A Taxonomy of Internet Censorship and Anti-Censorship

Draft Version December 31, 2010

Christopher S. Leberknight Princeton University Department of Electrical Engineering csl@princeton.edu	**Mung Chiang** Princeton University Department of Electrical Engineering chiangm@princeton.edu
Harold Vincent Poor Princeton University Department of Electrical Engineering poor@princeton.edu	**Felix Wong** Princeton University Department of Electrical Engineering mwthree@princeton.edu

ABSTRACT – Internet is supposed to be born free, yet it is censored almost everywhere, and severely censored in a few countries. The tug-of-war on the Internet between censors and anti-censors is intensifying. This survey presents a taxonomy on the principles, techniques, and technologies of Internet censorship and anti-censorship. It highlights the challenges and opportunities in anti-censorship research, and outlines a historical account via the lenses of news coverage in the past decade.

I. INTRODUCTION

Internet is supposed to be born free, yet it is censored almost everywhere, and severely censored in a few countries. *Censorship* is defined as the institution, system or practice of reading communication and deleting material considered sensitive or harmful [1]. Throughout history, various methods of censorship have been used to reinforce specific religious and political agendas. Technology has often served as a major obstacle and catalyst for mandating censorship. Even though technological advancement often ameliorates the inefficiencies and limitations of the past, it also can precipitate unforeseen consequences. The invention of the printing press in Europe in the 15th century is a prime example. The printing press not only increased the spread of information and knowledge but it also increased the practice and frequency of censorship. The task of maintaining the status-quo through effective censorship policies is undergoing rapid change due to the growth and diversity of different devices and networks including:

- Web traffic
- Email (e.g., Gmail)
- P2P file-sharing
- Video (e.g., YouTube)
- Texting and messaging (e.g., Twitter)
- VoIP (e.g., Skype)
- Social Networks (e.g., Facebook)

The interplay of the technological forces which promote anti-censorship and the policies used to enforce censorship is the main focus here. Specifically, the objective of this research is to increase awareness of the ramifications and negative consequences of Internet censorship and to advance the state of the art of circumvention technologies. Therefore, this research aims to stimulate critical thought and debate on Internet censorship by discussing four main themes.

Goals of this survey:

1. Provide an overview of research on censorship resistant systems and the different dimensions (political and technological) of online censorship.
2. Provide a review of the technological landscape and a taxonomy of anti-censorship technologies.
3. Discuss the most critical design features to enable a successful and effective anti-censorship system.
4. Discuss current trends and implications.

Through the discussion of these four main points we pose several open questions to the academic community:

Open questions

1. How can we quantify the efficacy of current online censorship technologies?
2. Which metrics can be used?
3. Are there fundamental limits to existing online censorship technologies?

Among the three main steps of censorship and its circumvention, (1) monitoring and surveillance, (2) blocking, filtering, and modifying content, (3) recording events, we will focus mostly on the middle step in this survey.

II. INTERNET CENSORSHIP

A. Principles

Internet censorship policies are primarily concerned with two main principles based on usability and censorship:

1. Limit the performance degradation
2. Enforce censors

The first principle is concerned with promoting usability. That is, the policy should attempt to censor information which may be disruptive to the status-quo without significant overhead or performance degradation. The second principle corresponds to achieving a certain level of accuracy with respect to restricting objectionable content.

While various forms of media have been used in the past to communicate and inform the public of current events none are as formidable to authoritarian regimes as the Internet. For example, the printing press helped to spread information by accelerating the publication and dissemination of books and newspapers, while radio and television broadcasting facilitated the rapid communication of events and helped to expand overall news coverage. However, the Internet enables a much more rapid generation and spread of information and ideas compared to previous technologies. In addition, the inherent characteristics of the Internet make controlling information on the network extremely challenging. One reason which makes information on the Internet difficult to control compared to other forms of media is that national borders are more permeable online: residents of a country that bans certain information can find it on websites hosted outside the country [8].

Another main reason which makes online information especially difficult to control has to do with the fundamental design and objective of the Internet. The essential requirement was to design a distributed system which was secure and would be less susceptible to failure and damage from a single point of failure. The very nature and advantage of a distributed system is that in the event there is some damage or failure in the network, transmission can be routed around the damage. In addition, to allow for communication between different systems a set of standard protocols would need to be developed to ensure interoperability. As a result, the exact characteristics such as robustness which make the Internet an ideal platform for communication and dissemination of information also make it very difficult to regulate the spread and access of information. Therefore, the combination of the ability to rapidly generate and share new ideas coupled with the complexity of controlling information flow, creates a viral effect which can pose significant risks to authoritarian regimes if the information contains subversive content which may influence the status quo or incite collective action and free thought. As a result, Internet censorship, which is defined as the control or suppression of the publishing or accessing of information on the Internet [8], has been steadily increasing in several totalitarian regimes. Even though censoring information on the Internet may be more difficult compared to other forms of media, several techniques have been developed and are in use in several societies such as China, Iran, and Syria.

The criteria of censorship include the following:

- *Cost:* both resource and opportunity cost, which directly impacts the availability of censors.
- *Scope:* the range of communication modes censored.
- *Scale:* the number of people and devices that can be simultaneously censored.
- *Speed:* the reaction time of censors.
- *Granularity:* the resolution at different levels, e.g., server, port, webpage, end user device, etc.
- *False negative:* the accuracy of censors.
- *False positive:* too high a false positive rate depletes the censor resources.
- *Circumvavility:* how easily can the censors be disabled.

Each bullet in the above list also presents an opportunity for the designers of anti-censorship techniques.

B. Techniques

A review of relevant literature has revealed that the most prevalent use or practice of Internet censorship is primarily conducted in authoritarian regimes, such as China, Cuba, Iran, North Korea, Syria, and Tunisia [10]. Overall, it has been reported that China has the most advanced and sophisticated censor network [10]. These countries have employed several new policies and technologies aimed at controlling access to information on the Internet. This research presents a taxonomy of Internet censorship technologies to help identify and explain the different strengths and weaknesses of various censorship strategies. The taxonomy can be broadly categorized by attack mode, filtering method, and target which are used to achieve a certain level of digital censorship.

The attack mode defines the sources of interest within the network topology and an associated action. The source of interest within the network consists of *nodes, users, and links*. The objective of a specific Internet censorship policy may identify an attack point and action within the network. For example, node attacks may consist of DoS, domain de-registrations or server takedown. To attack or censor a particular user the censorship organization may first decide to trace and record specific user activity prior to blocking any content. Therefore, there may be specific instances in which the operators of the censorship network wish to monitor and record activity as opposed to blocking or filtering specific content. Another mode of online censorship is to attack a link within the network, which can be accomplished using techniques such as IP blocking/filtering, DNS tampering, and/or HTTP proxy filtering.

With respect to filtering method, perhaps the most prevalent type of online censorship technology is a method known as IP filtering. IP filtering is used to block or filter objectionable content by restricting access to specific IP addresses. There are several different methods to filter content and while other totalitarian regimes have utilized one or more methods, it has been reported that only China exercises all of them [10]. The most popular filtering methods are depicted in Figure 1, which contrasts the tradeoff between the different filtering techniques in terms of their accuracy versus their operational cost. For example, IP filtering is least costly to operate compared to stateful traffic analysis, but it does not provide a high degree of accuracy. Since many websites may be hosted on one IP address, blocking the IP address to restrict access to a particular website which contains objectionable content also blocks all other websites which may not contain objectionable content. IP blocking is simple and cheap to implement by providing routers with specific IP addresses to block. However, it may unintentionally block websites containing valuable or useful information.

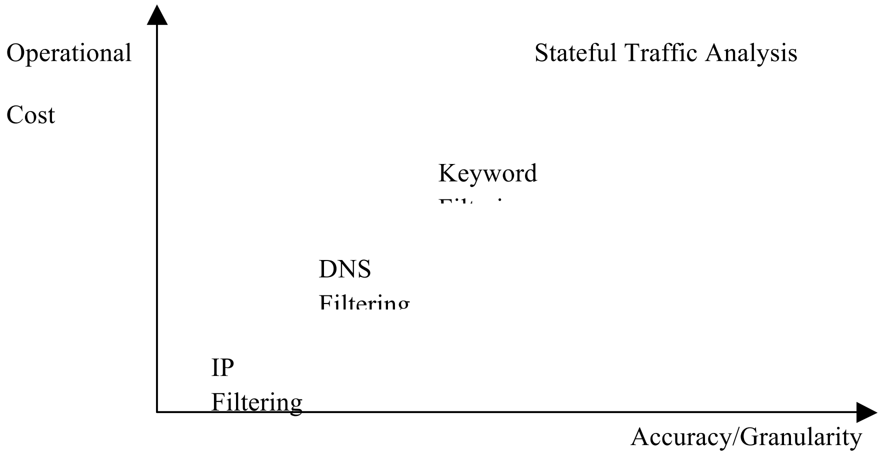


Figure 1. Internet Filtering Techniques

The next category in the taxonomy consists of the targets which are comprised of the technological devices and networks to be censored. The decision to enforce a specific online censorship strategy is primarily influenced by the task to be performed on the device or network application. For example, depending on the specific context, the censorship organization may block a device or network based on whether it is used to access or publish digital content. Deciding which filtering method or strategy to employ is not only based on operational cost vs. accuracy as depicted in Figure 1, but it is also based on several other salient factors as summarized in the last section's bullet list.

C. Technologies

Unlike circumvention technologies to be surveyed in Section III.C, there are not many commercially available censorship technologies. The technologies typically fall into two categories:

hardware and software. Software based technologies are primarily used filter and block content while hardware based technologies such as Deep Packet Inspection devices are used to classify network traffic and inspect packet headers and payloads. Much of the filtering software is developed internally. Smartfilter is a commercially available content filter which is developed by San Jose firm Secure Computing. In addition, several deep packet inspection devices are commercially available and are manufactured by companies such as Nokia, Siemens, and Allot Communications.

D. Implication for International Trade

While there are many examples of the social inequities brought about by Internet censorship [12][13][14] and even though all of these issues are great cause for concern, one area which has received far less coverage, and is becoming increasingly critical, is the implication of censorship on international trade. Essentially, by censoring online information domestic organizations can effectively discriminate against foreign suppliers [15]. For example, Google's decision to withdraw operations from China was due in part to non-compliance with Chinese censorship policies. Specifically, Google claims that as an organization which prides itself on being the source for information, cannot rightfully adopt a policy which enforces censorship. This clash between Google and Chinese authorities culminated with Google's decision to cease business operations in China on March 22, 2010 [16]. However, even though China insists Google must comply with Chinese censorship policies by blocking access to certain objectionable content, the Chinese search engine Baidu, often returns the same content. Therefore, it seems the use of online censorship to oust foreign competition may be another factor in play.

Furthermore a study published in 2009 by the European Centre for International Political Economy (ECIPE) concluded that after examining World Trade Organization's (WTO) official regulations and the current status of Internet censorship in various countries, the WTO has a strong case against governments involved in blanket Internet censorship. Blanket Internet censorship, or disproportionate censorship, involves permanent bans and entire blockages of websites [17]. The report by the ECIPE [17] as well the events surrounding Google's withdraw from China underscores the increasing importance and impact of online censorship not only on international trade but also on foreign policy. These two examples as well as the crackdown on Internet communications during the Iranian elections in 2009 have prompted the United States State Department to make unrestricted access to the Internet a top foreign-policy priority [18]. This new doctrine in addition to the surge in news coverage serves as formidable evidence that the future and spread of Internet censorship is a cause of great concern. This is not only a concern for citizens governed by authoritarian regimes, but there is also evidence which suggests online censorship is spreading to more liberal societies as well [50]. To further illustrate the prevalence of Internet censorship an analysis of news articles was extracted from http://news.google.com. The results in the Appendix Figure 1 depict the number of news articles containing the keyword "internet censorship" during a 17 year period from January 1993 to January 2010. The reports which indicate that China maintains the most advanced and sophisticated censorship network [11][12][13] and the results from Tables 1-3 in the Appendix further underscore the broad reach and implication for Internet censorship. As the size of the Internet market grows and the contribution to the global economy becomes more pronounced online censorship will require both a technological and political solution.

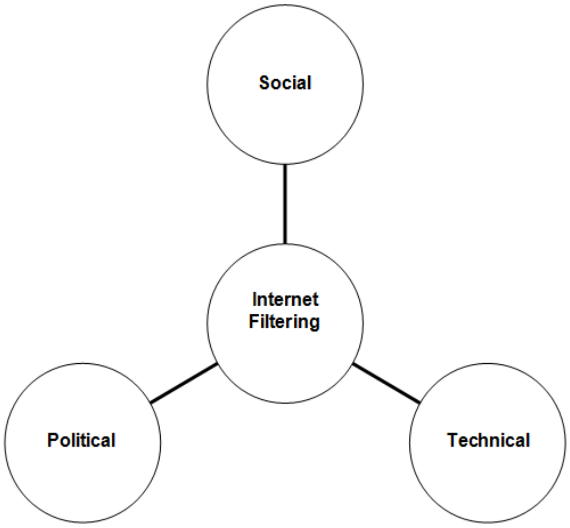


Figure 2. Three Aspects of Internet Censorship

While there have been many technological efforts and methods aimed at circumventing Internet censorship very few have had the capability to transform and influence public policy. That is, for circumvention or anti-censorship technologies to be truly effective it must be designed with the intention to not only provide free unrestricted access from a social perspective, but it must also be designed with the goal of transforming public policy. Internet censorship is a social, political and technical problem and each of these domains, depicted in Figure 2, interact in ways that can strongly reinforce one another. Therefore, some authoritarian regimes have discovered that the successful implementation and sustainability of Internet censorship not only requires advanced technologies, but also requires social or self-censorship which can be enforced through harsh punishments and political ideologies which encourage acceptance of the status quo. Consequently, to defeat Internet censorship the same three structures must be attacked. The question is whether technology can serve as the catalyst to create a domino effect.

From a technological perspective there are two methods destabilize information control. The first method requires that foreign supporters of anti-censorship lead in the technology evolution. This, however, will only be a short term solution based on whether or not the anti-censorship community can financially sustain itself. This would be highly unlikely if not backed by government financing as Internet censorship is sponsored by authoritarian regimes. Therefore, unless governments in favor of free speech and expression openly support and finance the destabilization of Internet censorship in foreign countries leading in the technology evolution will be difficult to realize because of the uneven playing field. A more likely approach and technological solution may be to design technologies that maximize the cost to operate censorship networks or increase foreign awareness of the consequences of Internet censorship. This was accomplished during the Iranian elections in 2009 [18] when Iranian citizens used YouTube and Twitter to bring images of violence and terror to the rest of world. With respect to increasing operational costs, based on the information in Figure 1, keyword censorship and stateful traffic analysis are the most expensive technologies to operate and therefore may lead to potential anti-censorship opportunities. Some underexplored anti-censorship technologies essential design characteristics will be discussed in the following sections. However, to understand the role of technology as a vital part in the overall solution to Internet censorship a review of previous research is provided in the next section followed by a taxonomy of anti-censorship technologies.

III. ANTI-CENSORSHIP - PREVIOUS RESEARCH

Extant literature on circumvention technologies discusses several techniques and strategies for deigning censorship resistant systems. There are two main dimensions: free access to information and free publication of information, the second being even more challenging than the first.

A key component for any censorship resistant system or circumvention technology is to ensure *privacy* by enabling users to communicate undetected in a censorship network. This is often accomplished by incorporating certain techniques such pseudonymity and anonymity into the system. However, previous research suggests that current techniques to ensure privacy still reveal a significant amount of identifying information [19]. Rao and Rohatgi (2000) indicate that techniques from linguistics and stylometry can use the identifying information to compromise pseudonymity. They suggest some countermeasures to address syntactic and semantic leaks of information. With respect syntactic leaks the authors suggest using a thesaurus tool, which could prompt the user to use alternatives while composing messages thereby reducing variations in vocabulary. For semantic leaks, they suggest translating the message to another language and then back again to the original language [19]. In addition to addressing the limitations for ensuring privacy using tools other research has introduces four properties: anonymity, unlinkability, unobservability and pseudonymity, and a set of anonymity metrics, which can be used to improve the design and evaluation of censorship resistant systems [20]. Expanding on research which has introduced tools, properties and metrics for ensuring privacy and specific applications to censorship resistant systems privacy via anonymity has also been explored by investigating the limitation of different *network topologies and document storage techniques*. Due to the single point of failure or ability to conduct denial of service attacks on centralized designs, network topologies such as peer-to-peer approaches for addressing anonymity have been suggested. One example is a peer-to-peer protocol that guarantees both anonymity and censorship resistance in semantic overlay networks [21]. Other literature describing peer-to-peer methods document storage techniques for protecting access and publication of documents have also been investigated. For example Serjantov (2002) discusses a peer-to-peer architecture for a censorship resistant system with user, server and active-server document anonymity as well as efficient document retrieval [22]. In addition, research by Waldman and Mazi`eres (2001) introduce a unique document storage mechanism known as entanglement in which newly published documents are dependent on the blocks of previously published documents [23].

So far the analysis of previous research has identified two main challenges for designing censorship resistant systems. These challenges include research focused on content protection and anonymity to ensure privacy. Other research has proposed novel *routing protocols* to address these two main challenges.. One notable example of such research was conducted by Katti et al (2005). Their research presented a protocol that uses a combination of information slicing and source routing to provide anonymous communication similar to Onion Routing but without a public key infrastructure [24]. Subsequently, a paper by Sovran et al (2008) presents a peer-to-peer system of relays which enables users within a censored domain to access blocked content using restricted service discovery [25].

In addition to content protection and anonymity other approaches for designing censorship resistant systems have centered on issues related to *content filtering.* For example, previous research has proposed a variation of censorship resistance (CR) that is resistant to selective filtering even by a censor who is able to inspect (but not alter) the internal contents and computations of each data server, excluding only the server's private signature key [26]. Further examples of research involving content filtering include a paper by Wolfgarten (2005)[27]. Wolfgarten (2005) analyzes large-scale, countrywide Internet content filtering and discusses techniques to effectively defeat censorship based on results from several tests. In addition, Crandall et al (2007) presents an architecture for maintaining a censorship "weather report" for determining which keywords are filtered over time [28]. Subsequently, a recent study by Park et al (2010) provides results from measurements based on filtering HTTP HTML responses in China. Their results suggest that the distributed nature of the Chinese filtering system and the problems inherent to distributed filtering are likely among the reasons it was discontinued, in addition to potential traffic load problems [29].

Therefore the main technical approaches for addressing challenges with designing censorship resistant systems include: (1) anonymity, (2) content protection, and (3) content filtering. In addition to the technical approaches and research on censorship resistant systems discussed above, several *social and behavioral methods* have also been investigated. For example, the first economic model of censorship resistance based on conflict theory and node preferences in a peer-to-peer system was presented by Danezis and Anderson (2004)[30]. Their model assessed how two different design philosophies (random and discretionary distribution of resources) resist censorship. The main finding was that, under the assumptions of their model, discretionary distribution is better. The more heterogeneous the preferences are, the more it outperforms random distribution. Pachinko and Pimenidis (2007) also explore social and behavioral methods to address challenges with designing effective censorship resistant systems [31]. In their research they define a model of a censorship resistant system based on a trusted directory which is used in order to prolong contacts among peers based on their reputation in a way, that honest members get contacts only to other honest peers and colluded members remain isolated.

As surveyed above, many different approaches to design censorship resistant systems have been proposed. The approaches so far have consisted of possible solutions from both technical and social perspectives. However, after a thorough analysis of previous research no single approach has proposed a solution which synthesizes both perspectives. Based on the information in Figure 2, a comprehensive and successful Internet censorship strategy involves collaboration and coordination among various social, political and technological entities. Therefore, a solution to Internet censorship must attempt to exploit the vulnerabilities within each entity. A solution to Internet censorship may evolve from a technological perspective provided it is designed with the optimal combination of features including an underlying or indirect motive to destabilize social and political structures. The rest of this paper will present an overview and taxonomy anti-censorship systems followed by a discussion of critical design features, concluding remarks and future research directions.

IV. TAXONOMY OF ANTI-CENSORSHIP TECHNOLOGIES

A. Principles

The primary objective of an anti-censorship system is to connect censored users to the uncensored Internet securely and anonymously. This can be accomplished by adopting one of the two anti-censorship principles:

1. Make it too costly to censor
2. Lead in the technology evolution

There are essentially four approaches of anti-censorship:

1. Volume-based
2. Speed-based
3. Covert channel-based
4. New technology-based

Similar to our list of 8 dimensions of censorship criteria, we have a list of 7 dimensions for anti-censorship:

- *Availability:* there is no use of an anti-censorship technology if the target users cannot acces it.
- *User-friendliness:* an often under-explored dimension, given the large population of users who are not technology-savvy.
- *Verifiability:* how can a user verify that the software is not a monitoring tool from the government.
- *Scope:* how many modes of communication can be covered.
- *Security:* this is the most obvious dimension of an anti-censor.
- *Deniability:* if caught, how can a user deny her involvement.
- *Performance:* how much will throughput and delay be degraded by using the anti-censor.

These metrics trade-off against each other, and the analysis and comparison of anti-censorship techniques can be carried out in the tradeoff space, e.g., scope-deniability vs. performance-availability, or user-friendliness vs. deniability.

A typical anti-censorship system is comprised of many components working together [32]. Conceptually, a censorship network can be viewed as a set of filters or a firewall and possibly coupled with manual processes which restrict users from accessing or publishing certain content. Figure 3 provides an illustration of the typical components comprising an anti-censorship system. The process to circumvent the censors can involve several steps as follows: censored users (1) use circumvention client software (2) on their computers to connect to circumvention tunnels (4), usually with the help of a tunnel discovery agent (3). Once connected to a circumvention tunnel, a user's network traffic will be encrypted by the tunnels and penetrate the firewall (7) without being detected by the censors (6). On the other side of the firewall, the network traffic will enter a circumvention support network (8) set up and operated by anti-censorship supporters (9). The computers, sometimes called nodes, in the circumvention support network act as proxies to access content from the unobstructed Internet (10) and send the information back, not necessarily taking the same route, to the censored user's computer [32]. Initially if a censored user knows nothing about the other side of the firewall, it is necessary to get them boot-strapped by employing out-of-band communication channels (5). Such channels include emails, telephone calls, instant messages, and mailing of CD-ROMs. Sometime users can also take advantage of these channels to locate circumvention tunnels (4), if the client software in use does not have a tunnel discovery agent (3) [32]. The key component in the overall system which facilitates covert communication within the censor network is primarily based on the software or tools and the underlying circumvention methodology.

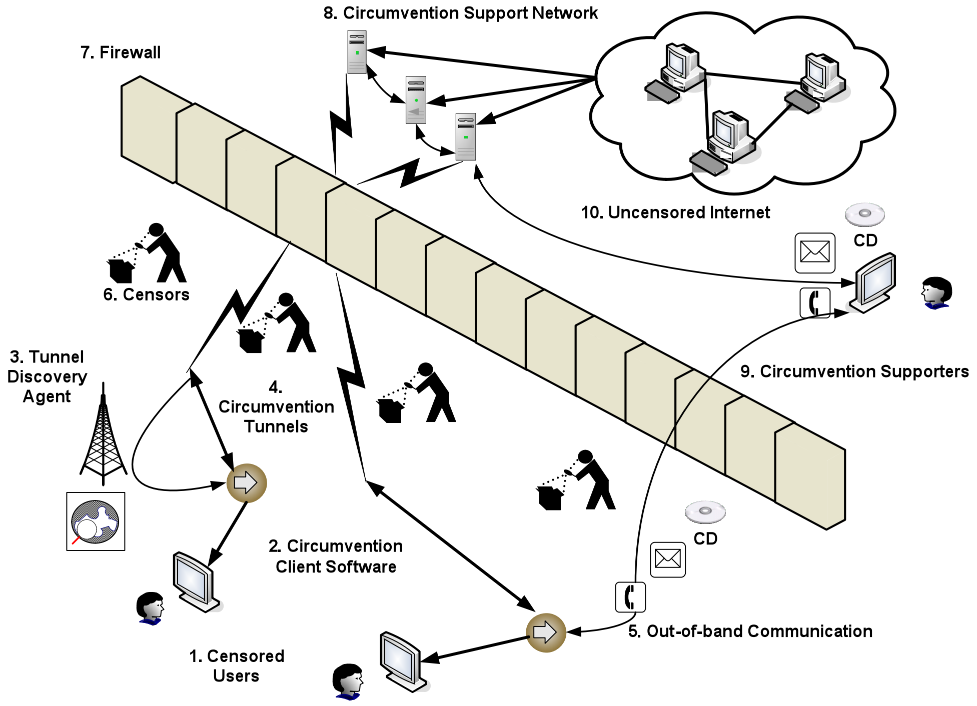

8. Circumvention Support Network
7. Firewall
10. Uncensored Internet
CD
6. Censors
3. Tunnel Discovery Agent
4. Circumvention Tunnels
9. Circumvention Supporters
CD
2. Circumvention Client Software
5. Out-of-band Communication
1. Censored Users

Figure 3 Anatomy of Anti-Censorship System [32]

B. Techniques

It has widely been reported that China has the most advanced censor network [32][33][34][38][39]. Consequently, the majority of research on Internet censorship including the evaluation of circumvention technologies has been investigated by analyzing various aspects of the censorship network controlled by the Chinese government. Therefore, themethods and policies employed by the Chinese government can help to inform and guide less technologically advanced authoritarian regimes who are interested in increasing their ability to censor online content. To this extent, due the threat of diffusion and adoption of Chinese Internet censorship technologies and policies by less technologically-advanced regimes, it is imperative that the subsequent discussions focus on online censorship in China in an attempt to hasten and undermine free thought and freedom of expression.

The most common type of online content blocking strategies in China consist of IP address blocking, DNS Hijacking and Content Filtering such as keyword or URL blocking [32][33][34]. IP address blocking operates by restricting users from accessing content by blocking the IP address where the content is hosted. IP blocking has the undesirable effect of over blocking since many sites can be hosted on a single IP address. Blocking the IP address of the site which contains “objectionable” content will also block all other sites on the same IP address which may not contain “objectionable” content.

DNS hijacking is finer grained compared to IP blocking, but it still is susceptible to over blocking. DNS hijacking allows operators to block access to content by blocking the name of the site instead of the IP address. For example, DNS hijacking would block or redirect users to another site when they tried to access www.google.com. Therefore, if multiple sites are hosted from one IP address only the sites containing the name to be blocked will be restricted. However, in the event some news article needs to be censored on a particular website the contents of the article cannot be censored without blocking the entire site. To address this limitation advances in content filtering such as keyword or URL filtering have been implemented to enable a higher degree of accuracy and granularity. The tradeoff between operational costs versus accuracy was illustrated in Figure 1.

Several anti-censorship techniques have been developed to circumvent the aforementioned technical filtering methods. While there are many academic projects actively engaged in the development of circumvention technologies our focus for this research is on the most common and popular commercial applications used for Internet censorship circumvention. The variety of commercial anti-censorship applications is based on one of the following circumvention methods described in Table 2 [33].

Method	Definition
HTTP Proxy	HTTP proxying sends HTTP requests through an intermediate proxying server. A client connecting through an HTTP proxy sends exactly the same HTTP request to the proxy as it would send to the destination server unproxied. The HTTP proxy parses the HTTP request; sends its own HTTP request to the ultimate destination server; and then returns the response back to the proxy client
CGI Proxy	CGI proxying uses a script running on a web server to perform the proxying function. A CGI proxy client sends the requested URL embedded within the data portion of an HTTP request to the CGI proxy server. The CGI proxy server pulls the ultimate destination information from the data embedded in the HTTP request, sends out its own HTTP request to the ultimate destination, and then returns the result to the proxy client.
IP Tunneling	Some of the most common tools used for IP Tunneling include virtual private networks or VPNs. VPNs give the user client a connection that originates from the VPN host rather than from the location of the client. Thus a client connecting to a VPN in a non-filtered country from a filtered country has access as if he is located in the non-filtered country.
Re-routing	Re-routing systems route data through a series of proxying servers, encrypting the data again at each proxy, so that a given proxy knows at most either where the traffic came from or where it is going to, but not both.
Distributed Hosting	A distributed hosting system mirrors content across a range of participating servers that serve the content out to clients upon request. The primary advantage of a distributed hosting system is that it provides access to the requested data even when the original server cannot, for instance if the original server has been overwhelmed by traffic or even taken down by a denial of service attack

Table 2 Circumvention Methods [33]

C. Technologies

There is a wide range of anti-censorship technologies developed over the past 15 years:

- User anonymization (JAP, ANON, Tor onion router)
- Covert channel (Infranet)
- Deniable publishing (Publius, Tangler, i2p)
- Web proxy server (Triangle Boy, Garden, UltraSurf, DynaWeb, Gpass)
- Turn PC into encrypted server (psiphon, peacefire)

- Application tunneling (Relakks, Guardseter, HTTPTunnel)
- Web tunneling (Anonymizer, Freegate, GhostSurfer)
- Google Cache and RSS aggregator

Fundamental research questions remain to be addressed in many of the above. For example, web proxy server technologies are among the most often used anti-censors. Yet there is only limited quantification on how to use a large number of hidden proxies with independence, memorylessness, and costly discovery properties to achieve fundamental limits on how long can users discover proxies but censors cannot.

As soon as the Internet became available in China in 1996 the Chinese government began to develop technical filtering methods [40]. It has been reported that the three technical filtering methods used in China: IP blocking, DNS Hijacking, and Content Filtering first emerged in 1999 and 2002 [41]. Since 1999, many circumvention technologies employing one of the circumvention methods, described in Table 2, have been developed. A timeline of the most common circumvention technologies alongside the emergence of the different technical censorship methods are presented in Figure 4 and Table 3. The information in Figure 4 provides insight into the trends and evolution of the various circumvention technologies and with respect to the emergence of the three technical filtering methods. During the time this manuscript was written, no detailed information was available for the most recent technologies developed in 2010. Consequently, the proceeding analysis excludes the Anonymizer Universal and the Xi Xiang Project which was both developed in 2010. Out of the remaining 15 circumvention tools only 11 are still actively used, suggesting that circumvention technologies have a very short lifespan. After the acquisition of the company which developed Triangle Boy, development for the tool ceased to exist and was no longer supported after 2003 (lifespan 1993-2002) [32].

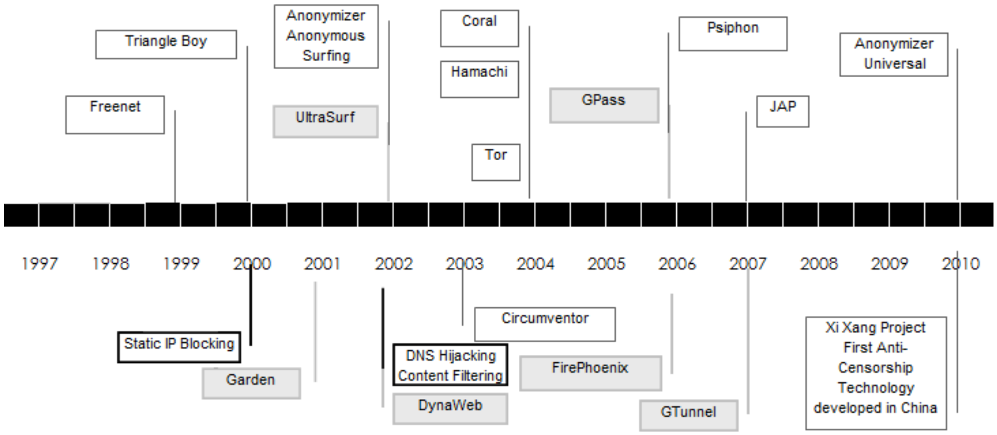


Figure 4 Timeline of Circumvention and Censorship Methods

In addition, TOR has declined in use and popularity following the blocking of 80% of its relays in September 2009 (lifespan 2004-2009)[36][42], and GPass abruptly stopped working in March 2009 with no explanation from the developers regarding the present or future status of the software (lifespan 2006-2009) [37]. The inactivity and availability of these systems over time indicates the average lifespan for circumvention technologies are approximately 4 ½ years (4+6+5+3/18). This is a rough approximate due to the limited availability of data, but it may be inferred that, while there may be many reasons for the decline of a particular anti-censorship system, one definitive contributor is China's advanced filtering methods and ability to adapt its methods to combat new circumvention technologies. To increase the ability to combat online censorship several organizations have joined an alliance known as the Global Internet Freedom Consortium (GIFC) which was formed in 2006. The GIFC was is an alliance of organizations that develop and deploy anti-censorship technologies for Internet users residing in oppressive regimes [32]. This alliance allows members to combat online censorship through technical advancements, promotion and support by leveraging the combined strengths and capabilities of each member organization. The grey shaded boxes in Figure 4 consist of anti-censorship systems which are all provided by the GIFC (http://www.internetfreedom.org/). The boxes outlined in bold font indicate the year in which the Chinese Government instituted technical filtering methods. Together the side by side comparison of anti-censorship technologies and Internet censorship filtering methods, in Figure 4, provides some indication of how anti-censorship technologies have evolved. Specifically, it can be observed that the disproportionate number of circumvention technologies compared to the limited number filtering methods implies either that these filtering methods are extremely effective and/or there are other factors at play which empower the Chinese governments' ability to enforce Internet censorship. Based on our analysis we believe the effectiveness of Internet censorship is due to a combination of strategies designed at various social, political and technological levels.

To further elucidate the way in which anti-censorship technologies are evolving, Table 3 presents the circumvention methods employed in each of the 15 tools presented in Figure 4, sorted by the year each tool was released. Two interesting observations can be made based on this information. First, it is evident that, overall, many more tools based on the HTTP proxy method have been developed compared to any other tool. While there are many factors that can influence the adoption and diffusion of any technology, the most likely candidate in this particular case is performance. A recent study evaluated the amount of delay or response time for a particular anti-censorship tool to return the content from a particular censored site. The results demonstrated that the tools based on the HTTP proxy method had the highest performance followed by the CGI proxy tools and the re-routing tools [33]. The second observation is that around 2003 there was a shift from developing HTTP proxy tools to IP Tunneling tools. This coincides with the time in which China began to implement and enforce content filtering. IP Tunneling is especially useful for coping with content filters since the specific nature and addressing of the original datagrams are hidden. In addition, when IP Tunneling is combined with IPsec it can be used to create a virtual private network [43]. This method therefore makes it very difficult for filters to inspect the actual contents of the communication and rely on more expensive methods such as stateful traffic analysis. However, even though some of the more recent tools are based on IP Tunneling, there is still some evidence which reports that older HTTP proxy tools, such as UltraSurf and DynaWeb are the most popular and widely used anti-censorship technologies [32]. In addition, to the superior performance of these two tools compared to the other tools evaluated another contributing factor for the success of UltraSurf and DynaWeb may the users trust in the system.

Tool	Year Released	HTTP Proxy	CGI Proxy	Re-routing	IP Tunneling	Distributed Hosting
Freenet	1999	✓				
Triangle Boy	2000	✓				
Garden	2000	✓				
Anonymizer	2002	✓				
DynaWeb	2002	✓				
UltraSurf	2002	✓				
Circumventor	2003		✓			
TOR	2004			✓		
Coral	2004					✓
Hamachi	2004				✓	
Firephoenix	2006				✓	

GPass	2006				✓	
Psiphon	2006		✓			
GTunnel	2007				✓	
JAP	2007			✓		
Total		6	2	2	4	1

Table 3 Circumvention Tools and Methods

Trust is one critical feature which should be considered when designing censorship resistant systems. The following section discusses the role of trust and several other essential design features.

V FURTHER DISCUSSION ON DESIGN CHARACTERISTICS

Understanding a user's perception of trust in an information system and especially in circumvention technologies is a critical factor influencing technology adoption and acceptance. Several IS studies have examined trust and trusting intention [45][46][47][48] but there has been little focus on trust relating to circumvention technologies. For circumvention technologies, trust can be broadly divided into two categories, trust within the user's social network and the user's trust within the technology. These determinants of trust are illustrated in Figure 5.

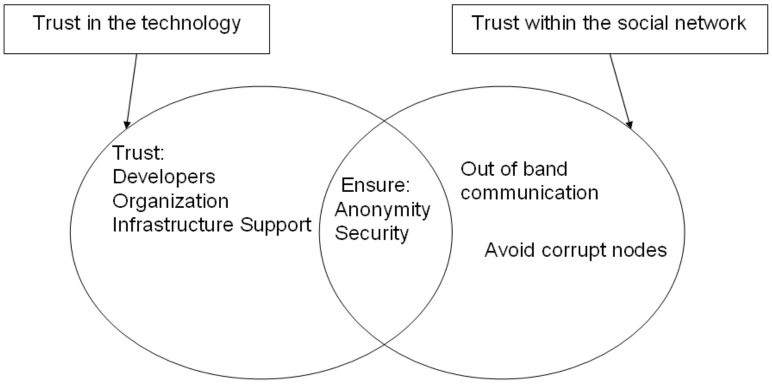


Figure 5 Determinants of Trust

Trust within the social network is influenced by out of band communication and the user's ability to avoid corrupt nodes. Often times to communicate new methods or new censorship resistant technologies individuals in a censored network rely on the exchange of information using emails, instant messaging, CD-ROMs, and telephone calls. This requires a strong degree of trust with other individuals in the network, and individuals who may be trying to penetrate or infiltrate the covert social network. Equally important is the user's trusting beliefs in the circumvention technology. This involves not only the user's ability to trust the organization or developers of the technology but also the user's trust in other aspects such as the organizations' supporting infrastructure and financial sustainability. For example, as the number of users increases, more investments will most likely be required to operate, expand, and maintain the organizations' server infrastructure [33]. Therefore, users must have trust in an organizations' financial sustainability. In addition, there are other technical aspects influencing a user's trusting intention that are related to the architecture and method employed for each circumvention technology. A decomposition of the different architectures, tools and the placement of trust are illustrated in Figure 6.

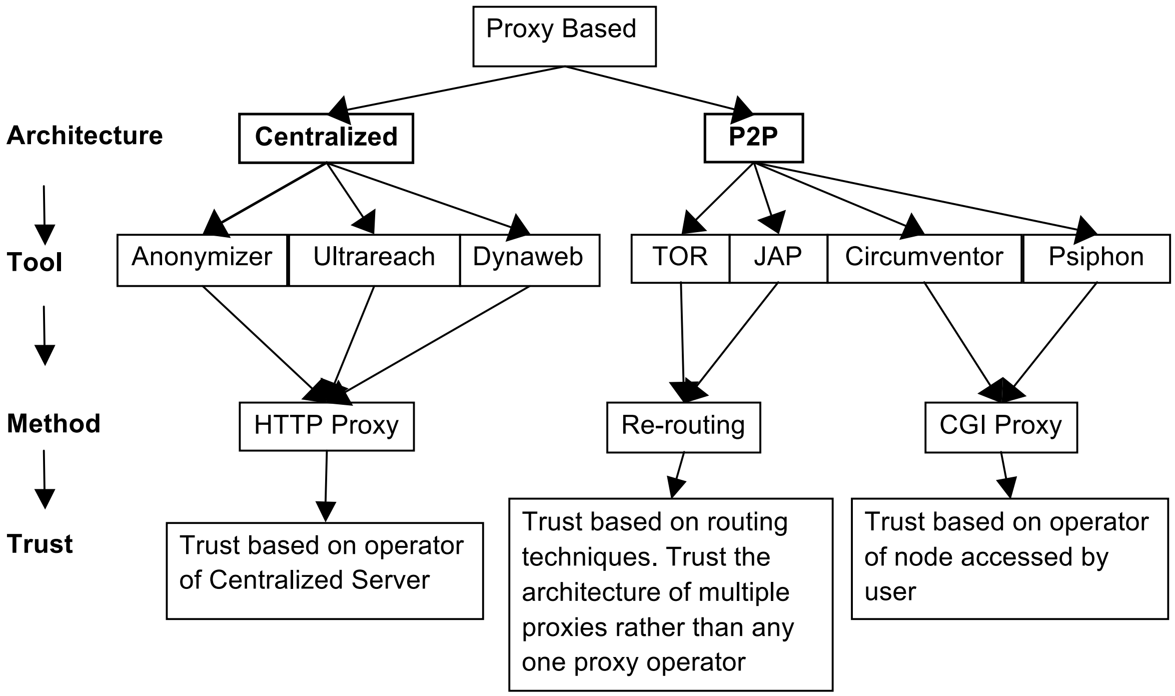
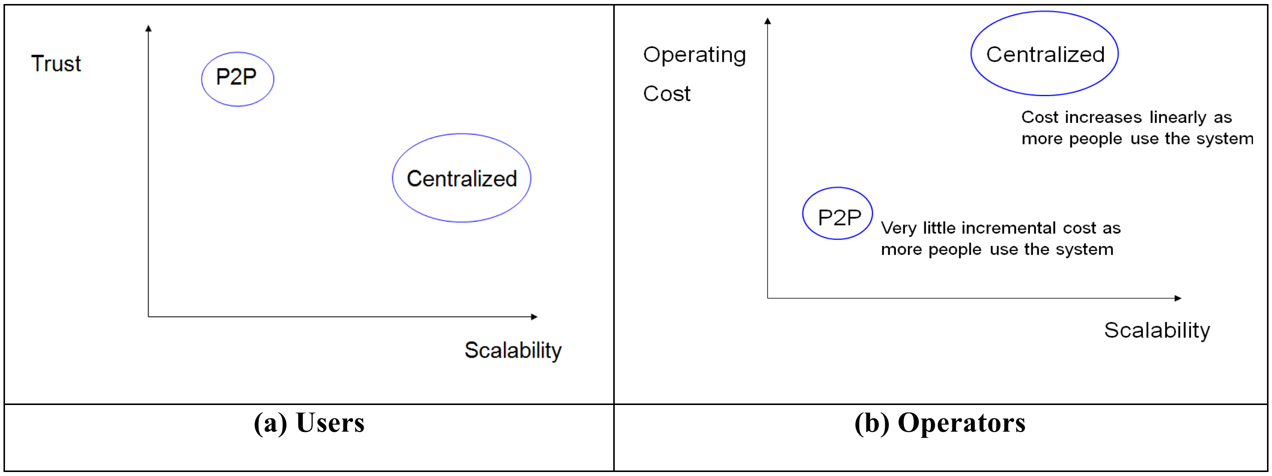

Proxy Based
Architecture
Centralized
P2P
Tool
Anonymizer
Ultrareach
Dynaweb
TOR
JAP
Circumventor
Psiphon
Method
HTTP Proxy
Re-routing
CGI Proxy
Trust
Trust based on operator of Centralized Server
Trust based on routing techniques. Trust the architecture of multiple proxies rather than any one proxy operator
Trust based on operator of node accessed by user

Figure 6 Trust based Architectures

Based on the information in Figure 6, existing circumvention technologies are either based on centralized or P2P architectures, and implement a variety of methods such as HTTP Proxy, Re-routing and CGI Proxy. Each technology or tool has different advantages and disadvantages regarding trust. For example, due to the distributed nature of P2P architectures, trust is displaced among many users. This may seem like a desirable feature since users will be more confident using a system if they believe several attacks on different peers will be necessary to compromise the system. However, if users had such faith in P2P systems then we would therefore expect to see a larger P2P user base compared to a centralized user base. However, this is not the case. It has been reported that centralized tools such as DynaWeb and UltraSurf are among the most popular circumvention technologies [32]. The P2P tools were all developed after 2003 (Table 3) and the centralized tools were all developed prior to 2003. This is the dividing point between pre and post content filtering methods. Based on a review of previous literature, centralized tools such as DynaWeb and UltraSurf outperformed other P2P based applications [32]. Therefore, while the newer P2P tools may be more capable at combating the latest content filtering methods, which were employed in 2003, it appears that performance still played a larger part in the users decision to use a particular tool. Even though P2P applications reduce the case of a single point of failure and they also distribute trust, P2P applications still suffer from one major disadvantage. The main disadvantage is that the authorities can infiltrate the network by masquerading as a collaborator. In a centralized system if the organization can demonstrate that they are trustworthy than there's less risk of infiltration from corrupt nodes. While P2P applications have the advantage of distributing trust they also have the disadvantage of not being able to enforce control over corrupt nodes. The ability to determine the exact amount of trust provided by a particular system is somewhat opaque, but this is the very nature of any covert communication system. Even though trust may influence an individual's decision to use the system, a more likely reason for determinant of behavioral intention to use is performance. Based on our analysis of existing research and popularity of different anti-censorship technologies, it seems that users are more likely to value performance over trust. This does not imply users are not concerned about trust, we believe that a certain degree of trust must exist but performance will still play a more integral role in a user's decision to use a particular technology. The tradeoff between trust and performance for users is illustrated in Figure 7a.


Trust
P2P
Centralized
Scalability
Operating Cost
Centralized
Cost increases linearly as more people use the system
P2P
Very little incremental cost as more people use the system
Scalability

(a) Users

(b) Operators

Figure 7 Centralized vs. P2P

Even though P2P applications are more susceptible to infiltration and corrupt nodes compared to centralized systems, overall, Figure 7a demonstrates they still offer a higher degree of trust. However, the higher degree of trust comes at the expense of performance and this is one likely reason why centralized systems have gained greater popularity. From the operators' perspective in Figure 7b, even though the centralized architectures scale much better they are much more costly to maintain as the user base increases. This may be one of the main reasons why more P2P tools were developed after 2003. While P2P technologies may be less costly to maintain, their poor performance has swayed users in authoritarian regimes that frequently experience online censorship to use centralized systems. However, such users will be severely impacted if organizations that develop centralized systems cannot obtain considerable financial support. Perhaps a consortium such as the GIFC is one strategy for addressing rising costs by combining resources and technical efforts. However, a better solution would be to lobby for government backed support and funding [44]. Fortunately, for many citizens in authoritarian regimes, recently there have been several U.S. backed efforts to support projects, organizations and technologies which promote Internet freedom [49].

While trust is one key factor influencing the adoption and use of anti-censorship tools, there are many other design factors that should also be considered, such as the list of 7 in Section IV.A. There are a number of circumvention or anti-censorship tools available, and the most appropriate tool is predicated upon several factors such as the specific task to be executed (i.e. access or publish content) and the user's requirements. A diagram illustrating all of the features we have identified and the features that are most crucial to the design and acceptance and success of any anti-censorship system is provided in Figure 8.

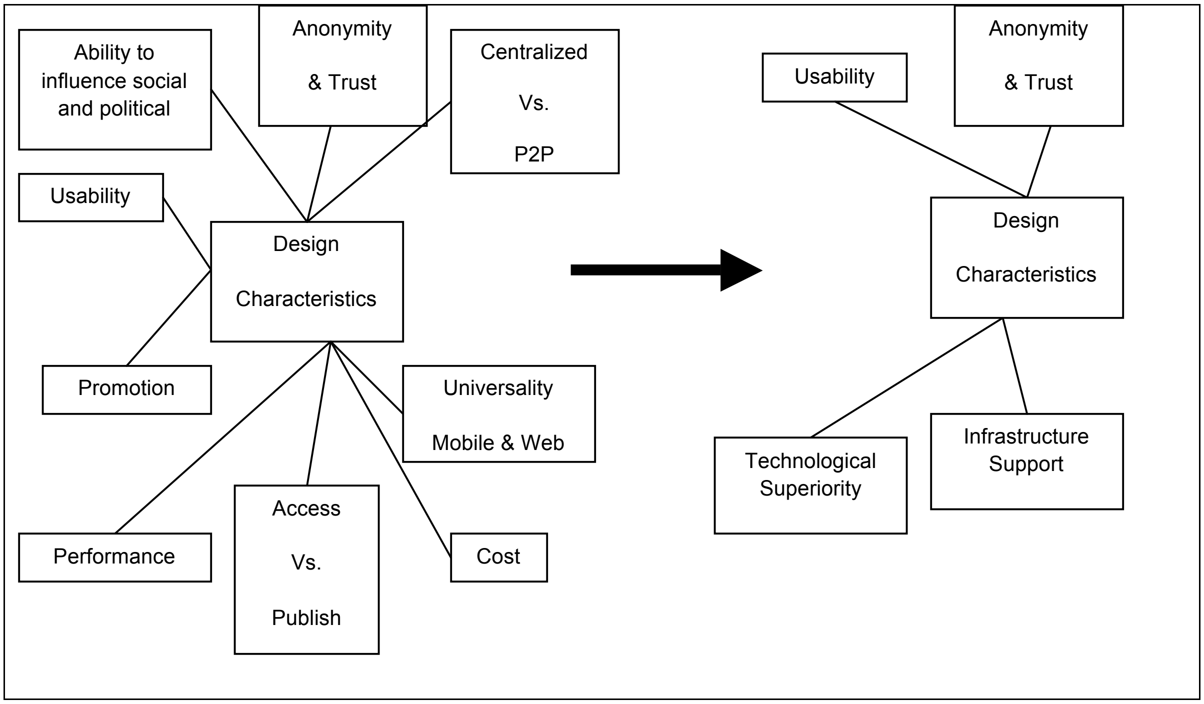


Figure 8 Design Characteristics for Anti-Censorship Technologies

VI. CONCLUSION

The main two objectives of this research are to provide an overview of online censorship including the interplay between technology and policy, and present a taxonomy and set of key design factors for anti-censorship technologies. With respect to the first objective, a historical review of information control including the evolution into online censorship and the corresponding commercial applications demonstrates how censorship has been practiced and employed throughout the world as a social and political tool to control free thought and expression. Many societies have reaped the benefits from technological advancements, however, in several totalitarian regimes, technology has often been perceived as a formidable opponent in the struggle to maintain the status quo. The intended benefits for new technologies are often met with unforeseen consequences. For example, technologies such as the printing press helped to streamline the publication and dissemination of information but it also inadvertently threatened closed societies who were intent on controlling which information should be allowed to proliferate. Throughout history there have been several other technologies which have improved communication and the spread of information. However, no single technology has posed as a great threat authoritarian regimes and censorship as the Internet.

U.S. Policy (US Congress)	Year	Technology (Initial Release)
• US House Policy: Bipartisan, Bicameral Bill Stops Internet Jamming- October 2, 2002 • US House Policy: Tear Down This Firewall - September 19, 2002	2002	Anonymizer Anonymous Surfing, **UltraSurf**, Dynaweb

• US House Representative Cox: House Passes Global Internet Freedom - July 16, 2003 · • US-China Economic and Security Review Commission: SARS in China: Implications for Information Control, Internet Censorship, and The Economy - June 5, 2003 · • 108th US Senate: Global Internet Freedom Act of 2003 - June 4, 2003 • 108th US House: Global Internet Freedom Act - HR 48 - January 7, 2003	2003	Circumventor
• Secretary of State Establishes New Global Internet Freedom Task Force - Feb 14, 2006	2006	**GPass**, **Firephoenix**, Psiphon, Twitter
• Global Online Freedom Act of 2007 - Dec 10, 2007	2007	JAP, **GTunnel**
• Testimony of GIFC in the US Senate Hearing on Global Internet Freedom (2008-05-20)	2008	
• U.S. State Department speaks to Twitter over Iran | Reuters - Jun 16, 2009 • Senators Push Digital Code of Conduct Forbes – Jun 25, 2009 • US to increase funding for 'hackivists' aiding Iranians - The Boston…Boston Globe – Jul 26, 2010 • US Supreme Court backs free speech on the Internet Seattle Times – Jan 22, 2009	2009	
• POLITICS BLOG: Senators announce formation of Global Internet Freedom Caucus-Mar 24, 2010 • US Government Works to Break Down Virtual Walls-Mar 19, 2010	2010	AnonymizerUniversal, Xi Xang Project (first anti-censorship technology software developed in China)

Table 4 US Policies and Release of Circumvention Technologies

Based on our analysis of online news articles in the Appendix, Internet censorship has been steadily increasing since 1993 with the largest incline occurring between 2007 and 2010. In addition to the social inequities brought about by censoring information, the sharp rise in online censorship during the last couple of years also presents major implications for international trade.

Events of such magnitude involving online censorship, limiting freedom of expression have prompted the U.S State Department to make unrestricted access to the Internet a top foreign-policy priority [18]. This climacteric move by the U.S. State Department marks a critical turning point in U.S. foreign policy which has taken several years to emerge due in part to many technological efforts. To highlight the interplay between technology and policy relating to online censorship a list of U.S. polices [52] concerning online censorship and the release of popular circumvention technologies is presented in Table 4. The data in Table 4 provides a timeline and side-by-side comparison of different technological advancements and efforts to crystallize political support. Based on the data in Table 4, significant political support and funding addressing online censorship issues was not achieved until 2009. This is a major milestone because even though there have been previous policies addressing censorship, adequate funding had yet to be awarded. This suggests that in 2009 Internet censorship became an important part of the U.S. political agenda. Consequently, government backed funding for anti-censorship projects and technologies will greatly improve the success and sustainability for efforts and technologies aimed at defeating online censorship.

The second objective of this paper is to provide a taxonomy of circumvention technologies and set of critical features to aid designers in developing new techniques to combat online censorship. There are many different tools and each tool has a specific function and advantage. Therefore, currently no tool exists which can be used to accomplish every task, but it is our intention to highlight which factors we feel are most critical to the success and adoption of the technology. Out of the 9 features presented in Figure 8, we suggest the following 4 features should be considered when designing censorship resistant systems: (1) anonymity and trust, (2) usability, (3) technological superiority, and (4) infrastructure support. Trust and anonymity play a very crucial role in an individual's decision to use the system. In addition, previous research has also suggested that usability or how easy it is to use the system is also a key determinant for behavioral intention to use. Subsequently, technological superiority must also be addressed to ensure the system operates at acceptable performance rates. Lastly, infrastructure support is listed as a key factor even though it is more of an operational consideration as opposed to a design factor. However, the lack of infrastructure support implies that as the user base increases so must the organizations ability to support the technology. Now more than ever this is becoming a reality due to recent announcements by the U.S. government to fund certain anti-censorship tools and projects [44][49].

While technology can play an integral role in combating Internet censorship, but ultimately success can only be achieved if the technologies are designed to exert economic or political pressures. For example, leading in the technology evolution may have some short term positive effects, but designing a system which increases the operational costs for the censorship organization will have a much greater impact. We provide a list of existing and potential future technological solutions in Table 5. Those techniques that are under-explored to-date are in bold. For example, based on our analysis of several online censorship methods illustrated in Figure 1, the exploration of strategies such as changing keywords using a randomized chain reaction will be costly to effectively filter.

Number	Anti-Censorship Techniques
1	Alternative DNS servers/names
2	Open proxy
3	Hopping IP servers and to popular servers
4	**Chopping up content across packet boundaries**
5	Conceal payload content by stegano/crypto methods
6	**Fountain code / network code / spreading code**
7	**Change keywords with randomized chain reaction**
8	Social media

9	Secure cloud computing (Google doc)
10	Remote log in to another computer
11	Secure phase for pre-agreement of protocol
12	**Timing covert channels**
13	**Extensive caching/translation**

Table 5 Future Anti-Censorship Technologies

In addition, the development of new anti-censorship techniques and tools will help to shed light on appropriate methods for quantifying the effectiveness of existing online censorship technologies. Our future research will investigate quantitative, new metrics to help understand the existence and achieving of fundamental limits for future anti-censorship technologies.

VII. REFERENCES

[1] Merriam Webster Dictionary http://www.merriam-webster.com/

[2] "History of Censorship." The Long History of Censorship.Mette, Newth. July 2001. Beacon for Freedom of Expression.. 15 Mar. 2009 http://www.beaconforfreedom.org/about_project/history.html.

[3] Government Control of Information , Cass R. Sunstein, California Law Review, Vol. 74, No. 3, Symposium: New Perspectives in the Law of Defamation (May, 1986), pp. 889-921 Published by: California Law Review, Inc.

[4] See K. DAVIS, ADMINISTRATIVE LAW TREATISE? 4.45, at 442-46 (2d ed. 1978).

[5] See, e.g., Pell v. Procunier, 417 U.S. 817 (1974); Saxbe v. Washington Post Co., 417 U.S. 843 (1974).

[6] W. WILSON, THE NEW FREEDOM1 13-14 (1913); see also id. at 130 ("Government must, if it is to be pure and correct in its processes, be absolutely public in everything that affects it.")

[7] United Nations General Assembly, Universal Declaration of Human Rights. Geneva: United Nations General Assembly; Resolution 217 A (III), UN doc. A/ 810 (III) (1948 Dec).

[8] http://en.wikipedia.org/wiki/Internet_censorship

[9] Ruthfield, Scott. “The Internet’s History and Development: From Wartime Tool to the Fish-Cam”, September 1995, http://www.acm.org/crossroads/xrds2-1/inet-history.html

[10] Roberts et al, “2007 Circumvention Landscape Report: Methods, Uses, and Tools,” The Berkman Center for Internet & Society at Harvard University, March 2009. Retrieved from http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/2007_Circumvention_Landscape.pdf

[11] Zittrain, Jonathan and Ben Edelman, *Empirical Analysis of Internet Filtering in China, IEEE Internet Computing*, March/April 2003 (Vol. 7, No. 2), pp. 70-77. http://cyber.law.harvard.edu/filtering/china

[12] Freedom on the Net: A Global Assessment of Internet and Digital Media, April 1, 2009. Retrieved from http://www.freedomhouse.org/uploads/specialreports/NetFreedom2009/FreedomOnTheNet_FullReport.pdf

[13] Internet Filtering in China, OpenNet Initiative, June 15, 2009. Retrieved from http://opennet.net/sites/opennet.net/files/ONI_China_2009.pdf

[14] Undermining Freedom of Expression in China: The role of Yahoo, Microsoft and Google, Amnesty International, July 2006. Retrieved from http://irrepressible.info/static/pdf/FOE-in-china-2006-lores.pdf

[15] Erixon, Fredrik and Lee-Makiyama, Hosuk, Jan. 6, 2010, Chinese Censorship Equals Protectionism, *The Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB10001424052748704842604574641620942668590.html

[16] Helft, Miguel and Barboza, David, March 22, 2010, Google Shuts China Site in Dispute Over Censorship, *The New York Times*. Retrieved from http://www.nytimes.com/2010/03/23/technology/23google.html

[17] Hindley, Brian and Lee-Makiyama, Hosuk, Protectionism Online: Internet Censorship and International Trade Law, European Centre for International Political Economy, Nov. 2009. Retrieved from http://ecipe.org/publications/ecipe-working-papers/protectionism-online-internet-censorship-and-international-trade-law/PDF

[18] Gorman, Siobhan, Jan 20, 2010, Web Access is New Clinton Doctrine, *The Wall Street Journal,* Retrieved from http://online.wsj.com/article/SB10001424052748703405704575015461404882830.html?mod=rss_Politics_And_Policy

[19] J. R. Rao and P. Rohatgi. Can pseudonymity really guarantee privacy? In *Proceedings of the Ninth USENIX*
Security Symposium, pages 85–96. USENIX, Aug. 2000. <http://www.usenix.org/publications/library/proceedings/sec2000/full_ papers/rao/rao.pdf>.

[20] George Danezis and Claudia Diaz. A survey of anonymous communication channels. Technical Report MSRTR-2008-35, Microsoft Research, January 2008.

[21] Backes, M., Hamerlik, M., Linari, A., Maffei, M., Tryfonopoulos, C., and Weikum, G. 2009. Anonymity and Censorship Resistance in Unstructured Overlay Networks. In *Proceedings of the Confederated international Conferences, Coopis, Doa, Is, and ODBASE 2009 on the Move To Meaningful internet Systems: Part I* (Vilamoura, Portugal, November 01 - 06, 2009).

[22] SERJANTOV, A. 2002. Anonymizing censorship resistant systems. In *Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS'02)*. MIT Faculty Club, Cambridge, MA.

[23] M. Waldman and D. Mazi`eres. Tangler - a censorship resistant publishing system based on document entanglements. In *Eighth ACM Conference on Computer and Communications Security*, Nov. 2001.

[24] S. Katti, , D. Katabi, and K. Puchala. Slicing the onion: Anonymous routing without pki. Technical report, MIT CSAIL Technical Report 1000, 2005.

[25] Y. Sovran, A. Libonati, and J. Li. Pass it on: Social Networks stymie censors. In 7th International Workshop on Peer-to-Peer Systems (IPTPS 08), Feb. 2008

[26] Ginger Perng, Michael K. Reiter, and Chenxi Wang. Censorship resistance revisited. In Jordi Herrera-Joancomarti, editor, Pre-Proceedings of the 7th International Workshop on Information Hiding, pages 279–293, 2005.

[27] S. Wolfgarten, "Investigating large-scale Internet content filtering," M.Sc. in Security and Forensic Computing 2005/2006, Dublin City University, Ireland.

[28] J. Crandall, D. Zinn, M. Byrd, E. Barr, and R. East. Conceptdoppler: A weather tracker for internet censorship. In *14th ACM Conference on Computer and Communications Security*, 2007.

[29] Park , Jong Chun and Crandall, Jedidiah R. Empirical Study of a National-Scale Distributed Intrusion Detection System: Backbone-Level Filtering of HTML Responses in China30th International Conference on Distributed Computing Systems (ICDCS 2010), Genoa, Italy, June 21–25, 2010

[30] George Danezis and Ross Anderson. The economics of censorship resistance. In The Third Annual Workshop on Economics and Information Security (WEIS04), 2004.

[31] Panchenko, A. and Pimenidis, L., 2007, in IPIP International Federation for Information Processing, Volume 232, New-Approaches for Security, Privacy and Trust in Complex Environments, eds. Venter, H. Eloff, M, Labuschagne, L., Eloff, J., von Solms, R., (Boston: Springer), pp. 409–420

[32] "Defeat Internet Censorship: Overview of Advanced Technologies and Products," Global Internet Freedom Consortium, Nov. 2007

[33] Roberts et al, "2007 Circumvention Landscape Report: Methods, Uses, and Tools," The Berkman Center for Internet & Society at Harvard University, March 2009

[34] R. Clayton, S. J. Murdoch, and R. N. M. Watson, "Ignoring the great firewall of china," I/S: A Journal of Law and Policy for the Information Society, vol. 3, no. 2, pp. 70–77, 2007

[35] Garden Networks, January 28, 2010. Retrieved from http://en.wikipedia.org/wiki/Garden_Networks

[36] Tor partially blocked in China, September 27, 2009. Retrieved from https://blog.torproject.org/blog/tor-partially-blocked-china

[37 GPass not working, March 29, 2009. Retrieved from http://www.how-to-hide-ip.info/2009/03/24/gpass-not-working/

[38] OpenNet Initiative (2005). Internet filtering in China in 2004–2005: A country study. Retrieved May 28, 2007
from http://www.opennetinitiative.net/studies/china/.

[39] MacKinnon, R (2008) Flatter world and thicker walls? Blogs, censorship and civic discourse in China. Public Choice. 134: 47–65.

[40] Freedom on the Net: A Global Assessment of Internet and Digital Media, Freedom House, April 1, 2009. Retrieved from
http://www.freedomhouse.org/uploads/specialreports/NetFreedom2009/FreedomOnTheNet_FullReport.pdf

[41] Internet Blocking Exposed, Global Internet Freedom Consortium, July 2002. Retrieved from http://www.internetfreedom.org/files/WhitePaper/InternetBlockingExposed.pdf

[42] Internet censorship in the People's Republic of China, May 2010. Retrieved from http://en.wikipedia.org/wiki/Internet_censorship_in_the_People's_Republic_of_China

[43] IP Tunneling, February 2010. Retrieved from http://en.wikipedia.org/wiki/IP_tunnel

[44] Ford, Caylan, "What Hillary Clinton, Google can do about censorship in China", *The Washington Post,* January 20, 2010. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2010/01/20/AR2010012002805.html

[45] Xin Li, Traci J. Hess, Joseph S. Valacich, Research contributions: Using attitude and social influence to develop an extended trust model for information systems, September 2006, ACM SIGMIS Database, Volume 37 Issue 2-3

[46] Grazioli, S. and Jarvenpaa, S. (2000). "Perils of Internet Fraud: An Empirical Investigation of Deception and Trust with Experienced Internet Consumers," *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, Vol.30, No.4, pp. 395-410.

[47] Gefen, D., Karahanna, E., and Straub, D. (2003). "Trust and TAM in Online Shopping: An Integrated
Model," *MIS Quarterly*, Vol.27, No.1, pp. 51-90.

[48] McKnight, D.H., Choudhury, V., and Kacmar, C. (2002). "Developing and Validating Trust Measures for E-Commerce: An Integrative Typology," *Information Systems Research*, Vol.13, No.3, pp. 334-359.

[49] Figliola, Moloney, Patricia, Nakamura, H., Kennon, Addis, L. Casey, and Lum, Thomas, "U.S. Initiatives to Promote Global Internet Freedom: Issues, Policy, and Technology", Congressional Research Service, April 5, 2010. Retrieved from http://www.fas.org/sgp/crs/misc/R41120.pdf

[50] Joining China and Iran, Australia to Filter Internet (December 2009) Retrieved from http://www.foxnews.com/scitech/2009/12/15/like-china-iran-australia-filter-internet/

[51] Klaus Krippendorff: Content Analysis: An Introduction to Its Methodology. 2nd edition, Thousand Oaks, CA: Sage 2004

[52] Ultrareach.com Accessed June 28, 2009. Retrieved from http://www.ultrareach.com/background_en.htm

[53] New York Times, December 10, 2010, Liu Xiaobo, Online http://topics.nytimes.com/top/reference/timestopics/people/l/liu_xiaobo/index.html

APPENDIX

Trends in Internet Censorship News Coverage

Figure 1 demonstrates that Internet censorship news coverage has been steadily increasing since 1993. The data points along the blue line represent the total number of articles or search results in online news containing the keyword “internet censorship” for each year from 1993 to 2010. The data points on the yellow line represent the total number of articles related to each search result. Together, these two trend lines provide a measure of importance or popularity of internet censorship in the news although the latter provide a more accurate and true representation of the trend in internet censorship. For example, during 2007 a search on http://news.google.com for online news articles containing the keyword “internet censorship” returned a total of 716 articles or search results, and 8,547 related articles. Subsequently, during 2008 there were a total of 710 articles containing the keyword “internet censorship” and 27, 821 related articles. Therefore, even though there were less articles or search results returned in 2008 (710) compared to 2007 (716) there were actually more articles related to the search results in 2008 (27,821) compared to 2007 (8,547). This may seem somewhat ambiguous at first, but the disparity between these two numbers is due to Google’s search engine optimization algorithms. Google attempts to return the most accurate search results for a given keyword and therefore only returns a small subset of results compared to the “total” number of results which actually exist. Google’s filtering process used to return the most accurate search results coupled with the assumption that users will most likely only review the search results on the first couple of pages greatly reduces the total number of returned search results and hence skews the visibility of internet censorship news coverage. Therefore, both trend lines and associated data points are provided to put a more accurate context on the online news coverage and spread of internet censorship.

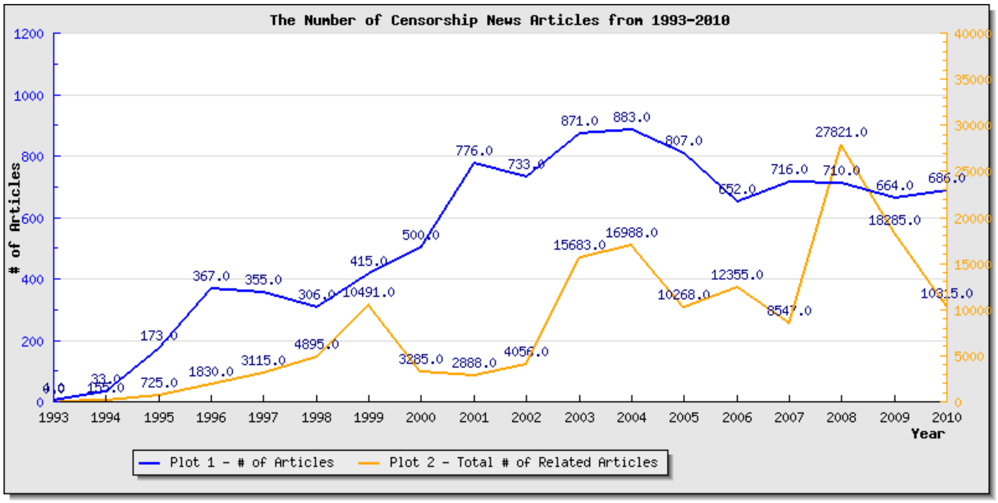


Figure 1 Online News of Internet Censorship

To further illustrate the impact of internet censorship in the news media, we extracted articles from our original data set corresponding to Figure 1 which contained the largest number of related articles. In total

17 articles were extracted and the associated number of related articles are presented in Figure 2. The graph demonstrates an undulation of data points specifically occurring between the 20th and 21 century with the majority of internet censorship news articles occurring in the 21st century. This is a strong indication of the increasing trend and by visually inspecting the graph it can be observed that the article with the highest number of related articles (5,386) occurred in 2009. The specific details for the top 6 articles, sorted by year, which emerged from our analysis, are provided in Table 1.

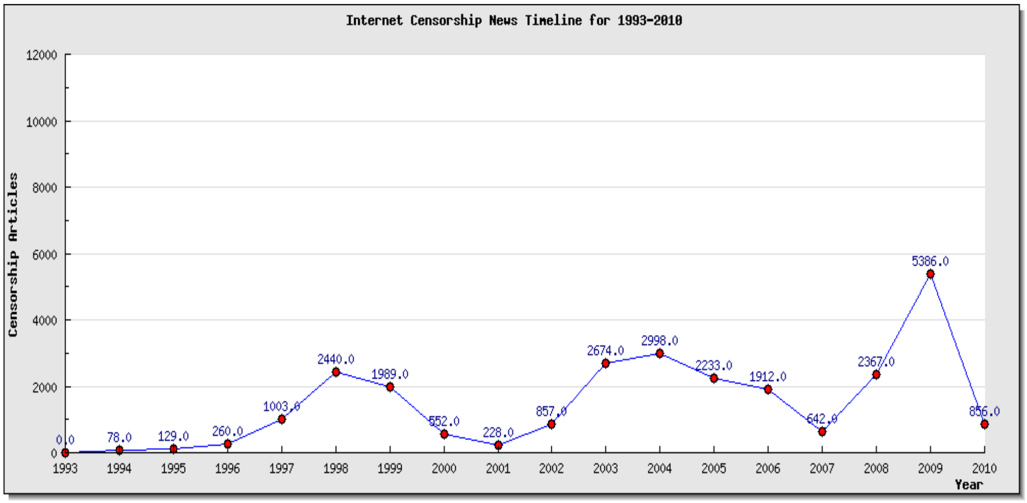


Figure 2 Search Results with Largest Number of Related Articles

While the examination of our data set does not constitute a through content analysis [51] it does provide some interesting and meaningful points of observation. For example, out of the top 6 articles in Table 2, 50% of the articles are related to China and Internet censorship with the article “Obama’s address censored in China” containing the largest number of related articles.

Title	Date	News Source	Related Articles
Starr report: Print follows the Web	1998	CNET	2440
Interview about fate of regime is telling	2003	nwsource.com	2674
What do we do now? - Salon.com	2004	Salon	2998
VATICAN – CHINA Singing the praises of a Pope who loves China	2005	asianews.it	2233
China Partially Lifts Great Firewall for Media, but Access Remains...	2008	PBS	2367

Obama's address censored in China	2009	Washington Times	5386

Table 1 Top 6 News Articles

The Case Study of 2010 Noble Laureate Liu Xiaobo

Liu Xiaobo is political essayist and democracy advocate who has been repeatedly jailed by the Chinese government for his writings. In October 2010 he was awarded the Nobel Peace Prize in recognition of "his long and nonviolent struggle for fundamental human rights in China" [53]. He is the first Chinese citizen to win the Nobel Peace Prize, one of two laureates to have received it while in prison, and the only one who was represented by an empty chair.

In protest of the award, the Chinese government initiated a campaign to censor any information reporting that a Chinese dissident had been awarded the Nobel Peace Prize. A description of a brief analysis of the censorship activity is provided below. A search for the keywords Liu Xiaobo (English and Chinese names) in Baidu (Chinese search engine) was performed in Chinese and overall only news articles about the government condemning the Nobel Prize appeared. Baidu explicitly said that some results are filtered because of legal issues. Searching for Liu Xiaobo in English revealed that censoring was not as prevalent and news articles were available although most were hosted on sites outside of China. In addition, most major Chinese websites removed any mentioning of the Nobel Prize this year (2010) including prizes which were awarded for physics, chemistry, and other fields. The following main points summarize the results of the analysis:

1. Censorship is more severe in Chinese than in English. If we search "Nobel peace prize 2010" in Baidu, we get some results mentioning it being awarded to Liu Xiaobo; but if we search "2010诺贝尔和平奖" (Chinese translation), nothing relevant appears.
2. Keywords are mostly created according to pronunciation, rather than synonyms or other transformations, possibly because this is the easiest for others to understand from the context (or by reading it loud).
3. Leaking censored content is more common in forums than in blogs, possibly because blogs take a "censor then publish" approach, while forums take a "publish then censor" approach. But eventually almost all sensitive content is removed.
4. Censorship is mostly done manually, and delegated to blog/forum admins/moderators, as seen from the fact that the removal times of sensitive content vary for different sites.
5. One way to get keywords is to use Baidu's keyword suggestion function: searching "刘 波" (leaving out second character of "刘晓波" to avoid Baidu's own censorship) gives the following list:
   刘x波刘xiao波波晓刘刘1晓1波刘和谐晓波
   刘x波是谁刘x晓x波刘j晓波中国刘x波刘0晓0波
   some interesting keywords found this way:
   诺基亚和谐奖 ("Nokia harmony prize", first and fourth character same as
   Chinese translation of "Nobel Peace prize")

刘晓庆 (a well-known Chinese actress, Liu Xiaobo's name differ only in last character)

刘晓庆弟弟 ("the actress's brother", to make it more explicit)

6. Baidu has cached most deleted blog/forum entries: from the cache we can view their content and posting times.
7. The most heavily used keywords I have seen so far are "LXB" (from Liu XiaoBo) and "刘晓庆" (clear from context, but also confusing enough for censors).

A search of the following keywords "Nobel peace prize", and "LXB Nobel" was performed using the Chinese search engine Baidu on 10/10/2010 3pm (US EST) and 10/10/2010 10:30pm (US EST), respectively. A total of 130 search results were analyzed. 52 results were not related to the search term and out of the remaining 78 results, 49 results were displayed in forums, 17 results were displayed in blogs and 10 results were hosted on sites outside of China (By identifying and classifying ip address). By clicking on the 78 search results and inspecting the cache in Baidu it was possible to determine if the content had been removed. The results are presented in Table 2.

	Forum	Blog	External Site
Removed?	28	12	0
Not Removed?	21	5	10

Table 2 Ratio of Blocked Content/Platform

In addition to restricting Chinese citizens to freely communicate information regarding Liu Xiaobo, self-censorship in major Chinese news sites is also widely practiced. A list of major Chinese news sites which removed content containing "Noble Peace Prize 2010" is presented in Table 3.

	10/8/2010 8pm (US EST)	10/9/2010 8pm (US EST)	10/10/2010 3pm (US EST)
yahoo china	removed	removed	removed
sina	removed	removed	removed
163.com	removed	removed	removed
tencent	exists	exists	removed
sohu	removed	removed	removed

Table 3 Self-Censorship on Chinese News Sites