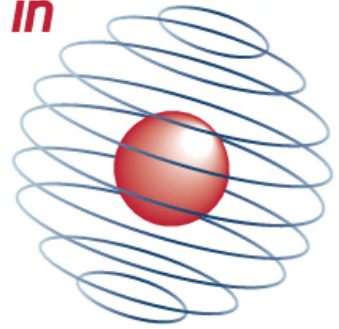




UNIVERSITY OF
OXFORD

CENTRE *for* DOCTORAL TRAINING *in*
**CYBER
SECURITY**



CDT Technical Paper

01/14

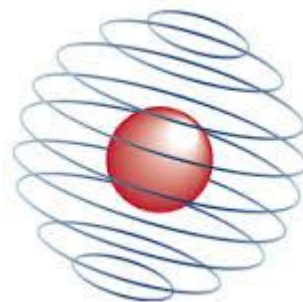
**A first look at Deep Packet Inspection
employed by the Golden Shield**

Oliver Farnan

**A first look at Deep Packet Inspection employed by
the Golden Shield**

6/27/2014

Oliver Farnan and Joss Wright



Abstract

We describe a series of tests on the capabilities of the Chinese Golden Shield. These tests focus on the Deep Packet Inspection capabilities of the Golden Shield, aiming to find out what is filtered and what is not. Our tests find that DPI triggering is not as easy to trigger as was expected, in contradiction of earlier research. We believe this is due to filtering optimisation of the Golden Shield in an effort to improve its efficiency given a limited technical capability. Our work joins a growing narrative that the Golden Shield is not able to fully monitor all network traffic in China, and makes sacrifices to focus primarily on key methods of information exchange, such as web traffic.

1 Introduction

Censorship of the Internet is increasing ^{[10][3]}. As implementation and enforcement of this censorship normally occur out of the public eye, it can be difficult for researchers to understand and analyse the censorship systems in place. This paper aims to give further insight into Internet censorship by examining the capabilities of the Golden Shield of China.

Deep Packet Inspection (DPI) is one of the most powerful yet least researched methods of Internet censorship. DPI offers censors the capability to perform more granular and thorough analyses of content than other methods. Whereas other forms of censorship may block an entire IP address, website or domain name, DPI allows censors to block a single file, such as a webpage or image. Additionally DPI offers more comprehensive censorship, allowing censorship of a broader set of technologies than other filtering methods.

The People's Republic of China's Internet filtering has long been considered one of the largest and most sophisticated in the world ^[2], and has been known to use DPI techniques ^[22]. This censorship network is known officially as the Golden Shield, and colloquially as the Great Firewall of China. It covers a range material the Chinese Government deems inappropriate, such as historical events, political ideologies and calls for collective action ^[13].

Most previous research on Chinese censorship has focused on higher level Internet protocols (such as DNS ^{[27][26]}, HTTP ^[4] and Tor ^[25]) and lower level filtering has not been examined to the same extent. Specifically, keyword filtering ^[21], HTTP GET request filtering ^{[4][16]} and DNS lookup filtering ^[27] appear to be the primary methods of content filtering employed by the Golden Shield.

In this paper we outline methods for evaluating DPI used within a censorship network such as the Golden Shield. These are potentially capable of blocking a wider range of content than other techniques, including data within a variety of network protocols from ICMP and UDP to FTP and IRC. Although work has covered these protocols in the past, they are frequently not the primary focus.

One of the drawbacks of DPI in contrast to other filtering techniques is its speed ^[29]. As DPI speed increases it is likely that this technology will become more prevalent. We believe this is an area that requires further research, and as such this paper lays out several methods and techniques that can be used for detecting, mapping and analysing DPI based filtering.

As DPI filtering is resource intensive, there is often pre-filtering of data so that only certain types are fully analysed. Indeed with the Golden Shield, there is evidence of filtering techniques adapted to

make better use of available filtering resources^[16]. One of the key areas of interest for DPI filtering is determining what is filtered and what is not, and how this corresponds to the requirements and desires of the censoring agency.

In this paper we're going to cover related work in section 2, discuss considerations for the research in section 3, and describe our tests and their outcomes in section 4. We provide our conclusions in section 5 and future work in section 6.

2 Related Work

There is a variety of previous work covering both censorship systems in general, as well as the specific techniques used within China.

In 2006 Clayton, Murdoch and Watson^[4] discovered how the Golden Shield filters connections. It does this by sending TCP RST packets to both hosts (the client and the webserver) when specific keywords are detected, fooling each host into believing that the other has closed the connection. They show how this could be used to perform a denial of service attack.

Park and Crandall look at HTML response filtering by the Golden Shield in 2010^[16]. They look at how the Golden Shield typically filters both GET requests and HTML responses (as covered by Clayton et al^[4]), and how each of these would be affected by being geographically distant from both end hosts. They conclude that because HTTP request packets detect banned keywords within an already connected data stream, it is more difficult for RST attempts (from the filter) to 'guess' the correct sequence number and so this often fails. In contrast, GET requests typically happen at the start of such a connection, so it is easier for the RST packets to arrive successfully at either host. They theorise that this is why HTML response filtering appears to have been discontinued, while GET request filtering is still in use. Crandall has published previous work proposing methods to monitor Internet Censorship^[26].

In 2011 Xueyang, Mao and Halderman have estimated a map of where China's Internet filtering is taking place^[28]. They do this by gathering a list of Chinese ASs, tracerouting through these and then merging the topology. With this map they then attempt to locate filtering devices and study their location on the network. They send HTTP GET requests containing filtered keywords with increasing TTLs to web servers within China, and wait for their connection to be blocked. This allows them to estimate where in the network the filtering is taking place. They find that most of the filtering occurs within the border ASs connecting China to the outside Internet.

Wright has done previous work studying regional variation in the Golden Shield^[27]. The focus of this work was on DNS filtering, where an incorrect response was returned to a filtered DNS request. He showed that although geographically dispersed DNS servers within China do not all use identical lists for filtering, there are clearly links between them and what they filter.

King, Pan, and Roberts look at what the authorities wish to censor within China^[13]. They find that contrary to previous thought, posts criticising the state, its leaders and its policies are not unconditionally censored. They find that instead censoring is most likely to occur when the content

supports or incites collective action or social mobilisation. This censorship occurs regardless of the context and reason for the action.

This body of research shows how the Golden Shield is changing over time. Repeating Clayton, Murdoch and Watson's ^[4] technique does not trigger filtering today. As well, Park and Crandall's ^[16] work shows change over time. In both cases this is a change for the Golden Shield filtering less data, instead of more. In [4] it is suggested that this is due to the lack of success of previous methods, combined with a limitation of the amount of data throughput capability that the Golden Shield is capable of monitoring.

3 Considerations

Some additional areas had to be considered before analysis of the capabilities of the Golden Shield could be performed.

3.1 String Matching

DPI filtering is typically based on string matching. This is where a string or regular expression is specified in the filter, and is then compared to the payload of the packet it is being applied to.

Fast string matching is dependent on quickly eliminating packets which do not meet filtering criteria. This can be done using techniques such as bloom filtering ^[6] or by specifying the characteristics of the packet (e.g. header data, incoming interface) prior to applying the string matching to confirm whether the packet matches the rule ^[19]. When there is no match at this stage, the packet can be ignored and forwarded to the next network node (assuming a blacklist approach is being used). If the criteria are met, then the string matching expression is applied to determine whether the packet should be filtered. If the packet fails both the pre-selection and the string matching, it is flagged and some filtering action can be taken. This could be simply dropping the packet, or (as is the case with the Golden Shield) taking other action to close the connection between hosts.

3.2 Filtering Techniques

Once a packet or stream has been flagged, there are different techniques that can be taken to filter the content that is to be censored. The most common method of filtering observed by the Golden Shield is to send TCP RST packets to both ends of the connection (e.g. to both the client and the webserver) ^{[8][28]}. This causes both hosts to end the TCP connection, preventing the server sending more data, and preventing the client application from displaying the data.

An advantage of this technique is that it can be performed out of band of the network traffic itself. The data can be copied to another node for analysis, which can perform the string matching, leaving the primary networking nodes free to continue to process data streams as fast as possible. If the matching criteria are met, the RST packets can be sent from the analysing node. This setup is described by Clayton et al ^[4].

3.3 Test String - Falun

For the tests performed in this analysis, the test string 'falun' was used to test the response to censorship. 'falun' is a highly censored word within China ^[11], and its use for testing filtering of the Golden Shield has precedence ^{[28][16]}. Censorship of 'falun' is due to the highly persecuted Falun Gong

movement. Falun Gong is a Chinese spiritual discipline that has come under heavy scrutiny from the Chinese government since its founding in the mid-1990s.

3.4 Stateful Filtering

Network filtering often has the capability to take the state of the network connection or packet into account. An example of this is whether the connection is going-into or coming-out-of the filtered domain (in this case China).

There are many different variables stateful filtering can take into consideration. While stateful filtering often covers simply the source and destination of a packet, it can also cover situations such as whether a full connection has already been established, or whether the packet is a request or a response.

3.5 Location of Tests

All of the tests described in this paper were carried out from hosts outside of China. This made it difficult to comprehensively test DPI filtering due to the probable stateful nature (at least partial) of the Golden Shield ^[16].

Effort was made to perform tests that could either circumvent the stateful nature of the filtering, or be agnostic to its stateful filters. Where possible, tests were performed trying both hosts (those within and outside of China) as both the source and destination for establishing connections. Additionally, protocols were used which made it more difficult for stateful filtering to determine where the request originated.

4 Our Approaches

Several methods for analysing DPI filtering were created and tested. They aim to build upon previous work by improving existing methodologies, as well as exploiting previous methodologies to focus specifically on DPI filtering. In section 4.1 we describe an improvement over previous methodologies for mapping out the Chinese Internet and AS topology, in section 4.2 we describe our method for testing DPI using FTP, and in section 4.3 we describe tests over HTTP GET requests to test the stateful nature of the Golden Shield.

4.1 Improved Network Mapping

To carry out mapping of censorship similar to that as performed by Xueyang, Mao and Halderman ^[28], a more thorough network map of China was needed. When we repeated their process, it was discovered that there was a more thorough way of mapping Internet Infrastructure than the previous methodology.

The first change to our methodology was to increase the numbers of address ranges scanned within an AS. Although address ranges within an AS will often be directly connected, this is not always the case. The previous methodology only used a single address to represent an entire AS, but by doing this we felt we were missing many potential paths. Instead we tracerouted to all of the address ranges belonging to the AS.

As well as this, our method found more ASs than the previous method, although it is likely that some of this is attributable to an increase in the number of ASs used within the country since 2010. The

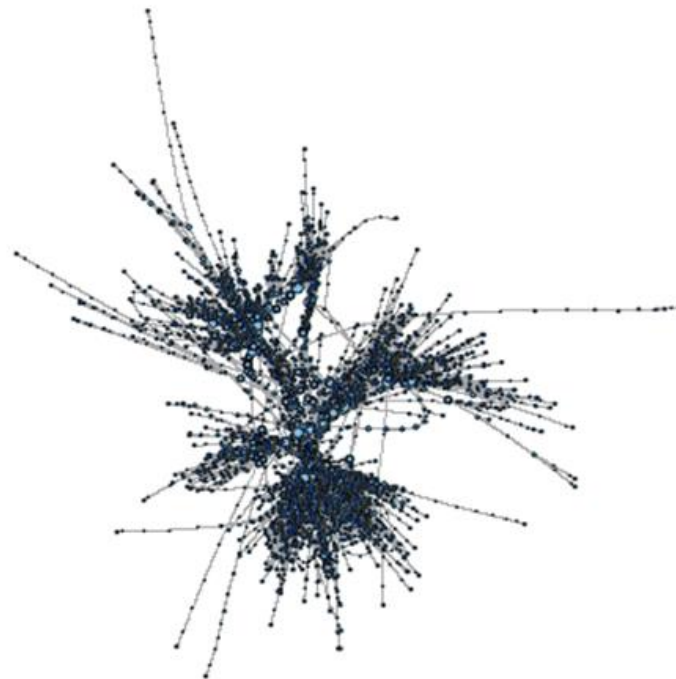
previous methodology attempted to find ASs by referencing the list of Chinese ASN against records within Routeview ^[24] and Ripe ^[17] databases. This returned address prefixes for 76 Chinese ASs. By repeating this methodology we discovered that many assigned ASNs were missing. Instead we crawled Robtex ^[18] and were able to discover IP address ranges for 121 ASs.

As a result of these changes, our mapping connected the routes to 27526 address ranges, compared to the 76 in the paper it was based on. This yielded a substantially different infrastructure map of China than if the previous methodology had been used.

Unfortunately within the timeframe of this project it was not possible to repeat these tests from hosts at other geographical locations, and instead the mapping was performed only from Oxford. This inevitably resulted in a tree branching out of Oxford towards the target ranges. As a result, while it demonstrates a large part of the Chinese Internet infrastructure, it is missing many joining connections between different AS and address ranges.

The data we obtained was loaded into igraph ^[12], a network analysis tool for R. This covers 78249 total hops, and 2241 unique nodes. Using a complete dataset (including mapping from hosts outside of Oxford) this will allow us to perform analyses on the Chinese Internet infrastructure. Diagram 1 is a map of the infrastructure routes in our data, produced by igraph. The larger the node on the diagram, the more routes passed through it.

Diagram 1 – Map of Chinese Internet (as mapped from Oxford)



The 27526 unique address ranges we discovered covered 374726879 unique IP addresses within to Chinese ASs. A breakdown of the address range of each of these can be found in Table 1.

Table 1 – Breakdown of IP Ranges in China

Address Range	Number of Occurrences Belonging to Chinese AS	Number of unique IPs per individual range	Total number of IPs within each range size
/32	7	1	7
/31	0	2	0
/30	20	4	80
/29	11	8	88
/28	15	16	240
/27	10	32	320
/26	16	64	1024
/25	8	128	1024
/24	10110	256	2588160
/23	3524	512	1804288
/22	2608	1024	2670592
/21	2046	2048	4190208
/20	2276	4096	9322496
/19	1347	8192	11034624
/18	1002	16384	16416768
/17	584	32768	19136512
/16	987	65536	64684032
/15	293	131072	38404096
/14	160	262144	41943040
/13	112	524288	58720256
/12	45	1048576	47185920
/11	13	2097152	27262976
/10	7	4194304	29360128
/9	0	8388608	0
/8	0	16777216	0
/7	0	33554432	0
/6	0	67108864	0
/5	0	1.34E+08	0
/4	0	2.68E+08	0
/3	0	5.37E+08	0
/2	0	1.07E+09	0
/1	0	2.15E+09	0
		Total unique IPs within Chinese ASs	374726879

4.2 FTP

To test DPI within China a variety of tests were performed using FTP. FTP was chosen as it provides a variety of ways of transferring data between two hosts, including options for the remote host to establish a reverse connection with the local host. This is important as it allows testing of stateful filtering.

This experiment was carried out on servers within the Chinese address space and within Chinese ASs. To find these addresses, the data from the mapping of the Chinese network was consulted (section 4.1). This method provided a list of 27526 address ranges, covering 374726879 different unique IP addresses.

These addresses were then scanned using Zmap. Zmap is a port scanning tool developed by Zakir Durumeric, Eric Wustrow, and J. Alex Halderman ^[7], optimised for performing a survey scan over a wide variety of hosts. Unlike earlier port scanners such as Nmap ^[15], Zmap is asynchronous and stateless. By gaining this speed it sacrifices some reliability (~2% false negatives using a single SYN packet). There are alternative fast asynchronous or stateless port scanners, such as MassScan ^[14], Unicornscan ^[23] and scanrand ^[20] which are arguably even faster ^[14].

Zmap sent a single SYN on TCP/21 to all 374726879 hosts, and gave an output for all of those which responded with a SYN-ACK. This resulted in 164135 hosts that responded in this way, indicating that they had a service running on this port. As this is the default port for FTP control, this is indicative but not proof of servers running this service.

This list was then checked to see which of these hosts were running FTP. We were specifically interested in hosts which gave access to FTP functionality with anonymous credentials. Anonymous FTP typically accepts null credentials, or either “ftp” or “anonymous” as the username. These credentials will give READ, WRITE or READ/WRITE access to the service.

Metasploit was used to test if the hosts responding on TCP/21 supported anonymous FTP. Metasploit tests this by attempting to log on to the service and requesting that a test file be uploaded, downloaded, and then deleted. Based on these responses (and the banner the service gives as you log in) we were able to identify 5401 (of 164135 responding on TCP/21) FTP servers providing anonymous functionality within the Chinese address ranges. Of these 539 also allowed anonymous WRITE actions. It should be noted that anonymous FTP is typically a default off feature, meaning hosts providing this service have gone out of their way to allow it.

Several tests were conducted on these hosts. Each of these tests covered different combinations of data to see if any filtering was in place. Hosts belonging to corporation organisations were selected for these tests, over those hosts belonging to individuals, or those which could not be identified.

The first test involved an innocuous file containing the string ‘test1’, which was uploaded, downloaded and deleted from the FTP server. This was our neutral test used to confirm no filtering of data containing innocent content.

The second test involved a file containing the string ‘falun Falun’, which was uploaded, downloaded and deleted from the FTP server. This test was performed with both active and passive FTP modes. Passive FTP mode transfers the file from the remote host to the local host using the TCP data stream

that has already been connected between the two hosts. Active FTP mode creates a new TCP data stream from the remote host to the local host. This data stream was used to test whether stateful filtering was in place, as it allowed testing of connections both going into and coming out of China.

The third test involved a file containing a fake HTTP GET request. HTTP GET filtering has been observed in China in the past ^[16]. As DPI filtering is based on string matching, depending on the exact regular expression used this could trigger any DPI used for filtering genuine HTTP GET requests. As active FTP mode was used this request appeared to come from a host within China, attempting to access blacklisted material outside of the Chinese network.

The file contained:

```
GET /falun.Falun HTTP/1.1  
Hostname:*****
```

No filtering or censorship techniques were observed on any of these tests. This indicates that if DPI filtering is taking place, it is either not taking place on the ports FTP is using (TCP/20, TCP/21 and the ephemeral ports), or that other criteria of the string matching are not being met.

4.3 HTTP GET Requests

Previous work has shown the use of HTTP GET request filtering ^{[16][28]}. To test the current functionality and statefulness of this filtering, we performed a series of HTTP GET request tests. These tests were performed on websites within China, from connections outside of China. The 100 sites chosen for this test were the Alexa Top 100 sites in China ^[1]. This test was based on the HTTP GET request tests performed by Jong Chun Park and Jedidiah R Crandall ^[16] and those by Clayton et al ^[4].

Several HTTP GET requests were made to each of the chosen sites. These requests attempted to request a series of websites from each server. Five requests were made to each websites, requesting the pages / (the top level page), /index.html, /hellohello.html, /falun.html, and /falunFalun.html. These requests were made using GNU Wget ^[9] and HTTP version 1.0. Altogether 500 HTTP GET requests were made to web servers within China from external hosts.

These requests were made so that an example of a valid request could be observed (typically / and /index.html), an example of a non-censored but failed attempt could be made (hellohello.html) which should result in a 404, redirect or related error, and an example of two potentially censored requests (/falun.html a /falunFalun.html). The failed requests were then compared to the censored requests to determine if there was any difference in the HTML response.

Both /falun.html and /falunFalun.html were used in case the word Falun was split between packets. Previous research has indicated that censorship systems (the Golden Shield in particular) have failed to prevent content when the filtered string is split over multiple packet payloads ^[16].

Censorship of these requests was likely to occur in one of two places: at the HTTP response (e.g. a fake 404, or an HTML page missing content) or at the TCP connection (e.g. with the connection set via a RST packet). Firstly, we analysed the HTML response content from the web servers. We looked for any differences between the pages of the invalid requests (/hellohello.html) and those using potentially

filtered requests (/falun.html, /falunFalun.html). No differences in the HTTP response were noted between the invalid group of innocent requests, and the invalid group of potentially filtered requests.

Secondly, during this period network traffic on the end hosts was recorded and evaluated. There were no signs of lower level filtering attempts (such as unexpected TCP RST packets) from either the servers themselves, or nodes between our local host and the webserver. All requests for non-existent content returned the same network traffic as those for potentially filtered requests. No unexpected RST requests were observed.

5 Conclusions

The DPI tests we ran connecting into China from outside connections came up negative. There was no censorship indicating DPI techniques that were present for the tests carried out for this research.

The big restriction of these tests was that they were performed from hosts outside of the Golden Shield. While efforts were made to perform meaningful tests given what was available, this placed limitations on what we were able to carry out.

The surprising result of these tests was that it was not as easy to trigger Golden Shield filtering as was expected. Carrying out tests which should have triggered filtering according to previous work ^[4] came up negative this time.

A possible explanation for this may be that there is an on-going effort to optimise what the Golden Shield is filtering, and that it does not have the capability to monitor all traffic within its domain. This fits with the narrative laid out in previous work ^[16]. DPI censorship capabilities are dependent on the speed and memory of the filtering tools. With this in mind, it makes sense for the operators of the Golden Shield to make efforts to limit the amount of work it has to do by ignoring certain data types, protocols and ports.

This may have manifested itself twice within our tests. Firstly, performing HTTP GET requests for known filtered content should have been filtered according to earlier research ^[4]. That these tests are now not triggering filtering could indicate that the Golden Shield has been changed to ignore requests from outside of China.

Secondly, the FTP active mode tests performed data transfers initiated from hosts allowing anonymous WRITE from within China. These tests sent data containing known filtered keywords from within China to hosts without it, over TCP ports 20, 21 as well as ephemeral ports, and there appeared to be no filtering of this content.

This indicates that either certain ports have been 'whitelisted' from filtering (e.g. those known for the transfer of large amounts of data), or certain ports have been 'blacklisted' for checking (those used for the spread of information). Given that we know that Chinese censorship is aimed at counteracting the flow of certain information ^[13], it makes sense for them to focus on those protocols (and ports) which are typically used for this. The main suspect of this would of course be TCP/80 and HTTP. Further experiments will be needed to identify whether they've gone for a 'whitelist' or 'blacklist' approach, and how comprehensive their coverage of other protocols is.

Overall, our work fits the narrative that there is a continual effort to optimise the efficiency of the Golden Shield. New research in this field frequently finds inconsistencies with older work, and it is likely due to this optimisation. Our work builds on that growing list of inconsistencies, and offers new techniques and methods for testing the Golden Shield and its capabilities.

6 Next Steps

There is a lot that can be done to clarify some of the findings in this paper. The most telling finding would be to repeat the tests using source hosts from within the domain of the Golden Shield. This could immediately confirm that filtering (specifically of web content) is now stateful, and ignores some source addresses from outside of its coverage.

To cover the possibility that filtering is done based on different ports and services, future tests should aim to look at filtering of data over different source and destination ports. Examples of known filtering should be compared to the same message and payload, but on a different port. One such test that could be carried out would be to repeat known filtered HTTP GET requests, but connecting over a different port instead of TCP/80. This will reveal whether string matching is looking simply for GET request attempts over all traffic, or whether (in an effort to optimise their filtering) it has been restricted to TCP/80. The test can then be repeated over a range of ports to determine which ports are filtered and which are not.

Similarly, filtering of other transport layer protocols should be explored. Where GET requests are filtered over TCP/80, would similar packets be filtered if the same data was sent over equivalent UDP ports? What if the data is stored within the payload of ICMP packets?

The tests described in this paper were all performed by scripts, and these scripts have been written to function over a variety of *nix based hosts. It is our hope that we will find an opportunity to repeat them from different geographical locations, including some within China itself.

References

- [1] Alexa.
<http://www.alexa.com/>
- [2] Bambauer, Derek E., et al. "Internet filtering in China in 2004-2005: A country study." Berkman Center for Internet & Society at Harvard Law School Research Publication 2005-10 (2005).
- [3] Bitso, Constance, Ina Fourie, and Theo JD Bothma. "Trends in transition from classical censorship to Internet censorship: selected country overviews." *Innovation: journal of appropriate librarianship and information work in Southern Africa: Information Ethics* 46 (2013): 166-191.
- [4] Clayton, Richard, Steven J. Murdoch, and Robert NM Watson. "Ignoring the great firewall of china." *Privacy Enhancing Technologies*. Springer Berlin Heidelberg, 2006.
- [5] Crandall, Jedidiah R., et al. "ConceptDoppler: a weather tracker for internet censorship." *ACM Conference on Computer and Communications Security*. 2007.
- [6] Dharmapurikar, Sarang, et al. "Deep packet inspection using parallel bloom filters." *High Performance Interconnects, 2003. Proceedings. 11th Symposium on*. IEEE, 2003.
- [7] Durumeric, Zakir, Eric Wustrow, and J. Alex Halderman. "ZMap: Fast Internet-wide Scanning and Its Security Applications." *USENIX Security*. 2013.
- [8] Ensafi, Roya, et al. "Detecting Intentional Packet Drops on the Internet via TCP/IP Side Channels." *Passive and Active Measurement*. Springer International Publishing, 2014.
- [9] GNU Wget
<https://www.gnu.org/software/wget/>
- [10] Gomez, James. "Dumbing down democracy: Trends in internet regulation, surveillance and control in Asia." (2004).
- [11] House, Freedom. "Freedom on the Net 2012." (2012).
- [12] iGraph
igraph.org
- [13] King, Gary, Jennifer Pan, and Margaret E. Roberts. "How censorship in China allows government criticism but silences collective expression." *American Political Science Review* 107.02 (2013): 326-343.
- [14] MassScan.
<https://github.com/robertdavidgraham/masscan>.
- [15] Nmap.
<http://nmap.org/>
- [16] Park, Jong Chun, and Jedidiah R. Crandall. "Empirical study of a national-scale distributed intrusion detection system: Backbone-level filtering of html responses in china." *Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on*. IEEE, 2010.

[17] RIPE NCC Projects

<http://www.ripe.net/projects/ris/rawdata.html>

[18] Robtex.

www.robtext.com

[19] Roesch, Martin. "Snort: Lightweight Intrusion Detection for Networks." LISA. Vol. 99. 1999.

[20] Scanrand.

<http://www.sans.org/security-resources/idfaq/scanrand.php>

[21] Tao, Mr. "China: Journey to the heart of Internet censorship." Investigative report sponsored by Reporters Without Borders and Chinese Human Rights Defenders (2007).

[22] Tor – Knock Knock Knockin' on Bridges' Doors. 2012.

<https://blog.torproject.org/blog/knock-knock-knockin-bridges-doors>

[23] Unicornscan.

<http://www.unicornscan.org/>

[24] University of Oregon Route Views Archive Project

<http://archive.routeviews.org>

[25] Winter, Philipp, and Stefan Lindskog. "How the Great Firewall of China is Blocking Tor." FOCl. USENIX Association (2012).

[26] Wolfgarten, Sebastian. "Investigating large-scale Internet content filtering." M. Sc. in Security and Forensic Computing 2006 (2005).

[27] Wright, Joss. "Regional variation in Chinese internet filtering." Information, Communication & Society 17.1 (2014): 121-141.

[28] Xu, Xueyang, Z. Morley Mao, and J. Alex Halderman. "Internet censorship in China: Where does the filtering occur?." Passive and Active Measurement. Springer Berlin Heidelberg, 2011.

[29] Yu, Fang, et al. "Fast and memory-efficient regular expression matching for deep packet inspection." Proceedings of the 2006 ACM/IEEE symposium on Architecture for networking and communications systems. ACM, 2006.