

Regional Variation in Chinese Internet Filtering

Joss Wright

Oxford Internet Institute,

University of Oxford

joss.wright@oii.ox.ac.uk

1 Introduction

The internet has become, for over two and a half billion people, a central tool for communication and access to information. The volume of data travelling over the internet and the number of individuals that rely on it make control of the internet a powerful tool for controlling information flow to society.

A large, and increasing, number of nations around the globe participate in some form of internet filtering or censorship (Deibert et al., 2008). Whilst filtering and censorship can, to a limited extent, be somewhat open and transparent, their nature tends towards secrecy. Due to the power of internet filtering, for individuals and society, it is crucial to understand how and to what extent filtering occurs around the world.

Several groups have investigated internet filtering, most notably Deibert (2008) and Herdict (2012). To a large extent, however, these groups have focused on national level filtering – studying the targets of filtering and the approaches taken by an entire country.

National-level filtering, however, is only the crudest form of such investigation. Whilst many states have national filtering policies, it is false to assume that the specific implementation of these policies are necessarily consistent between regions, networks, or even individual computers. In order to fully understand filtering and its role in the globally networked world, it is necessary to explore connectivity at a more geographically and organisationally fine-grained level.

To this end, therefore, this article investigates regional variation in internet filtering, specifically that occurring within China and its variance between cities. By exploring the extent to which filtering is applied, the varying targets of filtering, and the specific experience of users in different areas of China, insights can be gained into the technical implementation and limitations of the national scale filtering infrastructure, the policy decisions and mechanisms that underlie filtering choices, and the effects on individuals.

This article focuses on the case of China for a number of reasons. Firstly, China's national filtering network, the 'Golden Shield Project' (金盾工程, jīndùn gōngchéng) or 'Great Firewall', represents arguably the largest and most technologically advanced filtering mechanism in use today, operating over a total population of roughly 1.3 billion citizens and an internet population estimated at 513 million users as of 2011 (CNNIC, 2012), making the total number of Chinese internet users comparable to the entire population of the European Union.

Secondly, the Chinese state implements a strict and dynamic filtering policy in which globally popular websites, keywords and services are commonly blocked, both in the long-term and in response to political events. At the national scale, economies must be made in the mechanisms of filtering in order to limit the required computational and human resources to a manageable level. Finally, China itself presents a geographically and culturally diverse subject of study. These factors, among others, combine to make China illustrative of the possibilities of broad-scale and thorough internet filtering.

The central finding of this article is that filtering varies to a great extent across China, strongly suggesting that the specific implementation of filtering policies is deferred to local-level organisations. Despite this nation-wide patterns of specific filtering behaviour are clearly apparent, reflecting the existence of detailed filtering decisions that can be imposed on local organisations. A secondary outcome of this work is that certain of these behaviours strongly imply that the censorship mechanism is employed not only to block access to websites, but also to track attempted connections to blocked resources.

This paper proceeds as follows. Section 2 explores existing work that has looked at the theory and practise of filtering around the world, and in the case of China particularly. Section 3 examines approaches to mapping filtering and censorship, and the limitations of existing methods. Section 4 presents the experimental and methodological approach followed in this work. Section 5 presents the results of the experiments, with further discussion and analysis in Section 6. Finally, Section 7 summarises the findings of the experiments and proposes avenues for future work.

2 Existing Work

A number of studies, and ongoing projects, aim to investigate both global internet censorship and the specific case of China. Perhaps the most comprehensive study of global internet filtering has been presented in Deibert et al. (2008). This work is notable not just for its scope but for its focus on the sociological as well as technical aspects of filtering, covering the nature of filtered topics and the levels of state transparency in the filtering process. This work, as with most others, presents only a national view of internet filtering, without considering variations within given target states.

At a similar scale, the Herdict Project (Herdict, 2012) is an ongoing project that relies on users in filtered regions to report blocked websites. This data is aggregated and displayed to users according to country. The Herdict project also provides a web-based tool that presents users with potentially blocked websites for their region, and allows them to report whether the site is apparently blocked.

Morozov (2011) presents a global study of the effects of online surveillance and censorship, and its interaction with internet-based activism and the potential for political change. This work, presenting a pessimistic view of the potential of the internet as a tool for positive political change, explores the various ways in which political actors take advantage of the power of the internet to manipulate and stifle debate. Morozov argues that the balance of power in the use of the internet is typically weighted significantly towards those in power and that internet-based activism, through organisation via social media or more direct social disobedience, is a largely ineffective use of the effort of activists.

MacKinnon (2012) presents a more positive view of the potential for internet users to act collectively online in order to negotiate and demand change in their access to information and services. MacKinnon surveys the state of global information control and filtering, and the responses of individuals to the controls imposed on their access to information. The focus of the work is largely, but by no means entirely, on the actions of corporations as providers of services, and the broad scale decisions made either in collaboration with governments or otherwise. MacKinnon argues that such corporations should consider, and be held accountable for, the ramifications of their business decisions. This echoes arguments made, in a less adversarial setting, by Pariser (2011), who explores the algorithmic filtering imposed by corporations in an attempt to customise content for users, leading to a 'filter bubble' that restricts the opinions and informations sources presented to a user, and outside of which that users does not see. Both of these works largely assume the power and efficacy of the internet as a major source of information for individuals, in contrast to the more pessimistic view of Morozov.

In direct relation to the theme of collective action, King *et al* (2011) develop a theoretical framework for understanding the state-level motivations of internet filtering in China, through empirical analysis of the patterns of censorship and takedown occurring on Sina Weibo, the extremely popular Chinese microblogging service.

Weibo posts are directly removed by the Chinese authorities when they are deemed to have violated unwritten rules of conduct. By examining the nature of removed posts, and how quickly authority responds to particular topics of discussion, King *et al* argue that Chinese censorship is not, as typically assumed, occurring to prevent *state critique* of the Chinese Communist Party, but is instead predicated on preventing discourse that leads to collective action – a reducing the *collective action potential* of the society while still allowing critique of the state. This theory has interesting implications for the potential, investigated in this article, for regional variation in censorship to be adapted for regions with different social or economic makeup.

The use of the internet in the specific case of China is examined in detail by Yang (2009), who highlights the role of the internet as a force for activism and dialogue for Chinese citizens. Most interestingly, it is claimed that while the central government is intolerant towards speech that aims to criticise the state as a whole, it is far more lenient with criticism of local-level government and officials. The view, backed up with specific examples, is again a potential driver for the regional variation in filtering investigated in this article.

There have been several technical studies of the means employed to filter and surveil internet connections. A useful overview of the technical scope of filtering technologies was presented by Murdoch & Anderson (2008) who organise internet filtering according to four major categories with varying degrees of complexity and efficacy. These approaches, and in particular DNS filtering which is key to the work in this article, are discussed in greater detail in Appendix B.

A seminal study of one major technique employed by the Chinese national firewall was presented by Clayton *et al.* (2006), who identified the specific mechanisms by which the firewall instructed both ends of a communication to abandon their connection when objectionable content was detected, and proposed that by ignoring this instruction, specifically a TCP RST (“reset”), Chinese filtering could potentially be 'ignored' by appropriately modified systems.

Another early study of the Chinese filtering infrastructure was carried out by Crandall *et al* (2007) which made use of direct requests to identify particular keywords blocked by the Chinese firewall, along with an approach based on latent semantic analysis to identify previously unknown keywords related to know topics. The analysis in this work not only identified a broad range of topics that fell within the scope of the Chinese filtering regime, along with specific keywords, but also that different topics may be blocked for differing amounts of time according to the perceived severity of the topic. Importantly, this work also identifies that the Great Firewall acts, in the author's words, as a 'panopticon', with filtering occurring within China rather than simply being located on the interface between China and other countries. The strong implication is that China's 'firewall' is as concerned with control of information within the country as with information flowing across its borders.

Looking more directly at the infrastructure of filtering, Xu *et al* (2011) examine where filtering occurs in the China in terms of the structure of the networks that make up the Chinese internet, and how these networks interact with networks of neighbouring countries. Xu *et al* seek to identify particular networks and *border routers* – routers that are placed on the border between China and its neighbours – where the specific action of filtering takes place. These results can be used to identify, at a relatively coarse-grained level, which major internet organizations within China are responsible for hosting the majority of the filtering infrastructure.

The work of Xu *et al* identifies that the infrastructure of filtering is relatively well dispersed across China in both a geographic and organizational sense, albeit with a particular concentration in Guangdong. This work, which is perhaps closest in its nature to the work presented in this article, supports the view of a distributed, decentralized and partially 'outsourced' filtering infrastructure.

3 Mapping Filtering

To understand regional variation in internet censorship, it is necessary to obtain network measurements from multiple points within a target region. Several existing projects involve an aspect of mapping, either at a logical or

a geographical level. The Herdict project (The Herdict Project, 2012) allows users to report apparently blocked websites, via a browser plugin, using this data to present both a global map of filtered sites and a per-country breakdown of those sites most commonly reported. The Alkasir project (Al-Saqaf, 2012) combines user-based reporting of blocked content with an anti-censorship tool that attempts to penetrate such filtering. A relatively new project at the time of writing, the Open Observatory of Network Interference (OONI), seeks to develop and deploy an open network of monitoring tools managed by volunteer operators that would allow for active monitoring of global filtering (Filastò, 2012).

To bypass internet filtering, perhaps the most well-known technology internationally is the Tor Project (Dingledine et al., 2004), which allows users to reroute their connections through a global network of volunteer-run anonymising proxy servers. This network, originally designed to preserve the connection-level privacy of users, was found to be an excellent tool for bypassing national filtering and now invests significant resources in supporting this use. Similar tools include Psiphon (Psiphon Inc., 2012) as well as numerous Virtual Private Network (VPN) servers that allow users to evade national filters. All of these services work in a similar manner: by rerouting a connection through a server located in a different country, the user experiences the internet as if their connection originated in that country. Thus, a user from Saudi Arabia is able to route their connection through a US computer and bypass all filtering imposed by their state or organization, at the cost of some slowing of their connection and gaining any filtering or surveillance, if any, imposed by the US or the provider of the proxy.

From these examples, two major possibilities for studying internet filtering can be identified. The first is to ask users in a given country to report their experience, as exemplified by the Herdict project; the second is to make use of an available service, such as a Tor node or VPN server, in that country to experience the filtering directly. Both of these approaches have limitations when applied to large-scale measurement of filtered networks.

The advantage of using a system such as Tor, Psiphon or VPN services is that they allow a researcher to control traffic directly. Sites of interest and even specific patterns of traffic can be directly sent and examined, allowing a much more detailed examination of the technical measures employed on a given network. The approach taken by Herdict, however, cannot currently reproduce this level of sophistication. In the absence of a large network of experienced and technically capable users, user-level reporting only provides that a site appears to be unavailable, without reference to the conditions that cause the unavailability¹.

In order to achieve a fine-grained mapping of filtering, therefore, there are two major points of interest beyond those commonly considered by the most well-known current mapping projects. The first of these is the precise geographical location of a particular computer. The ability to determine the originating country of an IP address is relatively well known, and location to the level of an individual city can be achieved with some accuracy. Recent results from Wang et al. (2011) have proposed mechanisms that achieve a median accuracy of 690 metres, albeit within the US. This simple extension would provide a valuable source of data on the applications of filtering. In many cases it is also possible to determine which organisation has been allocated any particular IP address, to the level of an ISP or major company, as partially explored by Xu *et al* (Xu, 2011). Both of these pieces of information can be used to build up a much more detailed view of filtering.

The second point of interest is to study, in detail, the technical nature of the filtering that is imposed on a given connection in a given location. While work has been conducted into specific methods, as in the work of Clayton et al. (2006) relating to the Chinese national filter, most large-scale projects appear to be focused more on the existence of filtering rather than the details of its implementation.

1 The Herdict project does allow a user to express their opinion as to the cause of the blocking, but in the absence of direct experimentation this data has significant limitations.

4 Experimental Approach

This work seeks to discover and map variations in internet filtering between geographical regions of China, as well as to determine the particular nature of the filtering that occurs. This aims to provide insight into the nature of filtering decisions, in terms of centralization and control, as well as the specific implementation of filtering as it is devolved to local-level actors.

The approach taken is therefore largely data-driven, and focuses on obtaining a geographically diverse source of data regarding the filtering observed by Chinese internet users. To achieve this, in light of the various filtering approaches that exist, both within China and globally, drawing data from the DNS service provides a rich and relatively accessible source. The DNS provides a number of attractive features for technical reasons, but also avoids a number of legal and ethical concerns that are detailed in Section 5.6.

Due to their crucial role in resolving names to IP addresses, DNS servers are common and widespread, providing a desirable level of coverage that is difficult to match with most other direct approaches. DNS servers are also typically openly accessible, meaning that there is no technical restriction in making requests to these remote systems.

DNS servers are also attractive in that they are typically run either by internet service providers for the benefit of their customers, or by large organizations that run their own networks. The result of this is that the results returned by a given DNS server typically reflect the view of the internet, at the level of DNS, of a reasonably large class of users.

From a legal and ethical point of view, DNS servers have the advantage of functioning, at an extremely simple level, as a simple database of mappings between domain names² and IP address. As such, requesting information regarding a given domain does not cause any direct access to potentially sensitive resources on behalf of a third party³, as would be the case for the proxy services mentioned above. This, importantly, avoids involving any third parties, willing or otherwise, in experiments.

To obtain a useful sample for investigating DNS censorship across China, a list of DNS servers was retrieved from the Asia Pacific Network Information Centre (APNIC), the Regional Internet Registry responsible for allocating IP addresses and Autonomous System (AS) numbers across the Asia Pacific region. This organization maintain a database, known as a WHOIS database, that stores information regarding registered domain names in their region, including the authoritative DNS servers for each domain. From the WHOIS records, a list of 278 DNS servers apparently located in China, according to our geolocation service, was retrieved of which 187 were found to be available and responsive to remote queries.

In order to relate particular DNS servers to their physical location, the freely-available MaxMind GeoIP database (MaxMind Inc., 2012) was employed to resolve IP addresses to their city of operation, allowing identification of the location of almost all DNS servers in the test set. It is worth noting, as will be discussed later, that this does not represent the location of the *users* of that service; these users make DNS requests from their home network connection, and could potentially be located in almost any geographical location, but will in practice almost certainly be within China.

2 Strictly speaking, DNS servers return the IP for a particular *hostname*, many of which may exist under a given domain name. For the purposes of this article, the two may be considered functionally equivalent as request were not made for multiple hosts within a single domain.

3 For completeness, it should be mentioned that DNS servers function in a hierarchy, and may request information for unknown domain names from more authoritative servers. This normal function of the service would not, however, implicate any third party, and would in fact be directly traceable to the computer used in our experiments.

A list of potentially filtered domain names was obtained from the Herdict Project's website (The Herdict Project, 2012). As the Herdict project receives reports regarding filtered websites from crowdsourced reporting, sorted according to country, this provides a useful source of data for potentially blocked domain names. An alternative approach which is perhaps less efficient, but potentially more revealing, would be to obtain a list of the most commonly visited websites worldwide, such as the Alexa rankings (Alexa, 2012).

The Herdict project lists the most frequently reported blocked websites for each country, each list comprising the top 80 reported domains. In addition to this, five popular Chinese websites were included in the test set that, presumably, would not be blocked in mainland China. A full list of tested domains is given in Appendix A. To learn the scope and scale of blocking, each potentially-blocked domain name in the list retrieved from Herdict was requested from each DNS server retrieved from the APNIC WHOIS database. These results were recorded and analysed according to the nature of the DNS response received in each case.

In order to determine whether the results returned were genuine, an equivalent query was conducted on a self-managed DNS server located in a country that does not perform extensive internet filtering⁴. The results of the remote query were compared heuristically⁵ with the local result, any differences were noted.

In order to minimise genuine short-term network errors, the sequence of requests was repeated six times at one-hour intervals. The results from the different experimental sets were combined in such a way that timeouts, which could represent genuinely poor connectivity, were eliminated unless they were seen to be consistent across all result sets.

4.1 DNS Response Types

The categorization presented in Appendix B identifies four major filtering techniques, of which DNS poisoning is only one. DNS poisoning can, however, occur in a number of forms, some of which could represent genuine errors rather than censorship. The most important behaviours of a DNS server, for the purposes of this article, are discussed here.

Invalid Server Errors

A DNS server, on receiving a given query, may respond with an indication that it is not, in fact, a DNS server. In this case, the requesting party will not receive a mapping from the requested name to an IP address, and thus cannot proceed with making a connection. Clearly, such a response could also indicate that the requested party was genuinely not, in fact, a DNS server.

Timeout Errors

4 In the case of the experiments detailed here, this was the United Kingdom. Whilst the United Kingdom certainly does engage in national-scale internet filtering, it does not in general involve this particular form of DNS manipulation, and care was taken that such filtering would not affect the results of these experiments.

5 Specifically, the first two dotted quads of the IP addresses returned by the remote and the local DNS server were compared. If these differed, the response was marked as incorrect. Large internet services often make use of dedicated content distribution networks such as Amazon Web Services that employ a wider range of IP addresses. Experimental results were manually examined to detect any such networks, and any domains that resolved to these networks were assumed to be accurate responses. This automated approach does allow a chance of introducing both false positives and false negatives with respect to the existence of misleading DNS results, but this risk is relatively small and should not unduly skew the overall results.

A simpler behaviour, and one that is harder to categorize unambiguously as censorship, is for a DNS server to accept requests, but not to respond in any way for blocked domains. Eventually, the requesting party will exceed a given time threshold and abandon the query. This again prevents the client from learning the IP address of the requested host, and could be ascribed to a genuine network error. A secondary effect of such an approach is that the requesting party does not receive an immediate response, which may cause internet requests to blocked sites to pause until the timeout threshold is reached.

Unknown Domain Errors

The simplest form of direct DNS censorship is for the DNS server to deny the existence of the requested website, causing the requesting party to receive an error. For known existing domains, this response is easily identifiable as malicious behaviour on the part of the server.

Misleading Results

A more subtle approach to censorship is for requests for blocked websites to generate a valid DNS response, providing the client with an IP address for the requested hostname, but to provide false information in the form of an incorrect IP address.

This approach has several potential implications, which will be discussed further below. One potential outcome of such an approach is that the requesting party may be directed to a host that logs all attempts to access banned websites, allowing for a level of surveillance or monitoring of such requests.

Genuine Results

A final possibility is that a DNS server returns the correct IP address that corresponds to the requested hostname. While this particular piece of information may be accurate, censorship may, of course, occur through other means.

5 Results

The results of querying DNS servers across China for reportedly-banned domain name are presented below, along with a number of identified trends in the responses. A number of particularly unusual observed behaviours are also highlighted and explored.

5.1 Broad Trends

Overall, experiments were conducted on 187 DNS servers across China, 178 of these servers answered at least one query with a valid, but not necessarily truthful, IP address. Of the responding servers, 79 answered at least one query with a response that appeared to be accurate, meaning that 99 servers returned only invalid results for the requested domains.

A small number of servers were clearly either misconfigured, or deliberately providing invalid results to requests. Five servers consistently timed out on DNS requests, despite an allowance for an artificially long timeout period of 60 seconds. One server consistently produced an invalid nameserver error, despite apparently accepting DNS requests.

Widespread DNS Poisoning

The experiments provide evidence of widespread manipulation of DNS results, occurring in all the forms discussed in the previous section. Interestingly, individual DNS servers do not, in general, display consistent blocking behaviour across all domains, but may instead return an incorrect IP address for one domain, claim that a second domain does not exist, and refuse to respond to requests for a third domain.

Domain	No Domain	No Answer	No Nameserver	Timeout	True IP	False IP
www.backchina.com	0	0	13	7	5	162
www.ntdtv.com	0	0	23	7	0	157
www.open.com.hk	0	1	20	7	3	156
www.torproject.org	0	2	24	7	1	153
www.tibet.net	0	2	22	7	3	153
www.peacehall.com	0	1	20	7	6	153
www.6park.com	0	0	26	7	2	152
www.hotspotshield.com	0	1	29	7	2	148
www.boxun.com	0	1	29	7	2	148
wezhiyong.org	0	1	33	7	2	144

Figure 1: Top misdirected domains from experiments, showing DNS error result counts for each domain.

Figure 1 demonstrates the ten most widely misdirected domains observed in experiments. These domains were thus almost universally blocked across China at the DNS level. It should be noted that, in addition to an overwhelming majority of misleading results for each domain, the remaining servers were likely either to timeout or to claim not to be a valid nameserver for this result.

Domain	No Domain	No Answer	No Nameserver	Timeout	True IP	False IP
www.ahrchk.net	4	17	64	40	60	2
killerjo.net	4	17	65	37	62	2
www.x365x.com	3	17	65	41	59	2
www.websitepulse.com	3	18	65	36	63	2
www.voanews.com	3	17	64	38	63	2
www.tumblr.com	3	17	64	38	37	28
www.steves-digicams.com	3	17	65	36	64	2
www.scribd.com	3	17	65	36	38	28
www.pinyinannotator.com	3	18	67	36	61	2
www.newgrounds.com	3	16	64	36	66	2

Figure 2: Ten domains most often claimed non-existent.

Figure 2 lists the domains most often claimed not to exist by the tested DNS servers. As can be seen, claiming a domain to be non-existent is far less common than providing an inaccurate IP address result. It is worth noting that the domains listed in Figure 2 receive large numbers of timed-out requests, as well as both accurate and inaccurate IP responses.

These results suggest that approaches to DNS poisoning favour misdirection of domains over claims that the domain does not exist, and that allowing a request to timeout by not responding, as opposed to generating an error, is also common approach.

Timeout Responses

The prevalence of timeouts could potentially be explained by filtering occurring not directly at the DNS servers, but instead at other points in the network. An explanation for this would be that requests for certain domains are not allowed to pass across the Chinese network, and are instead silently dropped by intermediary filtering routers either before or after reaching the server in question.

To understand this, it is useful to comment on the underlying *transport* of the DNS protocol. DNS typically makes use of an efficient underlying internet transport protocol known as UDP; this is in contrast to the widely used TCP protocol employed by the majority of common internet services such as the world-wide web. UDP has the advantage of higher speeds and lower transmission overheads than TCP, but does not provide reliable data delivery, nor does it implicitly confirm the success of data transfers. It is therefore the case that, if DNS requests for particular domains were blocked or dropped in the network, it would be difficult to detect this fact; the result would be observed simply as a timeout.

Another alternative is that the DNS servers in question, upon receiving a request for a blocked domain, simply ignore the request. From the experiments detailed here, it is difficult to verify either of these claims. It seems likely, however, that filtering of DNS traffic in transit to block requests would be more complex and costly, and would result in a more homogeneous and extensive pattern of timeouts than were observed. As such, the argument for filtering at the server level, or some combination of both arguments, appears most likely.

Common Misleading IP Addresses

An examination of the results returned from the experiments show that, in the case that a DNS server returns an IP address that does not correspond to the requested domain, the returned IP address is drawn from a comparatively small pool of possible responses; misleading IP addresses are neither random nor returned on a per-server basis.

The experiments detailed here made requests for 85 domains to 187 DNS servers, resulting in a total of 15,895 requests in total. Of these requests 6658 gave a response that pointed to an IP address, 2258 of which were judged to be misleading. These 2258 misleading results each pointed to one of only 84 IP addresses, showing significant correlation between misleading IP addresses returned by DNS servers across the country.

Two possible explanations exist for this result. The first is that a centralized list exists that provides specific DNS poisoning instructions, including IP address, for DNS server operators. The second possibility is that the observed DNS responses, being conducted outside of China and therefore travelling across the nation's border routers that are known to engage in substantial filtering, were manipulated in transit. Investigation of these two possibilities is an intriguing subject for future work.

5.2 Domain Statistics

Poisoning of Uncensored Domains

In addition to the list of 80 domains obtained from the Herdict project, the experiments incorporated domain names for five popular Chinese internet services with the intention that these would be unfiltered. Surprisingly, in several cases results appeared to show misleading results for a number of these domains. This could be due to misconfiguration of DNS servers, deliberately invalid results returned due to the request originating outside of China, or some other cause.

An illustrative example is that of `renren.com`, a popular Chinese social network. In at least two cases, invalid IP addresses were returned for this service, as shown in Figure 3. In this example, the two servers are located in different cities, and are apparently operated by separate companies; the two servers are both logically and physically distinct. Despite this, both servers return the same list of IP addresses, none of which appear to belong to servers of `renren.com`.

On directly querying the addresses in question, a number of them appear to be running an unconfigured webserver. It is not known what, if any, the significance of these addresses may be.

Server	Location	Remote Result
202.95.0.10	China, Beijing	renren.com. 900 IN A 123.125.38.2 renren.com. 900 IN A 123.125.38.3 renren.com. 900 IN A 123.125.38.239 renren.com. 900 IN A 123.125.38.240 renren.com. 900 IN A 123.125.38.241
121.101.208.41	China, Chaoyang	renren.com. 900 IN A 123.125.38.2 renren.com. 900 IN A 123.125.38.3 renren.com. 900 IN A 123.125.38.239 renren.com. 900 IN A 123.125.38.240 renren.com. 900 IN A 123.125.38.241

Figure 3: Inaccurate results for `renren.com`. Distinct servers show identical incorrect results.

Purposeful Misdirection of torproject.org

The Tor Project produce a number of tools that aim to provide anonymous and untraceable internet communications, as well as to bypass censorship. As such, both the tool and the project website are commonly blocked in countries with extensive internet censorship, and are engaged in an ongoing filtering arms race with the operators of the Great Firewall.

On querying servers for the Tor Project's website, `www.torproject.org`, a set of consistent misdirections was found in DNS server responses from multiple organisations and geographical locations; a total of twenty-nine responses all redirected Tor Project traffic to a unique alternative domain. The subject of this misdirection was <http://www.thepetclubfl.net>, a pet grooming service in Florida. On contacting the webmaster of this site it was confirmed that the site has for some time been experiencing a previously unexplained large volume of Chinese traffic. This confirms that these particular redirections do not appear to undergo subsequent blocking within China. The possible purpose of these redirections is explored in Section 6.

The redirected domains is, of course, not in any discernible way linked to the Tor Project. The number of results from disparate servers all pointing to the same domain strongly imply some broader connection, either through direct instructions implemented by local servers or, potentially, through sharing of blocklists between multiple organisations.

It should also be noted that this redirection was not the only set of consistent redirections noted for the Tor Project. The owner of a second domain similarly receiving traffic destined for the Tor Project has requested the author not to publicise their domain. While this request has been respected, the domain in question is a particularly interesting case as the redirections occur not simply to a single domain name, but to a number of similarly-named domains with different suffixes, such as `<domain>.com`, `<domain>.net`, and `<domain>.org.ez-site.net`. This again implies that the redirection in question is in response to instructions from a third party.

5.3 Server Statistics

Misleading Results

The majority of servers queried returned a mix of result types, with varying degrees of misleading results. A small number of DNS servers, however, demonstrated an unusually extreme range of negative responses, and are thus demonstrative examples of the invalid responses given.

DNS Server 113.11.192.25

This server is apparently located in Beijing. Over the course of 85 domain name requests, this server responded with 'no answer' a total of 68 times. This included the five presumably unfiltered services, including Baidu and RenRen, in the test set, and may indicate discrimination against requests located outside of China.

A further 13 requests resulted in the return of a valid IP address. On examination, all of these IP addresses were found to be unassociated with the requested domain. The list of domains and associated IP addresses can be found in Figure 5.

DNS Server	Location
122.102.0.10	China, Chaoyang
159.226.8.6	China, Beijing
162.105.129.27	China, Beijing
182.50.116.252	(Unknown)
202.102.224.94	China, Henan
202.115.32.39	China, Chengdu
202.127.12.8	China, Nanjing
202.99.216.75	China, Xian
202.99.96.126	China, Tianjin
211.161.46.86	China, Beijing
221.13.28.234	China, Guiyang
221.7.92.99	China, Chongqing
59.63.158.124	China, Beijing

Figure 4: Example torproject.org requests resolving to alternative domain thepetclubfl.net.

The nature of the IP addresses in question are of some interest. There is no discernable pattern in these results; they point to seemingly-random hosts corresponding to domains and organisations that do not appear connected with each other or with the originally requested domain. It is notable, however, that certain of the blocked domains point to the same IP addresses, even though those IP addresses are not related to the domain in question. As can be seen from Figure 5, peacehall.com and wujie.net, and backchina.com, boxun.com and open.com.hk redirect to the same IP addresses as each other.

The remaining four domains requested from this server resulted in a claim that no such domain existed.

Domain	Returned IP
www.hotspotshield.com	8.7.198.45
www.tibet.net	159.106.121.75
www.boxun.com	46.82.174.68
www.wezhiyong.org	8.7.198.45
www.backchina.com	46.82.174.68
www.ntdtv.com	8.7.198.45
www.peacehall.com	59.24.3.173
www.wujie.net	59.24.3.173
www.6park.com	159.106.121.75
www.open.com.hk	46.82.174.68

Figure 5: Misleading IP addresses from a Beijing-based DNS server.

DNS Server 202.99.224.203

This server is apparently located in Baotou. Of 85 domains, the majority of results were to claim that the server was not valid for returning DNS requests. In total, requests for 14 domains resulted in one or more IP addresses being reported, none of which led to the appropriate servers. This behaviour, to appear invalid for some domains and to return fake results for others, was particularly strange.

It could once more be observed that, although invalid IP addresses were returned, these were not purely random but instead were consistent for each domain, and were drawn from a small pool of IP addresses that were used multiple times for different domains.

Localhost Redirection

An interesting choice of address to return when providing inaccurate IP addresses is to point the request back to the computer from which it originated. This can be achieved through use of the special 'reserved' IP address 127.0.0.1, which also has the DNS designation of 'localhost'.

Local redirection has the advantage of not requiring a genuine IP address to be selected from the internet, which can lead to undesirable behaviour. It also minimises traffic passing over the internet, as any further requests made to this connection remain on the user's computer without travelling over the general internet.

Despite this, the use of redirection to the localhost was not particularly widespread amongst the queried servers. Of the 187 servers queried only six servers returned results pointing to the localhost, of which four consistently returned the localhost for any DNS query. This could represent either a misconfigured DNS server, or a blanket policy for unauthorized or non-Chinese requests.

Two servers, however, with addresses 202.99.224.200 and 202.99.224.223 returned 127.0.0.1 for the majority of requests, but also resulted in an invalid nameserver error for seven domains. In a further 13 cases an IP response was given that, again, appears random, resolving to Azerbaijani, Irish, US, Italian, and New Zealand based hosts.

5.4 Geographical Distribution

Figure 6 presents an overview of the variations in filtering observed across the various cities covered by these experiments. Darker grey markers represent a greater percentage of misleading DNS responses compared to accurate responses. As results were obtained for potentially many servers within a given city, the median average percentage of all results observed for all servers in the city is represented. To indicated cities with a larger number of DNS servers, markers are log-scaled according to the number of servers tested, ranging from a single server in cities such as Dongguang and Harbin, to 72 servers in Beijing.



Figure 6: DNS queries across China showing median percentage of misleading results for queried domains, with darker points representing a higher percentage of misleading results. Circle size represents the relative number of servers queried in each city.

No overall pattern of filtering has been detected in the results for different cities or regions, however there is clear heterogeneity across the country. This supports the view, as suggested by Xu *et al* (Xu, 2011) and others that high-level controls over filtering are relatively loose in terms of implementation of filtering, with the technical details of blocking being decided at the local rather than regional or national level. The implications of this are discussed further in Section 6.

5.5 Experimental Limitations

There are a number of limitations to the experimental methodology employed in this work. The first and most obvious is that the experiments relied on a restricted list of DNS servers obtained from the APNIC WHOIS database. While the set of servers used provided a reasonable coverage of China geographically, with a notable bias towards the East of the country due to the higher density of population and development, there was a great disparity between the number of servers observed in each city. This figure ranged from 72 servers in Beijing, to only one server in several of the smaller cities. While this will, to some extent, reflect the realities of DNS server placement in China, it appears insufficient for a genuine analysis of the relative experience of internet users. A more fundamental limitation is that DNS servers are not necessarily, or usually, located in the same geographical area as a user. A DNS server is typically operated and managed by an ISP, and made available to its users automatically. It is therefore likely that a given ISP's customers, who may be widely dispersed, all use the same DNS server. As such, the results presented here arguably represent *organizational* variation, rather than geographical.

Further, the results in this article represent a snapshot taken in mid-2012, and as such cannot reflect changing patterns of censorship. Given the automated nature of these tests, however, and the relatively short time required

to conduct them, the gathering of time-series data is a relatively small step, and has the potential to reveal useful patterns of censorship over time, which would be of significant value in observing the relationship between real world events and the extent to which these dictate or affect filtering policy.

The final major limitation to this work is that it provides a purely technical view of one form of filtering occurring in China. These results provide a window into the limitations imposed on users' internet connections, but can provide little data with respect to the effects of censorship on users' browsing behaviour, social attitudes to various forms of content, choice of forums, modes and means of communication, and access to news sources. As such, the experiments detailed here provide only a limited first step in understanding the wider phenomenon of internet censorship.

5.6 Ethical Considerations

Deliberate misuse of a network service for the purposes of detecting internet filtering may be illegal in many jurisdictions; clearly, such misuse without a user's consent is unethical. Even when using openly available and general purpose services, however, there are serious considerations when attempting to access blocked content via a third party.

A user is unlikely to face repercussions for being seen to be attempting to access blocked content. The scale of internet use, even in smaller countries with low internet penetration rates, is simply too high for there to be serious policing of users who request filtered content. It is likely that, in the vast majority of cases, such attempts may not be logged at all, however, as will be discussed in Section 6, there is evidence that DNS requests for blocked services may be logged by the Chinese filtering apparatus.

By their nature the filtering detection mechanisms that have been discussed detect filtering through attempts to access filtered content: websites or IP addresses that are known, or are believed or likely, to be banned. It is possible that sufficiently high-volume requests for banned content may be considered worthy of further unwelcome scrutiny.

Volunteers that participate in research of this nature by running a filtering detection tool must do so having been fully informed as to the nature of the tool and the potential risks involved. From this perspective there is a significant added burden on the researcher to state to the participant, who may well not have any significant level of technical expertise, what the tool will do and what particular risks they run.

6 Discussion

The power of the internet as a tool for freedom of expression and communication has been widely recognized (Yang 2009; MacKinnon 2012), even though some doubt its efficacy at achieving genuine political or social change (Morozov 2011). What is clear in the case of China, through these experiments and others, is that the Chinese state is willing to expend significant resources in maintaining a level of active control over the flow of information across its national networks.

An interesting result in this respect is that, at the DNS level, there is a clear lack of *overt* filtering, which might be expected in the form of claiming that given domains are non-existent. Instead, across the vast majority of servers investigated, misleading IP addresses were provided in response to queries. The practical results of these misdirections, from the perspective of a user, are varied. In some cases, an alternative webpage may be displayed, as in the case of the Tor Project and its redirection to the websites of Tony Castro and a pet grooming service in Florida; in other cases, where the misdirection does not point to a valid web server, an error message will be displayed to the user. To a large extent, however, there is no explicit statement of filtering.

One feasible explanation for this behaviour is that when a DNS response results in an invalid domain, the user's computer immediately terminates the connection – it has no valid address to reach. By providing an alternative

address, the firewall ensures that the user's computer will make *some* connection. If, as the experimental results show, this connection is to a computer located outside China, it can be guaranteed that the user's connection will pass across the *border routers* of China, where a significant level of filtering is believed to occur (Xu, 2011).

This behaviour strongly supports the hypothesis that the purpose of such misdirection is to ensure that users who attempt to connect to banned services make observable connections through known routers. It has been established in the course of this research that, at least in the known examples, that connections to the falsely reported IP addresses do complete successfully. It seems likely that this behaviour allows for surveillance and monitoring of users attempting to reach censored information.

The focus of the experiments conducted in this work is on the regional variation in filtering. As clearly demonstrated in Figure 6, there is a wide diversity of filtering across China, although no overall geographical patterns have been identified in the broad-scale filtering results. The extent and variance in filtering, however, demonstrates a significant involvement of local actors in making the low-level filtering decisions, even if these reflect broadly-stated guidelines such as requests to block particular topics or websites.

While the patterns of filtering identified in these experiments are more interesting in terms of the correlations across the country, they also demonstrate a clear capability for differentiated filtering to occur on demand in different locations. An important aspect of future work in this area is to identify the development of filtering decisions over time, to observe the rapidity of response as filtering decisions pass from central government policy to regional implementation.

7 Conclusions

This article has proposed that it is, in general, false to consider internet filtering as an homogeneous phenomenon across a country, and that the practicalities of implementing a filtering regime are likely to result in geographical and organisational differentiation between the filtering experienced by users. The experiments conducted in the course of this work support this hypothesis – filtering varies widely from region to region. The study of these differences are of great interest in understanding both the technologies and the motivations behind filtering, and have proposed a number of mechanisms that could be employed to gain this understanding.

In response to the technical and ethical challenges of censorship research, these experiments represent a nationwide remote survey of the apparent filtering experienced by Chinese internet users, with specific reference to blocking attempts that occur through the Domain Name Service (DNS). These experiments have revealed widespread poisoning of DNS results, including invalid server responses, valid domains claimed to be non-existent, and the return of IP addresses that do not correspond to the requested domain.

Analysis of these results has revealed a number of trends in this filtering, most notably the prevalence of misleading responses for domains over claims that domains do not exist. Further, although the extent of filtering varies geographically, frequent correlations have been observed in the misleading IP addresses returned in response to requests for blocked domains by different servers, implying some level of top-down involvement in the behaviour of servers.

The nature of filtering experienced at the level of the DNS reflects a clear preference for misdirection rather than direct blocking; it is possible that this is partially due to a desire to 'soften' filtering by making it appear more akin to a network error than overt restriction. As suggested by the consistent misdirection to given IP addresses for certain domains, and the subsequent evidence that these connections are not themselves blocked, there is reason to believe that a level of surveillance operates in addition to more direct blocking.

The experience of internet filtering in China lends support to a limited form of Morozov's argument (Morozov, 2011) that the potential for activism via the internet is severely challenged by the capabilities that it provides to entrenched holders of power. Despite this the scope and extent of filtering apparent from these experiments, and the low levels of filtering seen in some regions, suggests that an ongoing fine balance is being struck between the desire to filter and the need to allow certain freedoms. This aligns with the technical data presented by Crandall *et al* (Crandall, 2007), and the theoretical developments of King *et al* (King, 2012), that demonstrate varying responses to different classes of censored topic.

It is clear from these experiments that the infrastructure of filtering in China is complex, and is managed and operated by multiple actors imperfectly communicating with each other. Centralized decision making is evident in some cases, local in others. While the system aims to control the flow of information both within the country and across its borders, it provides a potentially deep insight into the policies of China's decision makers, and the actions and reactions of its citizens.

Appendix A Tested Domain Names

A.1 Potentially Blocked

Retrieved from the Herdict Project:

www.torproject.org, www.google.com, mail.live.com, www.blogger.com, dropbox.com, www.wretch.cc, vimeo.com, www.scribd.com, anchorfree.com, developer.android.com, www.gmail.com, www.demonoid.com, www.bing.com, thepiratebay.org, piratebay.org, www.hotspotshield.com, www.box.net, mail.google.com, chinagfw.org, blogspot.com, wikileaks.org, www.tibet.net, www.boxun.com, www.bbc.co.uk, wezhiyong.org, www.bullogger.com, www.rfa.org, wikileaks.com, www.backchina.com, huffingtonpost.com, www.ntdtv.com, www.rthk.org.hk, www.aboluowang.com, www.voanews.com, www.wenxuecity.com, www.dw-world.de, zh.wikipedia.org, www.danwei.org, news.bbc.co.uk, www.peacehall.com, www.youtube.com, www.facebook.com, twitter.com, www.wujie.net, www.6park.com, www.steves-digicams.com, www.hotmail.com, www.x365x.com, www.wenku.com, picasaweb.google.com, www.camfrog.com, www.tumblr.com, www.foursquare.com, www.imdb.com, flickr.com, t.co, www.livejournal.com, www.twitzap.com, killerjo.net, www.paltalk.com, www.pinyinannotator.com, www.python.org, www.midwest-itc.org, www.cafepress.com, tar.weatherson.org, secure.wikimedia.org, theviennawilsons.net, www.gamebase.com.tw, www.newgrounds.com, angrychineseblogger.blog-city.com, www.open.com.hk, bbs.sina.com, www.mitbbs.com, www.parantezbaz.com, www.aixin119.com, english.rti.org.tw, www.ahrchk.net, mashable.com, www.siqo.com, www.websitepulse.com

A.2 Unfiltered Reference Domains

The following domains represent popular Chinese services that were anticipated not to be blocked within China, and were thus used as a test of the results returned by queried DNS servers:

baidu.com, qq.com, caixin.com, renren.com, chinaview.cn

Appendix B Filtering Techniques

Murdoch & Anderson (2008) categorise internet filtering approaches into four major families, largely according to the means by which traffic to be filtered is *recognized*, rather than the specific mechanism of blocking. These four categories, along with a fifth hybrid category, are discussed here:

TCP/IP Header Filtering

IP, the Internet Protocol, is the fundamental standard that largely determines the format and mechanisms by which computers are identified and located on the internet, and by which traffic passes from network to network. Fundamental to IP is the encoding of data into a series of discrete IP *packets*, which contain information such as the numerical address of the sending computer and the recipient, and the content of the data itself. Filtering may occur through inspection of the *header* of an IP packet, which details the packet's destination.

In terms of filtering, therefore, packets may be filtered according to lists of banned destination IP addresses. This method is simple and effective, but difficult to maintain due to the potential for services to change, or to have multiple, IP addresses. This approach may also incur significant collateral damage in the case of services that share IP addresses, causing multiple innocent services to be blocked along with the desired target.

TCP/IP Content Filtering

Rather than inspecting the header of packets, a filter may search the *content*, or *body*, of packets for banned terms. This presents a far more flexible approach to filtering, allowing packets to be blocked only if they include banned keywords or the traffic signatures of particular applications. This approach is also known as *deep packet inspection*, and is known to be employed to some extent by the Chinese national firewall. Deep packet inspection can be partially defeated by using encrypted connections, which obfuscate the data stored in the body of the packets, however filters may choose simply to block all encrypted connections in response, or to block traffic according to identifying traffic signatures that can occur even in encrypted protocols. The most significant limitation of this approach is that inspection of traffic content comes at a significant computational cost.

DNS Poisoning

The Domain Name Service, or DNS, maps human-readable names, such as `bbc.co.uk`, to numerical IP addresses usable by computers for routing data, such as `212.58.241.131`. The DNS is thus critical for most user-focused services such as the web. By altering DNS responses, returning either empty or false results, a filter can simply and cheaply block or redirect requests. This mechanism is simple to employ and maintain, but limits filters to entire websites, and can be relatively easy to bypass for technical users. DNS manipulation, or *poisoning*, is often employed as a first approach to web-based filtering, due to its low resource requirements and in spite of its ease of bypass; it has been noted that states typically begin with DNS filtering before graduating to more sophisticated filtering techniques over time (Deibert et al., 2008).

HTTP Proxy Filtering

A more sophisticated approach to filtering, to reduce resource requirements whilst maintaining flexibility, is to pass all internet traffic through an intermediary *proxy server* that fetches and, typically, caches information for users. This is a common internet service that can be used to speed up internet connections and reduce traffic. A suitably enabled proxy can, however, employ sophisticated filtering on certain destinations, whilst leaving other connections alone. This approach can, by ignoring the majority of traffic, be efficient on a national scale while still allowing for detailed filtering similar to TCP/IP content filtering.

Other Approaches

Various other means can be taken to regulate content on the internet. States can request that websites are removed from the internet, either by taking down their servers or by removing their names from the global DNS records. A state may also choose not to block a connection entirely, but to slow any connection to that site to unusable levels. It is also common for some nations to force the takedown of posts on internet forums and social

media websites in order to control debate; this is extremely common in the case of China and the Sina Weibo microblogging service (King, 2012). At a less technical level, legal and social constraints can be imposed to make access to certain services illegal or socially unacceptable.

References

- Aase, N., Crandall, J., Diaz, A., Knockel, J., Molinero, J., Saia, J., Wallach, D., & Zhu, T. 'Whiskey, Weed, and Wukan on the World Wide Web: On Measuring Censors' Resources and Motivations.' In *Proceedings of the 2nd USENIX Workshop on Free and Open Communications on the Internet. (FOCI 2012)*. Bellevue, Washington. August 2012.
- Alexa (2012). 'Alexa Ranking: Top 1,000,000 Sites'. <http://s3.amazonaws.com/alexastatic/top-1m.csv.zip>. Accessed January 15th, 2013.
- Al-Saqaf, W. (2008). 'Alkasir for Mapping and Circumventing Cyber-Censorship'. <http://www.alkasir.com/>. Accessed 8 August 2012.
- Anonymous. (2012). 'The collateral damage of internet censorship by DNS injection'. *SIGCOMM Comput. Commun. Rev.* 42, 3 (June 2012), 21-27. DOI=10.1145/2317307.2317311 <http://doi.acm.org/10.1145/2317307.2317311>
- Clayton, R., Murdoch, S. J. Watson, R. N. M. (2006). 'Ignoring the Great Firewall of China' in *6th Workshop on Privacy Enhancing Technologies*, Springer.
- CNNIC (2012). 'China Internet Network Information Center Homepage'. <http://www.cnnic.cn>. Accessed August 31st, 2012.
- Jedidiah R. Crandall, Daniel Zinn, Michael Byrd, Earl Barr, and Rich East. 'ConceptDoppler: A Weather Tracker for Internet Censorship'. In the *Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS 2007)*. Alexandria, Virginia. October 2007.
- Deibert, R. J., Palfrey, J. G., Rohozinski, R. Zittrain, J. (2008). *Access Denied: The Practice and Policy of Global Internet Filtering (Information Revolution and Global Politics)*, MIT Press.
- Dingledine, R., Mathewson, N. Syverson, P. (2004). 'Tor: The Second-Generation Onion Router', in *Proceedings of the 13th USENIX Security Symposium*.
- Ensafi, R., Park, J. C., Kapur, D. Crandall, J. R. (2010). 'Idle port scanning and non-interference analysis of network protocol stacks using model checking', in *Proceedings of the 19th USENIX Security Symposium*, USENIX Association, pp. 257–272.
- Arturo Filastò & Jacob Appelbaum. (2012). 'OONI: Open Observatory of Network Interference' In *Proceedings of the 2nd USENIX Workshop on Free and Open Communications on the Internet. (FOCI 2012)*. Bellevue, Washington. August 2012.
- King, Gary, Pan, Jennifer, and Roberts, Molly. (2012) 'How Censorship in China Allows Government Criticism but Silences Collective Expression.' *American Political Science Review*. Copy at http://www.apsanet.org/~/media/Files/Articles/2012/01/01_King_Pan_Roberts.pdf
- Lessig, L. 2006, *Code: And Other Laws of Cyberspace, Version 2.0*, Basic Books.
- MacKinnon, R. (2012). *Consent of the networked: the world-wide struggle for Internet freedom*. New York, Basic Books.
- MaxMind Inc. (2012). 'MaxMind GeoIP City Database', <http://www.maxmind.com/app/city>. Accessed August 8th, 2012.
- Morozov, E. (2011). *The net delusion: the dark side of Internet freedom*. New York, NY, PublicAffairs.
- Murdoch, S. Anderson, R. (2008). 'Tools and Technology of Internet Filtering', in R. Deibert, ed., *Access Denied: The Practice and Policy of Global Internet Filtering (Information Revolution and Global Politics Series)*, 2 edn, MIT Press, chapter 3, pp. 57–72.
- Pariser, E. (2011). *The filter bubble: what the Internet is hiding from you*. New York, Penguin Press.
- Psiphon Inc. (2012) 'The Psiphon Project', <http://www.psiphon.ca/>. Accessed August 8th, 2012.
- The Electronic Frontier Foundation. (2012) 'Tor Project Legal FAQ', <https://torproject.org/eff/tor-legal-faq.html.en>. Accessed August 8th, 2012.
- The Herdict Project. (2012). 'The Herdict Project', <http://www.herdict.org/>. Accessed August 8th, 2012.
- Wang, Y., Burgener, D., Flores, M., Kuzmanovic, A. Huang, C. (2011). 'Towards Street-Level Client Independent IP Geolocation', in *Proceedings of the 8th USENIX Symposium on Networked Systems Design and Implementation*, USENIX Association.

Yang, G. (2009). *The power of the Internet in China: citizen activism online*. New York, Columbia University Press.

Xueyang Xu, Zhuoqing Morley Mao, J. Alex Halderman. (2011) 'Internet Censorship in China: Where Does the Filtering Occur?' In *Proceedings of the 12th Passive and Active Measurement Conference (PAM 2011)*. Pages 133-142. Springer.