blacklist. Second, the filtering decision is based on individual webpages or even elements within a page.

However, a content filter has two difficult and somewhat competing requirements: intelligence to recognize an undesirable page or content within a page, and fast operation. Content filters typically use machine learning or AI techniques to classify webpages into a set of predefined categories, as Figure 1a shows; users or network administrators can choose to block or admit access to each category. Speed is important because people become impatient and leave a website if it returns requested contents too slowly.

Intelligent content filters examine various elements of a webpage for classification, including the metadata, links, text, images, and scripts.

Metadata in a Web document's header can reveal information about authorship and keywords.

Linkage analysis is based on "guilt by association"; pages are likely to be linked to other pages of the same type. For example, a news website is likely linked to other news sites. A page's links can be quite revealing about its subject.

Textual analysis is critical to content filters because many webpages are mostly text. Content filters typically parse the text, identify keywords, and apply machine-learning techniques to decide on the most appropriate category. This approach is not perfect without semantic analysis because it can be difficult to understand different contexts—for example, whether a page with sexual content has educational or pornographic purposes.

Image analysis is vital for classifying pornographic websites, but classification accuracy can be low. Image analysis usually looks for regions with skin color, but this task is complicated by the wide range of human skin colors and different lighting conditions.

Analysis of active content such as JavaScript, VBScript, and ActiveX controls is essential to detect and filter malicious webpages. In particular, JavaScript is frequently used for phishing and drive-by downloads. Unfortunately, heavy obfuscation of the code can make it unreadable, in which case it may be necessary to run the code in a JavaScript interpreter and observe its behavior.

Example: Green Dam

Green Dam is an example of a desktop Web filter that uses both approaches. It first checks a requested URL against blacklists and then scans text for obscenities and politically sensitive phrases.

To filter pornography, Green Dam includes an image scanner that appears to identify regions of skin color and inputs the region characteristics—for example, shape—into a trained support vector machine (SVM) classifier. An additional step attempts to recognize a human face, reportedly using the open source OpenCV image recognition software. If a human face is not the primary component, the filter classifies the image as pornographic and blocks the site, as Figure 1b shows.

Considering the algorithm's heavy reliance on skin color and face recognition, it is not surprising that tests of Green Dam have found the porn filter to be fallible, mistakenly filtering Garfield the cat, Johnny Depp, Paris Hilton, and flesh-colored pork.

Most countries use some form of Web filtering.

STATE-SPONSORED WEB CENSORING

Given the long history of governments trying to control traditional mass media, state-sponsored Web censoring could be expected. However, its prevalence was unknown prior to the OpenNet Initiative.

Filtered content

There is little disagreement that certain Web content is harmful and should be filtered, such as fraud, Web spam, malware, child pornography, and terrorism. State-sponsored Web filtering generally targets four types of content:

- political dissent;
- social deviance—for example, hate speech, pornography, and gambling;
- national security threats—namely, terrorism; and
- certain Internet services such as anonymizers, blogs, peer-to-peer sharing, and social networks.

Government Web filters in China and Iran are considered to be among the strictest, but most countries use some form of Web filtering. Even in the US, which is widely believed to have the fewest restrictions, Web filtering is common at public Internet access spots such as schools and libraries, as well as on private corporate premises.

Example: China

In 2003, the Chinese Ministry of Public Security initiated the Golden Shield Project, also called the "Great Firewall of China," a comprehensive network to monitor and control access to both domestic and foreign websites. All foreign traffic enters China at three choke points—international gateways or Internet exchange points (IXPs): Beijing-Qingdao-Tianjin in the north, from Japan; Shanghai on the central coast, also from Japan; and Guangzhou in the south, from Hong Kong. A small amount of traffic also enters by satellite, but satellite connections are slow and expensive.

The two main backbone providers, China Netcom and China Telecom, use IP blocking at the international gateways. The backbone providers also filter on the basis of URLs containing certain keywords in the URL path. Through DNS tampering, DNS queries for these URLs may return a "site not found" error.

While most countries depend solely on IP or URL filtering, China also exercises content filtering via its state-licensed ISPs. Content filters purchased largely from foreign security companies examine packet payloads for forbidden content—for example, sensitive keywords such as "democracy" or "64" (for June 4, the day of the Tiananmen Square incident in 1989). The filters may cut connections to websites with prohibited contents by means of TCP resets,

and then temporarily block further connection attempts to the same IP address. Prohibited websites within China require official registration and may be shut down.

Web filtering is one tool in the broader monitoring and censorship program. It is believed that the Chinese government employs a large cyber police force of mostly university students to continually inspect websites such as blogs and discussion forums. They search for prohibited content, especially political dissidence; participate in online discussions to influence public opinion; identify and track down the authors of subversive content; and pressure the public to report on individuals engaging in prohibited online activities. A system of electronic access cards deployed in Internet cafes identifies users before they can go online. Perpetrators are subject to fines, job dismissal, or imprisonment.

Web censoring in China involves an extra twist due to its lack of transparency. In contrast, countries such as Saudi Arabia publish their Web censoring policy and rationale. A user who requests a prohibited website is presented with a blockpage explaining that the site is disallowed, as Figure 2 shows. Citizens can submit requests for URLs to be blocked or allowed.

No one is quite certain of the extent of Web censoring in China. The level of censoring appears to vary with political events. Users may see connection errors but no blockpages with clear explanations. In addition, content filters based on sensitive keywords may block a webpage one day but allow access the next day if the page contents have changed.

Consequently, Chinese users are uncertain if a website's inaccessibility is due to deliberate filtering or random congestion in the Internet. This unpredictability actually makes Web censoring more effective because people cannot learn how the filters are working and thus are unsure how to circumvent them.

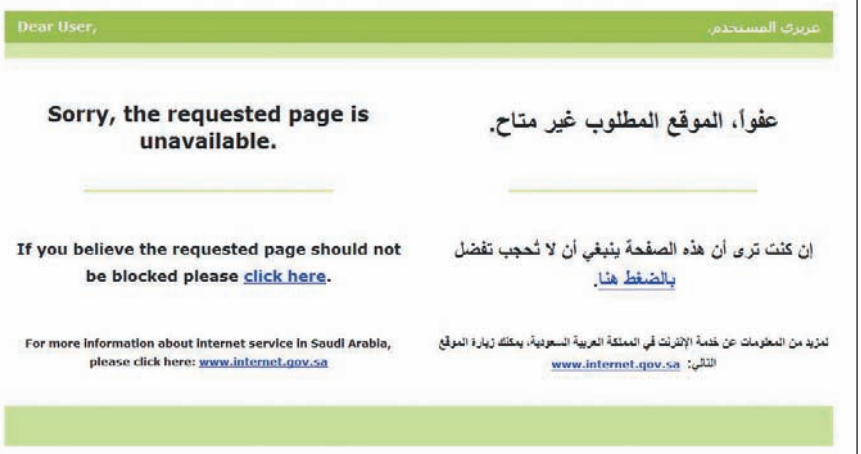

Figure 2. State-sponsored Web censoring. In many countries—in this case, Saudi Arabia—a user who requests a government-prohibited website is presented with a blockpage explaining that the site is disallowed.

FILTER CIRCUMVENTION

Given the Chinese government's effective covert Web censoring program, its mandate for the desktop Green Dam filter was puzzling. Although the filter would have been installed on every PC, it could easily be disabled.

Sufficiently motivated individuals can circumvent a Web filter—for example, they can bypass a URL blacklist by going directly to the server's IP address. In the case of China, Golden Shield does not attempt to read or censor any encrypted traffic such as used in virtual private networks (VPNs). Virtually all foreign businesses in China depend on VPNs, and blocking them would unacceptably impact commerce.

Chinese citizens also use various proxy tools, including Garden Networks' GTunnel (www.gardennetworks.org), Psiphon (www.psiphon.ca), and UltraSurf (www.ultrareach.com). These products use a combination of proxies, encryption, and onion routing to offer anonymized and uncensored Web access.

If Web filtering is easy to circumvent, then what purpose does it serve? In the case of China, the apparent goal is to impose significant obstacles on the country's 350 million Internet users to discourage the majority from accessing certain foreign websites and thereby keep public attention on tightly regulated domestic sites. Only a small minority have the technical know-how to find ways around government censorship.

For those concerned with the prevalence of Web censoring, the Herdict Web project (www.herdict.org) collects reports of websites around the world that are inaccessible to many people and thus a possible sign of censorship. As Thomas Jefferson once wrote, "The price for freedom is eternal vigilance."

Thomas M. Chen *is a professor of networking in the School of Engineering at Swansea University, Wales. Contact him at t.m.chen@swansea.ac.uk.*

Victoria Wang *is a postdoctoral researcher at the Centre for Criminal Justice and Criminology in the School of Human Sciences at Swansea University. Contact her at 257387@swansea.ac.uk.*

Editor: Simon S.Y. Shim, Dept. of Computer Engineering, San Jose State Univ., San Jose, CA; simon.shim@sjsu.edu