

12

THE INTERNET TREATS CENSORSHIP AS A MALFUNCTION AND ROUTES AROUND IT?

A New Media Approach to the Study of State Internet Censorship

Richard Rogers

THE WEB AS A SET OF DISCRETE SITES?

The research approach described here is a contribution to the study of state Internet censorship. It seeks to move beyond the dominant treatment of the Web as a set of discrete sites, which are blocked or accessible. Here the Web is considered to be an information-circulation space. In a sense, a conceptualization of the Web as circulation space as opposed to a set of discrete sites is more of a new media than old media starting point.

In an old media way of thinking, there are, say, single books that are censored, just as there are now single sites. There may be types of books, or types of sites, that are censored (e.g., dating, religious conversion, or human rights). But if censorship research work is considered from a new media perspective, the methods, techniques, as well as the research output may change.

On the Internet, part of a single site may have circulated, and that content may be available elsewhere. The information on sites that are censored may be syndicated, and fed by RSS, or it may have been scraped, in an automated or semi-automated form of copying and pasting. Additionally, snippets of censored content may also have been grabbed, and subsequently annotated, commented on, or similar, for example in the blogosphere. That

“new media apparatus” may be available. Finally, there may be “related sites” and “related content”—related because they are in surfers’ topical paths. (Alexa provides such “related sites.”) Thus, single sites may be censored but portions of the same or related content, and its apparatus, may be unblocked.

Revealing the unblocked content shifts the focus of the work from the analysis of single sites to that of information circulation. It also shifts the research away from the policies of the censor to the Web knowledge and skills of the censored site owner. For example, site owners cognizant of censorship have been known to change their domain names repeatedly, striving to keep a step ahead of filtering software and censor’s blacklists. The day-to-day competition between the censor and the censored is not so unlike that between search engine companies and search engine optimizers. The optimizer, like the censored, is striving to find out whether the new sleight of hand that keeps the information in the right space has been discovered.

Demonstrating the techniques of circulatory forms of censorship circumvention has implications for both censors as well as the censored. For example, the filtering software companies subscribe to proxy list providers’ notifications. Proxies are machines serving as gateways, and are used by surfers in censored countries (among others) to have a different geographical (Internet provider [IP]) point of entry to the net. (They also are used by censorship researchers to check sites in countries known to censor the Internet. One connects to the Internet in Iran (through an Iranian proxy), and fetches sites in order to see the connection statistics, and/or to capture screen grabs of blocked sites. Censors and filtering software companies also make use of proxy lists, adding them to their blacklists. Just as filtering companies may subscribe to alerts from proxy list providers, censors could pull in site feeds, query them in engines, and refresh the blacklist according to the engine returns.

URL LISTS AND INTERNET CENSORSHIP RESEARCH

One of the more comprehensive (and open source) blacklists of sites is coupled with the Dans Guardian filtering software, listing some 56 categories of sites blocked (at urlblacklist.com) from “kids time-wasting” to “weapons” (see Table 12.1). There is also the ability to register both suggestions for blocking as well as complaints about blocked sites. A well-known filtering application in the proprietary arena, SmartFilter by the Secure Computing Corporation, advertises 73 categories of blocked sites (see Table 12.2). In the past filtering companies’ lists have been cracked, and circulated, leading to great consternation about the editorial skills and orientations of the list-makers.

TABLE 12.1
URL Black List Categories and Descriptions for the
Dans Guardian Open Source Filtering Software, 23 March 2007

CATEGORY	DESCRIPTION
Ads	Advert servers and banned URLs
Adult	Sites containing adult material such as swearing but not pornography
Aggressive	Similar to violence but more promoting than depicting
Anti-spyware	Sites that remove spyware
Artnudes	Art sites containing artistic nudity
Banking	Banking Web sites
Beer liquor info	Sites with information only on beer or liquors
Beer liquor sale	Sites with beer or liquors for sale
Cell phones	stuff for mobile/cell phones
Chat	Sites with chat rooms, etc.
Child care	Sites to do with child care
Clothing	Sites about and selling clothing
Culnary	Sites about cooking et al.
Dating	Sites about dating
Dialers	Sites with dialers such as those for pornography or trojans
Drugs	Drug-related sites
E-commerce	Sites that provide online shopping
Entertainment	Sites that promote movies, books, magazine, humor
French education	Sites to do with French education
Gambling	Gambling sites, including stocks and shares
Gardening	Gardening sites
Government	Military and schools, etc.
Hacking	Hacking/cracking information
Home repair	Sites about home repair
Hygiene	Sites about hygiene and other personal grooming-related information
Instant messaging	Sites that contain messenger client download and Web-based messaging sites
Jewelry	Sites about and for selling jewelry
Job search	Sites for finding jobs
Kids time wasting	Sites kids often waste time on
Mail	Web mail and e-mail sites
Naturism	Sites that contain nude pictures and/or promote a nude lifestyle

TABLE 12.1
URL Black List Categories and Descriptions
for the Dans Guardian Open Source Filtering Software,
23 March 2007 *(continued)*

CATEGORY	DESCRIPTION
News	News sites
Online auctions	Online auctions
Online games	Online gaming sites
Online payment	Online payment sites
Personal finance	Personal finance sites
Pets	Pet sites
Phishing	Sites attempting to trick people into giving out private information
Porn	Pornography
Proxy	Sites with proxies to bypass filters
Radio	Non-news-related radio and television
Religion	Sites promoting religion
Ring tones	Sites containing ring tones, games, picture, etc.
Search engines	Search engines such as Google
Sexuality	Sites dedicated to sexuality, possibly including adult material
Sports news	Sports news sites
Sports	All sports sites
Spyware	Sites that run or have spyware software to download
Update sites	Sites where software updates are downloaded from, including virus sigs
Vacation	Sites about going on vacation
Violence	Sites containing violence
Virus infected	Sites that host virus-infected files
Warez	Sites with illegal pirate software
Weather	Weather news sites and weather-related
Weapons	Sites detailing with or selling weapons
Web mail	Just Web mail sites
White list	Contains site specifically 100% suitable for kids

Source: urlblacklist.com

TABLE 12.2
SmartFilter (Rich feature-set, March 23, 2007)

FILTERING OPTIONS

73 individual categories of Web sites
 Both URL and IP addresses
 http and https traffic
 File type (jpg, MP3, etc.)
 Granular key word searches/search engine key word blocking
 Time of day
 Day of week
 Default filtering policies available

FILTERING ACTIONS

Group users or workstations under a common policy
 Deny, allow, warn, but allow, exempt, delay, or report only
 Authorized override—authorized users can bypass the filter
 for a specified amount of time
 Global block/allow

FILTERING CUSTOMIZATION

500 user-defined categories
 Create unique filtering response message for end users
 Add, delete, or exempt sites from categories
 Pattern matching: build dynamic rules for granular custom filtering

Leading researchers of Internet censorship have had a similar point of departure. Until recently, the work has been devoted to building a global list of URLs, with some 37 categories in all. Once the lists are in place, the censorship researchers fetch the URLs through a browser in each of the countries under study (see Tables 12.3 and 12.4). As an initial check, proxy servers located in countries that censor the Internet may be used. If the http return codes are 403 (forbidden) or 504 (server gateway time out), the sites are tagged as suspected blocks. (Other http return codes may provide indications of censorship.) Researchers on the ground subsequently check each URL (suspected or otherwise). Lists are made of blocked sites, per category, across the set of countries under study. Country levels of censorship by site category (with specific lists of blocked URLs) constitute a main research

output. State censorship policy is described, as are the censorship techniques (e.g., gateway time outs in China), including the identification of particular software packages in use (e.g., SmartFilter in Saudi Arabia).

So far the main thrusts of Internet censorship research have been described, also in the context of filtering software more generally—list creation, URL fetching, and http return code monitoring. Now, I describe the means by which one may contribute to the creation of URL lists, and gradually fill in the notion of new media Internet censorship research, with its emphasis on the Web as a circulation space. In particular, I describe three Internet censorship research techniques: related site dynamic URL sampling (URL list-making with hyperlink analysis), redistributed content discovery (through key word searching, key phrase parsing, and additional searching), and surfer re-routing (through route map-making).

TABLE 12.3
Open Net Initiative's Categories in the Global URL List for State
Internet Censorship Research

Alcohol	Humor
Anonymizers	Major events
Blogging domains	Medical
Drugs	Miscellaneous
Dating	News outlets
E-mail	Person-to-person
Encryption	Porn
Entertainment	Provocative attire
Environment	Religion (fanatical)
Famous bloggers	Religion (normal)
Filtering sites	Religious conversion
Free webspace	Search engines
Gambling	Sexual education
Gay/lesbian/bisexual/transgender/ queer issues	Translation sites
Government	Terrorism
Hacking	Universities
Hate speech	Weapons/violence
Human rights	Women's rights
	Voice over Internet protocol

TABLE 12.4
Open Net Initiative's Country List
for State Internet Censorship Research

ASIA AND SOUTH ASIA	LATIN AMERICA
Burma	Cuba
China, Hong Kong	Venezuela
India	
Malaysia	MIDDLE EAST AND AFRICA
Maldives	Afghanistan
Nepal	Algeria
North Korea *	Bahrain
Pakistan	Egypt
Singapore	Eritrea
South Korea	Ethiopia
Thailand	Iran
Vietnam	Iraq
	Israel
EASTERN AND CENTRAL ASIA	Jordan
Azerbaijan	Libya
Belarus	Morocco
Kazakhstan	Oman
Kyrgyzstan	Saudi Arabia
Moldova	Sudan
Russia	Syria
Tajikistan	Tunisia
Turkmenistan	United Arab Emirates
Ukraine	Yemen
Uzbekistan	Zimbabwe

Related Site Dynamic URL Sampling

The current method in Internet censorship research for compiling the global list of URLs is editorial. For an initial URL trawl, directories may be used, such as Yahoo's, Google's or Dmoz.org's. Subsequently, country experts are consulted, and URLs of interest only for one or more particular country are collected. These are the so-called high-impact sites, such as opposition par-

ties. Generally speaking, between 1,000 and 2,000 URLs are checked per country. However, Julian Pain, head of the Internet Freedom desk at Reporters Without Borders, has indicated that the quantity of sites censored in particular countries may be much greater. In Saudi Arabia, “the official Internet Service Unit (ISU) is proud to tell you it’s barred access to nearly 400,000 sites and has even posted a form online for users to suggest new websites that could be blocked.”⁶²⁵

In its own form of a new media style (user-generated content), Saudi Arabia, like urlblacklist.com, “crowd-sources” URLs to bring to the attention of the ISU, using the many-eyes approach over the assumingly few eyes of the censors. If there are 400,000 sites being censored, however they are all sourced, and the Internet censorship researchers are checking only some 2,000, questions arise. How should URLs be sourced? How should the list be made more sizeable? An important consideration concerns the people on the ground in each of the countries who fetch the URLs on the lists through browsers. The time it takes to run the lists may be considerable; care also needs to be taken for personal security reasons. Thus the additional URLs put on the list to be checked should be vetted for relevance.

In a post-directory era, where in Google the directory is no longer a main tab (and three clicks away) and in Yahoo no longer the default search engine, relevance follows from counting links, and boosting sites either through freshness (in a pagerank style) or through votes (in a user ratings style). Here, initially, the link-counting strategy is employed, where a set of sites point to other sites to which they collectively link. Using the URL and site-type data furnished by the Internet censorship researchers, I crawled one category of sites in one country—the “political, social and religious” sites on the Iranian list. The sites’ hyperlinks (external links) are harvested, and co-link analysis is performed, where those sites with two links from the initial list of sites are retained. Once the network of interlinked sites is found, all the sites are cross-checked with the Internet censorship researchers’ lists of known blocked sites, ascertaining which sites are already known blocks. All newly discovered sites are fetched through proxies in Iran, in order to ascertain their status. The result is a map showing political, religious, and social sites blocked and unblocked in Iran, with pins indicating newly discovered blocks (see Fig. 12.1). Of particular interest is the case of the British Broadcasting Service (BBC). The Internet censorship researchers had the BBC news homepage on its list of sites to check (<http://news.bbc.co.uk>). The link analysis turned up a deep page on the site, the BBC’s Persian language page (<http://www.bbc.co.uk/persian>). In Iran, the BBC news page is accessible, as the researchers had found, but the Persian-language page is not. In all some 37 censored sites were newly discovered through what we termed a dynamic URL sampling method, which relied on an analysis of hyperlinking for related site relevance as opposed to the editorial process—directories and experts.

REDISTRIBUTED CONTENT DISCOVERY

Research into state censorship in Pakistan has found, among other things, that two groups seeking autonomy (the Balochi and the Sindhi) have their sites routinely blocked. The Internet censorship researchers have lists of blocked sites for the two groups, one of which (the Balochi) served as starting points for the URL discovery method just described—the crawling of sites, the link analysis, and the proxy checking. With two newly discovered censored sites added to the list through a hyperlink analysis, the overall question concerns the extent to which the blocked content has been redistributed to sites that are not blocked in Pakistan. The case study concerns the killing by the Pakistan military of the Baloch tribal leader, Nawab Akbar Khan Bugti. A special Google query for “Nawab Akbar Khan Bugti,” which excludes known blocked sites in Pakistan, shows some 900 results. (Google only serves up to 1,000 results per query.) The teaser texts of the returns are analyzed for unique phrases, and sorted by date (see Fig. 12.2). When listed chronologically, from June to October 2006, the parsed phrases appearing before and after Nawab Akbar Khan Bugti tell a story.

The following “story” describes of the death of “Nawab Akbar Khan Bugti,” the Baloch tribal leader, from parsed Google (teaser text) returns, June 26 to October 12, 2006. Baloch-authored content, not blocked by Pakistan Internet censorship, appears in italics.

He’s 80 years old, but Nawab Akbar Khan Bugti, a feudal lord in Pakistan’s rugged Baluchistan province, wants to fight to the death.

The irony was that Nawab Akbar Khan Bugti served to help the federal government when he was appointed as Governor of Balochistan by Mr. Zulfikar Ali Bhutto

“Nawab Akbar Khan Bugti was directly attacked. Luckily he survived all attacks and is safe,” said Khan, rejecting rumours that Akbar Bugti’s grandson

have claimed to have killed Nawab Akbar Khan Bugti, one of the founding fathers of the Baloch independence struggle, and 36 other freedom-fighters

The martyrdom of Nawab Akbar Khan Bugti is a loss for Pakistan and a gain for Baloch nationalist movement

It was the third attempt on the life of Nawab Akbar Khan Bugti. After the interception of satellite phone communication, the Nawab’s location was pin

LEAKY CONTENT: AN APPROACH TO SHOW BLOCKED CONTENT ON UNBLOCKED SITES IN PAKISTAN – THE BALOCH CASE.

Pakistan censors Websites related to Balochistan.

How to find blocked content on unblocked sites? A case study related to the killing of a Baloch tribal leader.

Step one:

Query Baloch-related sites known to be blocked in Pakistan for "Nawab Akbar Khan Bugti"

http://oldmanclub.persianblog.com	http://www.balochclub.com
http://balochestan.com	http://www.balochistaninfo.com
http://balochwarna.org	http://www.balochistaninfo.com/balochanitawar
http://www.baloch2000.org	http://www.balochmedia.net
http://www.balochfront.com	http://www.balochitawar.net
http://www.ostomaan.org	http://www.balochunitedfront.org
http://balouch.blogspot.com	http://www.balochvoice.com
http://dochebaloch.persianblog.com	http://www.bso-na.org
http://ngaran.blogfa.com	http://www.eurobaluchi.com
http://www.payambaloch.persianblog.com	http://www.zrombesh.org
http://www.rahimjaandehvari.blogfa.com	http://www.hazzaam.com
http://www.radiobalochi.org	http://www.balochunity.org
http://www.sarbaaz.com	

Step two:

Retain teaser text from Google results, and retrieve phrases appearing on more than one site. Examples of unique phrases obtained from teaser text:

He's 80 years old, but Nawab Akbar Khan Bugti, a feudal lord in Pakistan's rugged Baluchistan province, wants to fight to the death.

The irony was that Nawab Akbar Khan Bugti served to help the federal government when he was appointed as Governor of Balochistan by Mr Zulfikar Ali Bhutto

Step three:

Query Google for those phrases from blocked Baloch sites in Pakistan. Exclude known blocked sites from query in order to find the same content on other sites.

Sample query:

site:oldmanclub.persianblog.com "have claimed to have killed Nawab Akbar Khan Bugti, one of the founding fathers of the Baloch independence struggle, and 36 other freedom-fighters"

Step four:

Verify (through Pakistani proxies) that the newly found sites containing Bugti-related phrases are accessible in Pakistan.

Nawab Akbar Khan Bugti was directly attacked. Luckily he survived all attacks and is safe, said Khan, rejecting rumours that Akbar Bugti's grandson

Step five:


Distinguish between Baloch authored and non-Baloch authored content.

In a statement on the first Sabbath after the martyrdom of Nawab Akbar Khan Bugti, Shaheed-4-Balochistan and former governor and chief minister of

Step six:

Show blocked content on sites accessible in Pakistan, with a timeline of the story of the killing of Bugti from Baloch and non-Baloch sources. Resize phrases according to frequency of mentions.

ABOUT THE BALOCH CASE



The Baloch are an Iranian people inhabiting the region of Balochistan in Iran and Pakistan as well as neighboring areas of Afghanistan and the southeast corner of the Iranian plateau in Southeast Asia.

The Baloch were designated by the British as a "martial race." Martial race is a designation created by officials of British India to describe "races" (peoples) that were thought to be naturally warlike and aggressive in battle... Some of the peoples designated by the British as belonging to a martial race: Baloches, Chakmas, Jats, Rajputs, Awans, Gujars, Pashtuns, Marathas and Gurkhas.

Balochistan is Pakistan's largest province, and is said to be the richest in mineral resources. It is a major supplier of natural gas to the country.

On 15 June 2006, an estimated 600 fighters, led by three commanders, agreed to lay down their weapons after talks with Shoaib Nazharwanji, Balochistan's minister for internal affairs, in Dera Bugti district. On August 26, Balochistan tribal leader Nawab Akbar Khan Bugti was killed by Pakistan Military in an operation designed to kill off opposition to Pakistan military.

Source: en.wikipedia.org (Website blocked in Pakistan)

Image by: Fotokrom.com (user: art) and OpenStreetMap. Accessibility: Richard Dwyer and Erik Brink. Design: G3: Alan Davidson.

© 2006 Georgetown, and 2008

FIGURE 12.2.

Nawabzada Hyrbair Marri on Monday rejected government's claims that Nawab Akbar Khan Bugti had died because of the collapse of his cave hideout

Nawab Akbar Khan Bugti buried in Balochistan without the presence of his family

Nawab Akbar Khan Bugti in a military operation, prominent Baloch leaders and Pakistani human rights activists said it spelt doom for the country's unity and

In a statement on the first Sabbath after the martyrdom of Nawab Akbar Khan Bugti, Shaheed-i-Balochistan and former governor and chief minister of

Baloch Nationalist leader Nawab Akbar Khan Bugti, who was murdered by Pakistani military

Balochistan, Nawab Akbar Khan Bugti, the highest elected official to be killed by the Pakistan Army. Since March 27, 1948 when Balochistan was forcibly

The killing of Baloch leader Nawab Akbar Khan Bugti in August 2006 sparked riots and will likely lead to more confrontation. The conflict could escalate if

In fact, Nawab Akbar Khan Bugti has lighted the candle

But I wonder why journalists, brought in on a military helicopter to witness Nawab Akbar Khan Bugti being buried by a dozen common labourers, couldn't ask

blooded murder of their great leader Nawab Akbar Khan Bugti

Nawab Akbar Khan Bugti had played a significant and controversial role in Pakistani

Speaking at a condolence reference for the late Nawab Akbar Khan Bugti at the Hyderabad press club under the aegis of Sindh National Party (SNP),

tensions have increased since the killing of a veteran nationalist politician, Nawab Akbar Khan Bugti, in a military offensive in August.

The status of Nawab Akbar Khan Bugti, the octogenarian chieftain of a tribe in the restive southwestern province of Balochistan, almost reached the mythical

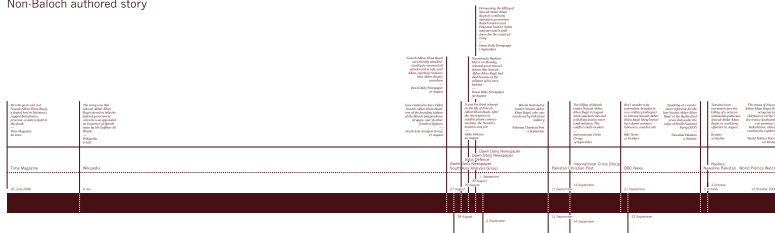
Note there is Baloch-authored (in *italics*) and non-Baloch-authored content. The research questions relate to the amount of Baloch-authored content accessible in Pakistan, as well as the level of distinctiveness of the story of his death from Baloch-related sites vis-à-vis non-Baloch. Where the first question is concerned, it is remarkable, in some sense, how “well” Pakistan appears to be blocking Baloch-authored content, for so little is redistributed. Phrased differently, the content circulation is relatively low. In the depiction of the “leaky content,” where the Baloch and non-Baloch content are resized according to frequency of returns, the Baloch-authored story size is small (see Fig. 12.3). Among the scant number of sites carrying a Baloch-authored story, often with redistributed content from blocked sites in Pakistan, are *Gedrosia.blogspot.com*, *Intellibriefs.blogspot.com*, *Ezboard.com*, *Thechosenpeople.blogspot.com*, *Thebalochpeople.org*, *Dc.indymedia.org*, and *Baltimore.indymedia.org*—blogs, forums, and indymedia sites. (Later, the account used on *Ezboard.com* by “hinduunity” was “locked down” under the site’s terms of use, after a threatened lawsuit, recounted on *intellibriefs.blogspot.com*.⁶²⁶ *Hinduunity.org* now has its forum hosted on its own site.) To take up the second question, the difference in the Baloch and non-Baloch versions of the story of the death of Nawab Akbar Kan Bugti is stark for the Baloch reference to murder as opposed to killing.

LEAKY CONTENT: AN APPROACH TO SHOW BLOCKED CONTENT ON UNBLOCKED SITES IN PAKISTAN – THE BALOCH CASE.

A timeline of the story of the death of Baloch leader, Nawab Akbar Khan Bugti, from Baloch authored and non-Baloch authored sources. Phrases resized according to frequency of mentions on sites accessible in Pakistan.

Timeline

Non-Baloch authored story



Baloch authored story



Timeline with story snips resized according to frequency of appearance



Data: Do You Know International and Conflict Relations. Analyzed by Richard Rogers and Edin Sarić. Design: by Nuno Tomazinger

© 2006 Geocomp and IBM

FIGURE 12.3.

Surfer Re-Routing

The famous quotation about how the Internet treats censorship—a version of which is the title of this chapter—is attributed to John Gilmore, co-founder of the Electronic Frontier Foundation. In the notes to his 1998 paper, “Why the Internet is Good,” Internet law scholar, Joseph Reagle, has the following annotations for the original quotation (in bold):

“The Net interprets censorship as damage and routes around it.”
 John Gilmore (EFF). [source: Gilmore states: “I have never found where I first said this. But everyone believes it was me, as do I. If you find an appearance of this quote from before March ‘94, please let me know.” Also in NYT 1/15/96, quoted in CACM 39(7):13. Later, Russell Nelson comments (and is confirmed by Gilmore) that on December 05 1993 Nelson sent Gilmore an email stating, “Great quote of you in Time magazine: ‘The net treats censorship as a defect and routes around it.’”] ⁶²⁷

The technical thought behind the quotation refers to packet switching, as another legal scholar, James Boyle wrote in 1997:

The distributed architecture and its technique of packet switching were built around the problem of getting messages delivered despite blockages, holes and malfunctions. Imagine the poor censor faced with such a system. There is no central exchange to seize and hold; messages actively ‘seek out’ alternative routes so that even if one path is blocked another may open up. Here is the civil libertarian’s dream. ⁶²⁸

There are now technical means to route around censorship, such as the circumventor by peacefire.org, a proxy service. Lists of proxy servers are updated frequently, in the ongoing race to stay a day or two ahead of the updates furnished by the content filtering software companies to their clients. Peacefire.org claims that filtering companies are routinely three to four days behind in updating their blacklists of proxies, so peacefire’s fresh proxy lists are useable on any given day. The intensive censorship and anti-censorship work behind the scenes is telling for how the discourse has changed for “route arounds.” Rather than being built into the infrastructure of the Internet, routing around should be described as labor-intensive and semi-manual work—proxy detection, list updating, alert sending. Thus, the discourse of routing around censorship is changing from the reverence of the Internet architecture and the far-sighted architects of the end-to-end principle to governance as well as to artful technique.

In an effort to show the routes, not for packets, but for content surfers, the Internet researchers' global list of women's rights sites was employed to make a surfer route map, as it may be called initially. As in the URL discovery method described earlier, the sites' outlinks were captured, and a network graph generated, showing the clusters of women's rights sites disclosed by inter-linking. The route map has sites and paths annotated in red and green, with red indicating known blockage (see Fig. 12.4).

The map plays on reworked ideas from hypertext theory whereby a surfer "authors" a path through Web space, one that eventually may be retrieved in the browser history. It also harkens to the art of surfing as opposed to mere searching. If one were to think of a surfer in China moving through a women's rights space (largely in English owing to the URLs on the Internet censorship researchers' global list), and authoring some sense of a story, first, from the seed list, hrw.org/women, ifeminists.com and womenofarabia.com (now offline) would not figure among the sources, for they are blocked.⁶²⁹

Which issues and stories about women's rights in China are discussed on hrw.org/women and ifeminists.com? Is there a path to similar or related content on unblocked sites? ifeminists.com have 10 entries on China: 3 of the 10 deal with the one-child policy, and the disproportion of boys born. Another follows from the "shortage of women," and reports the trafficking of North Korean women, sold to Chinese "husbands." A syphilis epidemic is discussed in two further stories, and the others deal with sexual harassment, online porn, easier divorces and AIDS, respectively. In discussing South Asia, China, and South Korea, Human Rights Watch, whose entire site is censored in China, writes about preferences for boys, "sex-selective abortions" as well female infanticide.⁶³⁰

In order to find surfer content routes, the actor sites on the women's rights map are queried initially for China-related topics discussed by [Ifeminists](http://ifeminists.com) and Human Rights Watch: "one-child policy" China, syphilis China, "shortage of women" China, AIDS China, "online porn" China, divorce China, and "sexual harassment" China. (Queries are made in English, for there is less censorship for English-language terms than for Mandarin.) Of the 88 nodes in the women's rights network, approximately one-third of unblocked sites return content on those key-word issues. The map organizes a women's rights-related content space, and through the choice of a map, as opposed to a list or a site tag cloud, suggests pathways. Because China has search engines delist sites and also performs key-word blocking, it is also important to cross-check known blocked words. A search through the known key words blocked in China did not turn up any of the above words.⁶³¹

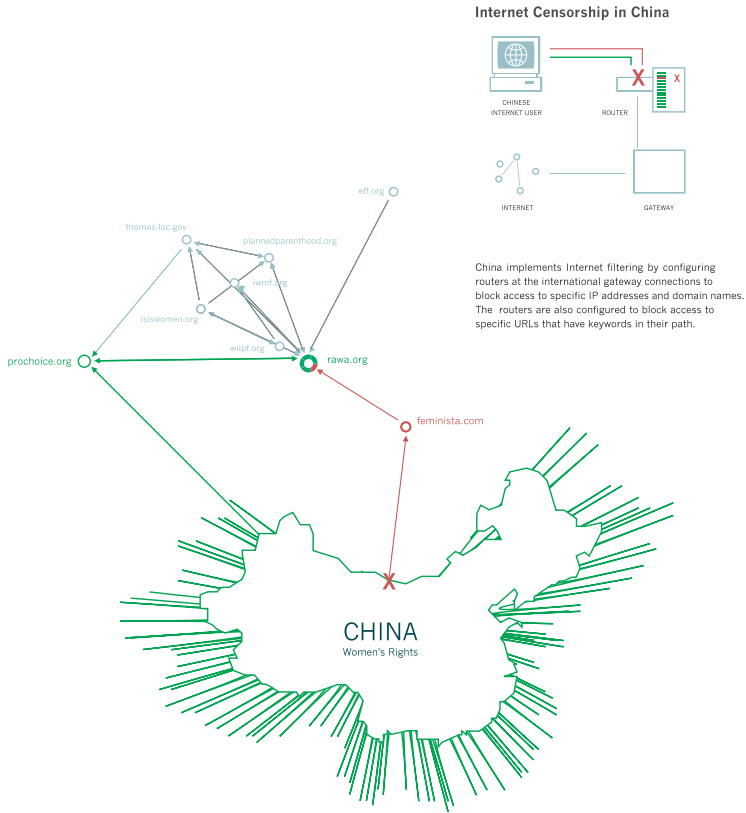
(Fire) Wall in China

The Internet treats censorship as a malfunction and routes around it?
A semi-manual approach to censorship circumnavigation.

Issue network location technique and circumnavigation map preparation

Including the dynamic sampling of URLs related to the issue (women's rights in China).

A list of women's rights URLs is entered into issuecrawler.net, which locates additional sites in the same or in related categories through hyperlink analysis. Results of the dynamic URL sampling technique are loaded into the map, showing how a surfer could find a route to uncensored sites.



INACCURATE DATA, DEMONSTRATES METHODOLOGICAL CONCEPT ONLY.

Internet Censorship Visualizations	2006
(Fire) Wall in China	06
Map by Issuecrawler.net Analysis by Richard Rogers Design by Alan Townsley	
© 2006 AndersonMedia and Government.org	

FIGURE 12.4.

CONCLUSION

State Internet Censorship is evolving from the directory-editor model, described earlier as old media-style for it assumes a Web constituted of institutions or actors operating single sites. The new media style, conversely, follows the movement of content around the Web (“circulation space”), and concentrates less on the policies of the censor than on the skills of the content movers, and how the results of those skills may be captured.

The new media style of Internet censorship research concentrates on describing both specific skills of the content movers as well as the techniques to measure the extent of the content movement. Importantly, the idea that the Web 2.0-style of content redistribution (scraping, feeding) is the new infrastructure of the Internet for routing around censorship appears to be in its infancy, however. The Baloch-authored story of the death of the tribal leader, it was found, is underredistributed on sites accessible in Pakistan. Although present and available (in English) information about the consequences of female infanticide in China (e.g., the “shortage of women” and the trafficking of North Korean “wives” to Chinese “husbands”) should not be considered abundant.

To date, digital journalism studies have focused on such subjects as newspapers going online and whether gatekeeping will be lessened owing to “interactivity.” Also treated are the relationship between blogs and mainstream news (who’s following whom) as well as the challenges of the amateur, where the Saddam Hussein hanging video appears to have greater claims to veracity owing to its mobile-phone graininess than news accounts of it filmed in a studio with anchorpersons. There is less emphasis on how information may become separated from its sources, and the consequences of the untethering for the distribution of attention.

When researchers and others consider the Web as circulation space, often there are particular connotations—Web as rumor mill or blogosphere as echo chamber, for example. Working with these assumptions, the “good journalist” would then be asked to trace the story back to a source. Source tracing, whether thought of in an archeological or genealogical sense, becomes the techno-epistemological practice, with an emphasis on source page date stamps. Here the practice is just as technical, however much the commitment changes to the expanse of the spread or “sharing,” as it’s sometimes called in participatory culture studies.

ACKNOWLEDGMENT

I acknowledge the support and assistance of the OpenNet Initiative, the project by the Citizen Lab at the University of Toronto, the Berkman

Center at Harvard University and the Advanced Research Group, Security Studies, University of Cambridge. The figures are by the Govcom.org Foundation, Amsterdam, with special appreciation extended to Erik Borra, Marieke van Dijk and Auke Touwslager. Laura van der Vlies provided valuable assistance.