Invisible Barriers: Identifying restrictions affecting New Zealanders' access to the Internet

by

Shadi Esnaashari

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the
requirements for the degree of
Master of Science
in Computer Science.

Victoria University of Wellington
2014

Abstract

The Internet is an important technology worldwide. People use the Internet for research, communication, shopping, entertainment, etc. In addition to these benefits, the Internet provides access to dangerous or illegal material. Because of this, some content and services may be blocked by governments, Internet Service Providers, organizations, or individuals. This blocking, whether for security or for network efficiency, has significant effects on people's access to services and information, which may not be considered when implementing restrictions. Although studies have been conducted on Internet blocking in many countries, no one has yet examined what is being blocked in New Zealand. In this thesis, we measured the prevalence of Internet blocking in New Zealand and the reasons leading to a decision to block access to websites or Internet services. Although several different tools existed, they could not be used directly because they either concentrated on a narrow range of services or did not work in an environment where some services they depended upon were blocked. For this reason, we developed our own tool called WCMT based on the issues identified from previous tools. We conducted our study using WCMT in order to identify blocked websites and services in our quantitative analysis, complemented by interviews with key informants in our qualitative analysis.

Produced Publications and Presentations

1. Shadi Esnaashari, Dr Ian Welch, Dr Brenda Chawner , "Invisible Watchers", (full paper-New Zealand Computer Science Research Student Conference NZCSRC. Hamilton, April 2013).

2. Shadi Esnaashari, Ian Welch, and Brenda Chawner," WCMT: Web Censorship Monitoring Tool"', Australasian Telecommunication Networks and Applications Conference (ATNAC 2013). Christchurch, New Zealand, Nov 20-22, 2013.

3. Shadi Esnaashari, Ian Welch, and Brenda Chawner, "Invisible Barriers: Identifying restrictions affecting New Zealanders' access to the Internet", Australasian Women in Computing (AWIC), Auckland, New Zealand, Jan 20-23, 2014.

4. Shadi Esnaashari, Ian Welch, and Brenda Chawner, "Restrictions affecting New Zealanders' access to the Internet: ALocal Study", The 28th IEEE International Conference on Advanced Information Networking and Application (AINA 2014), Victoria, Canada, May 2014.

Acknowledgments

In the completion of my Master's thesis I have been honored to work with brilliant professionals. I would like to thank my first supervisor Dr. Ian Welch. His great encouragements always motivate me, so does his magnificent support as my supervisor during my Honors project. I also extend my highest appreciation to my second supervisor, Dr. Brenda Chawner. I strongly appreciate her help in the previous year while I was writing up my thesis.

I was also lucky to have my husband beside me during this time. He was my support all through my studies. He supported me emotionally and encouraged me to successfully finish my studies. I would also like to offer my special thanks to my parents for always supporting me and making this course of study possible in the first place.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

"I believe there is something out there watching us. Unfortunately, it's the government."

(Woody Allen)

The Internet is a means of communication and a source of information. Although many people benefit from access to global communication, research, shopping, and entertainment, some use it for illegal activities. Because of this, content or specific services may be filtered by governments, organizations, Internet Service Providers, or individuals. Internet filtering, which can be considered a form of censorship, is a growing concern around the world and affects a large number of people [12].

The motivations for blocking access include politics and power, social norms and morals, security concerns, and protecting intellectual property rights and economic interests.

Blocking can be implemented by filtering the content or manipulating the traffic through different techniques such as packet filtering, deep packet inspection, URL blocking, service blocking, DNS tampering, BGP (Border Gateway Protocol) route manipulation, etc [9, 50]. Various types of blocking, whether for security or for network efficiency, may affect peo-

ple's access to services and information, which may not be considered when implementing restrictions. While many people support filtering, others see filtering as a denial of their rights. They believe that network neutrality should be respected [31].

There are cases where blocking goes beyond simple Internet filtering. Some Internet Service Providers (ISPs) or organizations may alter network traffic based on specific keywords, IP addresses, or URLs. They may prioritize some types of traffic or block others. ISPs adopt different blocking practices. Individual ISP settings may affect the performance of applications, such as Voice over IP, online games, or peer to peer sharing. This type of traffic management is not transparent to users, and users may not be aware that it is happening.

When these ISPs discriminate against some services for users they use a range of techniques, such as traffic shapers, blockers, and firewalls in order to monitor and manipulate their users' traffic. Finding such discrimination is difficult for users. Previous research has shown that some ISPs change or block users' traffic such as BitTorrent, VoIP, etc [71].

Many studies have been conducted to determine the level of censorship done at a governmental level in different countries [74, 70, 76, 4, 18]. However, we are not aware of studies focusing on the organizational level in New Zealand. In particular, we are interested in what restrictions are placed on wireless Internet access provided by organizations such as libraries, museums, tourist organizations, academic institutions, etc. When people use these services, it is often not clear what is allowed and disallowed in terms of access to web content and access to wider services. In addition, those organizations cannot always articulate their own policies when asked. For example, German visitors to Te Papa have found access to a moderate left-wing political site blocked because it was classified as Japanese pornography [54].

In this thesis, we describe a study that addresses the question of what is blocked and why by providers of wireless access in the Wellington region.

This required us to develop a system called WCMT (Web Censorship Monitoring Tool) that allowed detection of both blocked content and services. Although several different tools existed [53, 36, 38, 58, 25, 4, 1, 21, 6, 76] they could not be used directly because they either concentrated on a narrow range of services or did not work in an environment where some services they depend upon were blocked. To start our experiment we chose Open Observatory of Network Interference (OONI). OONI is a censorship monitoring tool. Unfortunately we could not use OONI since it used Tor and Tor was blocked in almost all of our experimental networks. There were other issues with OONI such as giving lots of false positives and not considering service blocking. These issues led us to implement WCMT.

We used WCMT to carry out a quantitative study of 8 providers with a follow up qualitative study where we investigated the organizational reasons underlying their blocking policies.

1.1 Research goals

The overall goal was to understand what and why content and services were blocked at free wireless access points provided by organizations in Wellington. We investigated, uncovered and analyzed Internet filtering in New Zealand in order to make them transparent for the users and ask for support in this area. We employed an approach that consisted of:
A. Developing a technical tool and core methodologies for the study of Internet filtering in New Zealand.
B. Finding content and service blockings in different organizations.
C. Interviewing IT professionals in organizations where we experimented our tool to identify reasons for implementing censorship.

1.2 Thesis structure

The rest of this thesis is organized as follows. Chapter 2 presents background of the problem and related Work. Chapter 3 looks at the design and implementation of our new tool. Chapter 4 looks at our preliminary testing. Chapter 5 presents quantitative measurement of blocking. Chapter 6 presents our qualitative investigation of polices. Chapter 7 discusses blocking and issues around it. Finally, Chapter 8 gives our conclusion, contributions, and suggestions for future work.

Chapter 2

Background and related Work

In this chapter, we start with a discussion on Internet filtering and provide a discussion on background and related work. We also review tools for investigating Internet filtering and use the results of this review to define the requirements for our new tool called Web Censorship Monitoring Tool (WCMT).

2.1 Content filtering

The terms Internet filtering [65] or blocking refers to the technical approaches that controls access to specific information on the Internet. This action happens without the cooperation of the content provider. The reasons for filtering are different. It can be for the safety of children, political, the industry's responsibility to keep their data private, etc.

2.2 Traffic manipulation

When the Internet was created all types of traffic were treated equally, a concept known as network neutrality [79]. However, as the amount of Internet traffic has increased, all data packets have not been treated equally,

and some ISPs give priority to certain types of traffic over others, a practice known as traffic manipulation. Network neutrality and traffic manipulation are controversial topics.

Supporters of network neutrality believe that controlling Internet access goes against the principles of the Internet. They argue that access to the Internet is now a de-facto human right, and allowing other people or organizations to determine which types of traffic receive priority is a form of discrimination.

Opponents of network neutrality include some ISPs, governments, and individuals who believe that network neutrality is harmful for both their businesses and the health of society. In particular, they argue that bandwidth intensive applications, such as video applications, use a disproportionate amount of capacity, and therefore, disadvantage simpler formats. To counter this effect, some ISPs prioritize specific traffic and deny other types of packets permission to pass [34].

Zdravko et al. [79] indicated that two issues of network neutrality are capacity and traffic management. Using different network applications has increased the use of Internet but the capacity of the Internet is limited. When Internet use increases, the network provider's cost will increase accordingly. Moreover, some of the network applications are very sensitive to delay, jitter, and packet loss. For the reasons discussed above, some ISPs prioritize specific traffic and do not permit some packets to pass. This is the cheapest way for ISPs' to prevent extra cost. Otherwise there is nobody to pay the ISPs the extra costs. Providers make decisions based on traffic type, source, and destination address and they use traffic management techniques to block or prioritize specific traffic in order to get better results.

The first network neutrality issue emerged in the USA in early 2000s [66]. Both the US Telecom and cable operators blocked specific traffic. US Telecom and cable operators blocked access to VoIP services.

Some organizations such as communication companies may apply packet

changes in the SKYPE or WhatsApp in order to prevent customers from using them. Telecommunication companies know that if customers use these applications, they will not use telecommunication centers' services. Hence, these companies will lose revenues. Communication companies may prioritize special packets. For example, they can give permission for the pay services to go faster than free services. This is another example which weakens network neutrality [51].

The problem affects users when the ISPs block certain types of packets and the novice user wants to use a new application. If the novice user cannot run the application, he/she will not know whether there is a problem with the application or the network. The novice user may think their application is malfunctioning which gives a sense of confusion to the users which is frustrating.

Different ISPs have different regulations and blocked different applications. Users who wish to access the blocked application might therefore be forced to change their ISP. But if all ISPs filter a specific application, then, there will be no other choice for the user. In addition, finding other ISPs will be time consuming and costly.

2.3 Implementing filtering

Implementing content filtering and traffic manipulation has occurred through different mechanisms such as packet filtering, DNS tampering, and BGP route manipulation [10]. A description on each of them has been given in the following sections [69].

2.3.1 Packet filtering

Packet filtering will happen at the routers or any other devices on the path of the communication. They will look at the header of the packet and will decide whether to allow or deny further passage to the packets [10]. Filter-

ing applies in either layer three or layer four of the Open System Interconnected (OSI) model [41]. The OSI model contains seven different layers, namely, physical, data link, network, transport, session, presentation, and application. The lowest layer is the physical layer. The highest layer is the application layer. The action of packet filtering will be done by firewalls.

Firewalls [17] examine all traffic routed between two networks in order to check whether they meet the criteria or not. If it does not, it is stopped. Firewalls can filter both inbound and outbound traffic. It is used to manage public access to private network as well as logging all attempts to enter that network.

Filtering the packets in firewalls will be performed based on their source and destination addresses and port numbers. Specific types of network traffic can also be filtered at firewalls. Firewalls are divided into three categories. An overview of each has been given below.

Packet filtering gateways: Packet filters provide a simple level of gateway security. This kind of gateways drops packets based on the source address, destination address, or port numbers. Packet filtering will be applied at layer three and four of the OSI model.

Layer three filtering: Layer three filtering happens at the IP header. This header contains information about the machine sending the packet. If the goal is to deny access to the special host, all the information coming in or out of that host will be denied.

The first kind of firewall refers to filtering the packet. The packet filter will happen at level 3 of OSI model (network layer) by looking at source and destination IP address. This kind of filtering can also happen on the protocol field in IP header.

There are lots of techniques to trick this kind of filtering such as spoofing, fragmenting and other ways of passing the traffic. But the

benefit of this technique is that it is faster compared to other techniques.

Layer four filtering: In layer four filtering, the data inside the IP packet in addition to the layer three filtering will be used. Port numbers are retrieved in this layer. So with this ability, the filtering can be applied not only on the IP address but also on the port number.

The problem with IP filtering is that if one IP address is blocked, every domain name which is hosted by that IP address will be blocked too.

Deep packet inspection firewalls: Another kind of filtering is through deep packet inspection (DPI). This method is stateful inspection plus visibility to application layer. This method allows the firewall to see the data passing through and not just the connection information itself. DPI can be used against buffer overflow attacks, denial of service attacks, intrusions, and worms.

Although DPI is very useful for Internet management, it has an effect on network neutrality and it reduces the openness of the Internet. Most of ISPs use DPI for advertisement, quality of services, copyright enforcement, etc. Most of the governments also use DPI for the purpose of surveillance and censorship.

DPI helps when a client connects to a webserver and wants to propagate a worm or when a website tries to install malware on the system via HTTP session.

Application level firewalls: This kind of firewall is far more secure compared to the other kinds. In this kind of filtering, there is no need for a general purpose mechanism to let many different kinds of traffic pass or deny. In contrast a special code can be used for each special application. In this kind of filtering there is no need to be worried

about different conflicting rules in the firewall as well as holes in thousands of hosts offering nominally secure services. In contrast to the others, this kind of filtering is very easy in terms of working with the logs and controlling the input and output traffic.

These firewalls are used by users in order to enable them to use a proxy to communicate with secure systems. It also hides valuable data and servers from potential attackers. In a nutshell these kinds of firewalls are very easy for users to work with.

2.3.2 DNS tampering

The Domain Name Service (DNS) is used to translate the URL address to the IP address so that computers can communicate with each other. DNS tampering makes inaccessible all the services offered by a specific domain name. These services could be web access, chat, or any kind of files that the user has requested from the server. The effect of DNS tampering is very similar to the effect of IP filtering. One of the examples of using this kind of filtering is that when a request from a client is received, it will be redirected to the server under control of censorship instead of providing the user with the correct IP address. China is one of the countries using DNS tampering in the "Great Firewall of China". This practice prevents access to websites which they may find objectionable.

2.3.3 BGP route manipulation

Border Gateway Protocol (BGP) is an inter-domain routing protocol used for global routing at the level of Autonomous Systems (ASes). BGP select the route based on path, network policies and rule sets. Routers on the Internet change BGP information to find out the available path for sending the packets. Routers also use this information for updating their forwarding tables. By disabling the routing process on the border routers or overwhelming BGP, a large portion of network will be affected or even

unreachable. This technique was used by Egypt and Libya. At the time of black out the government of Egypt and Libya completely cut off all routers.

2.4 Effects of content filtering and traffic manipulation

There are different strategies which can be used to implement Internet censorship. Different countries have different rules for implementing their filtering. Figure 2.1 shows three different possible zones where filtering could take place, for example, on individuals' computers, organizations, ISPs, or backbone. Depending on the location it is applied to, it will have an effect on different numbers of people. For example, if filtering is applied to the backbone, it affects the entire country. If it is applied at the organization or ISP level, it affects that portion of the network. If it is applied at the individual level, it affects that specific computer. A description of each level is given below [2].

2.4.1 Internet backbone

The blocking practice may happen at the backbone. This method of filtering will affect the entire country. International gateways are the places where this level of filtering will be applied.

2.4.2 Internet service providers

Most of the filtering that governments intend to apply will be conducted through Internet Service Providers. It depends on the ISPs to select filtering methods.

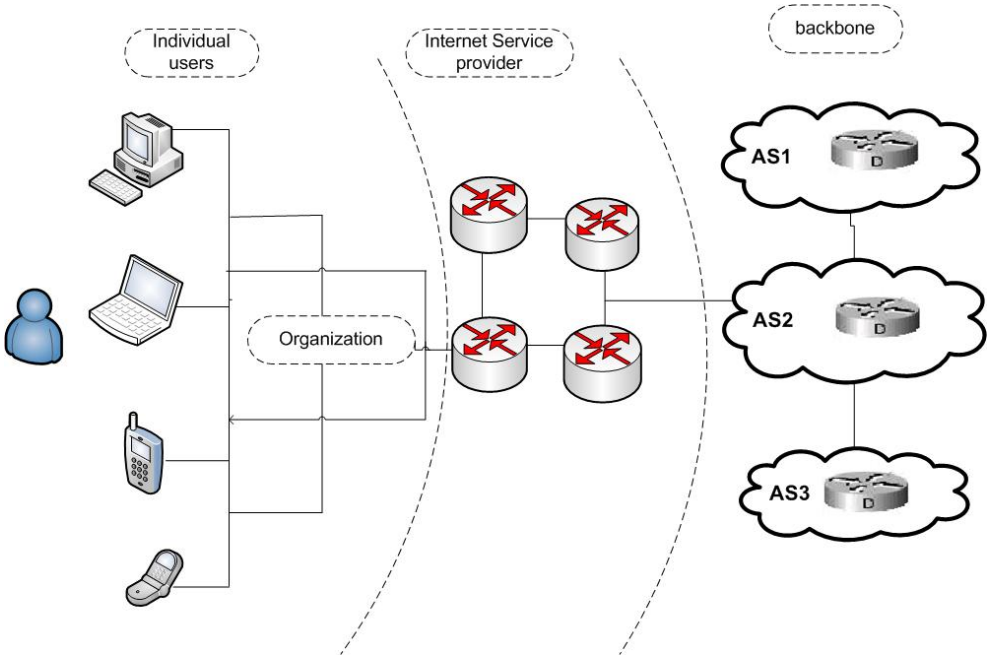

Figure 2.1
Different placement zones of censorship

2.4.3 Organizations

Some organizations have their own internal objectives and accessibility for their users. In this situation, the organization will use its own filtering method. For example, there are different rules for places such as companies, government organizations, schools and cyber cafes.

2.4.4 Individual computers

The restriction can be applied to individual computers. Censorship can happen through installing filtering software on a specific machine. There are different kinds of applications for parents for the purpose of controlling and monitoring [42]. These applications can do content blocking. Some of these applications are Net Nany 2.0, AVG family safety, K9 Web protection browser, etc.

These applications are different in their operations. For example, by using Net Nanny, parents would operate like traditional content blocking censorship. By using AVG Family safety, parents can monitor the web activity of their children. By using K9 web protection browser, the content of the porn and malware websites will be blocked.

2.5 Circumvention techniques to bypass censorship

When content blocking happens, the original version of the blocked content is not removed from the Internet. There are different Internet censorship circumvention technologies by which accessing blocked content would be possible. Some people use circumvention tools to pass the filtering and gain access to the content. But the others cannot use the same tools. There are different techniques to circumvent [73] filtering such as mirroring, additional DNS, changing IP address, the port number, and search engines which has been briefly explained below.

Mirroring: Mirroring refers to a technique which makes the contents available from different sources. Mirroring technique is helpful for bypassing IP filtering, DNS tampering, and filtering HTTP proxies. If users intend to use this technique, they should know the address of the mirrors.

Additional Domain Names: Additional Domain Names refer to the technique in which different domain names point to the same content. This technique is helpful for bypassing the DNS tampering and filtering HTTP proxies. If users decide to use this technique, they need to know the address of the DNS.

Changing IP address: In changing IP address technique, the IP address should change several times a day. Therefore, it is a sophisticated

technical task in order to bypass censorship.

Port changing: Port changing refers to changing the port number in order to bypass layer four filtering as well as filtering HTTP proxies. By using this technique, user will achieve the content through different port numbers.

Search engines: Search engines can be used in order to circumvent censorship. For example, Google has a feature called Google cache which can be used to achieve the cache version of the blocked content.

Proxy server: One of the very common ways to bypass censorship is using a computer outside of the country as a proxy server. It is easy for a computer to work as a proxy. Psiphon [48] can be used by any computer outside the border of censoring countries. By using this software the computer can work as a proxy for the other computers within the censoring territory.

VPN server: The other way to bypass censorship is through using a virtual private network (VPN) [24]. VPNs provide secure data transmission across the Internet. There are different VPN servers around the world. These numerous VPN servers will allow users to connect anonymously with different IP addresses.

Email: There are other techniques for gaining the desired censored content. For example when the user does not have access to specific content, he/she can send an email to www@web2mail.com [23] and the content will be provided through email.

TOR (The Onion Router) [37] is an example of a circumvention tool. It is an online anonymity system which conceals a user's location by routing the client's traffic through a network of servers. This means that people who conduct traffic analysis or censorship will not understand client's

activities such as visiting the web sites, online posts, and other communication forms. Anonymity tools like Tor aim to protect users' personal freedom and privacy.

There are other systems for the purpose of bypassing filtering, such as Freegate [19], Ultrasurf [57], and Psiphon [48]. All of these systems allow the user to be connected to one proxy outside the territory of censorship in order to gain the desired content. There are other approaches to circumvent censorship such as semantic overlay network [67]. Backes et al. [67] have used clouds in order to provide anonymity instead of using peers.

2.6 Related work

Free and open access to the Internet is an increasingly important topic. Many studies have recently been conducted in order to detect whether Internet content is being blocked or manipulated.

John-Paul Verkamp et al. [74] studied censorship in 11 countries around the world namely China, Bahrain, Malaysia, South Korea, Bangladesh, Thailand, Iran, Saudi Arabia, Russia, India, and Turkey. They used Planet Lab nodes [44] in the countries where these nodes were available. And in other countries where Planet Lab nodes were not available, they recruited volunteers through their personal contacts. They have found that different countries use different techniques for implementing censorship. For example, censorship at DNS level was used by Malaysia, Russia, and Turkey. They tested other ways of implementing censorship such as censorship at IP address which is used by China and Saudi Arabia. The other countries in their test were doing censorship using IPs, URLs, or keywords. They tested to check whether the URLs were accessible or not through using individual computers and Planet Lab nodes. For each website they tried to perform DNS resolution in order to get the IP address. If they did not get the same result, they would know that the content was being censored.

Dainotti et al. [70] studied Internet outages which were caused by cen-

sorship in Egypt and Libya. Their analysis was based on BGP inter domain routing control plane data, data plane traffic, trace route measurements, and RIP delegation files and MaxMind's geolocation databases. The governments in these countries shut down the Internet by disabling the routing process on critical routers or by suppressing transmission of BGP information. In order to observe the effect of blackout in these two countries they used the Internet numbering resources by finding IP address, BGP prefixes, and Autonomous System (AS) number of these two countries. They used Network Telescope in order to analyze the arriving traffic.

Andreas Sfakianakis et al.[76] improved the way of detecting Internet censorship by finding a way to differentiate network censorship from network failure. They could find not only the censorship but also the technology used to implement censorship. In order to run the experiment they used Planet Lab nodes around the world. They ran CensMon agent on the Planet Lab nodes. Users entered the URL plus the agent they wanted to send the request to. They used Google alerts to insert the interested URLs to their system. They also used Twitter [56] and Google Hot trends[20] in order to extract the popular trends.

Another study was conducted by the OpenNet Initiative [4]. This study compared different levels of filtering in 60 countries. They compared the level of filtering in political, social, security, and overall aspects. But unfortunately this study was not conducted in New Zealand. The result of this study is shown in Figure 2.2 and 2.3.

The developers of this system checked it by two lists of websites: a global list which was constant for each country and a local list which was different for each country. They implemented their software. The software designed to query the lists of URLs was prepared for a specific country. The list was accessed over HTTP both in a country with Internet filtering and a country with no filtering. Then the data gathered from the country with no filtering was compared with the data from the country with filtering. They tried to distinguish the connectivity errors from intentional

Country	OpenNet Evidence of Filtering Levels								Freedom House	Overall Rating
	Political		Social		Security		Overall			
	2007	2009	2007	2009	2007	2009	2007	2009		
Armenia	-	M	-	L	-	L	-	M	-	Medium
Australia	-	NE	-	M	-	NE	-	M	-	Medium
Azerbaijan	L	L	NE	L	NE	NE	L	L	-	Low
Bahrain	M	-	L	-	NE	-	M	-	-	Medium
Belarus	NE	L	NE	L	NE	L	NE	L	-	Low
Brazil	-	-	-	-	-	-	-	-	Low	Low
China	H	H	M	M	H	H	H	H	High	High
Cuba	-	-	-	-	-	-	-	-	High	High
Egypt	-	NE	-	NE	-	NE	-	NE	Medium	Medium
Estonia	-	-	-	-	-	-	-	-	Low	Low
Ethiopia	M	-	L	-	L	-	M	-	-	Medium
France	-	NE	-	NE	-	NE	-	NE	-	NE
Georgia	-	L	-	NE	-	L	-	L	Medium	Medium
Germany	-	NE	-	NE	-	NE	-	NE	-	NE
India	NE	-	NE	-	NE	-	L	-	Medium	Medium
Iran	H	H	H	H	M	M	H	H	High	High
Italy	-	NE	-	L	-	NE	-	L	-	Low
Jordan	L	-	NE	-	NE	-	L	-	-	Low
Kazakhstan	NE	L	NE	L	NE	NE	NE	L	-	Low
Kenya	-	-	-	-	-	-	-	-	Medium	Medium

Figure 2.2
Meta-analysis of international surveys of filtering (1)[4]

tampering. This study illustrated that the countries with more filtering were China, Cuba, Myanmar (Burma), Oman, South Korea, Sudan, Syria, Tunisia, Turkmenistan, United Arab Emirates, Uzbekistan, Vietnam, and Yemen.

Freedom House who studied censorship in 47 countries [18] demonstrated that Internet censorship was growing in many countries. Their study also proved that with the improvement of the methods of controlling the Internet, it was more sophisticated to pinpoint censorship. They concluded that different attacks against bloggers, political surveillance, manipulation of web content, and restrictive laws regulating speech online were threats to Internet freedom. They divided the level of censorship to four levels, namely free, partly free, not free, and no data as shown in Figure 2.4 [18]. Green color is showing free. Yellow color is showing partly free. Purple color is showing not free. Gray color is showing that there is no data for that area.

Kyrgyzstan	-	L	-	L	-	NE	-	L	-	Low
Libya	M	-	NE	-	NE	-	M	-	-	Medium
Malaysia	-	-	-	-	-	-	-	-	Medium	Medium
Moldova	-	L	-	NE	-	NE	-	L	-	Low
Morocco	NE	-	NE	-	L	-	L	-	-	Low
Myanmar	H	H	M	M	M	M	H	H	-	High
Oman	NE	-	H	-	NE	-	H	-	-	High
Pakistan	L	NE	M	M	H	M	H	M	-	Medium*
Russia	-	L	-	L	-	NE	-	NE	Medium	Medium
Saudi Arabia	M	-	H	-	L	-	H	-	-	High
Singapore	NE	-	L	-	NE	-	L	-	-	Low
South Africa	-	-	-	-	-	-	-	-	Low	Low
South Korea	NE	NE	L	M	H	H	H	H	-	High
Sudan	NE	-	H	-	NE	-	H	-	-	High
Syria	H	-	L	-	L	-	H	-	-	High
Tajikistan	L	L	NE	NE	NE	NE	L	L	-	Low
Thailand	L	-	M	-	NE	-	M	-	-	Medium
Tunisia	H	-	H	-	L	-	H	-	High	High
Turkey	-	L	-	L	-	NE	-	L	Medium	Medium
Turkmenistan	-	H	-	L	-	L	-	H	-	High
UAE	L	-	H	-	L	-	H	-	-	High
Ukraine	-	NE	-	NE	-	NE	-	NE	-	NE
United Kingdom	-	NE	-	NE	-	NE	-	NE	Low	Low
United States	-	NE	-	NE	-	NE	-	NE	-	NE
Uzbekistan	M	H	L	L	NE	L	M	H	-	High*
Viet Nam	H	-	L	-	NE	-	H	-	-	High
Yemen	L	-	H	-	L	-	H	-	-	High

Key: Collapsed ratings from different studies into one more general rating:

For OpenNet:

L, M, H = Low, Medium, High;

NE = no clear evidence, including 'suspected filtering';

Low = evidence of selective filtering;

Medium = 'substantial filtering';

High = 'evidence of pervasive filtering' (Faris and Villeneuve 2008:

For Freedom House Ratings:

Low = 10-26, rated 'Free';

Medium = 27-55, rated 'Partly Free';

High = over 55, rated 'Not Free' (Freedom House 2009: p. 20).

A '-' indicates that the country was not covered by the respective study / in the respective year.

'*' indicates that the rating for the respective country has changed from 2007 to 2009.

Figure 2.3
Meta-analysis of international surveys of filtering (2)[4]

Marcel Dischinger et al. [72] studied BitTorrent traffic over 17 weeks by 47300 end users in 1987 ISPs. They found that ISPs in countries like USA and Singapore blocked uploaded BitTorrent traffic more than other countries. It was their first step toward making transparency for users. BitTorrent which is a popular peer to peer file sharing protocol uses a large portion of data bytes over the Internet which is costly for the ISPs. Therefore, ISPs used different strategies to reduce the generated BitTorrent traffic by users.

Marcel Dischinger et al. [71] studied ISPs' performances and found that high packet loss and jitter had an effect on the transport protocols that relied on round trip times and consequently on the performance of the ISPs. Also real time application such as VoIP application would be affected from large queue size that ISPs set for their traffic settings. They

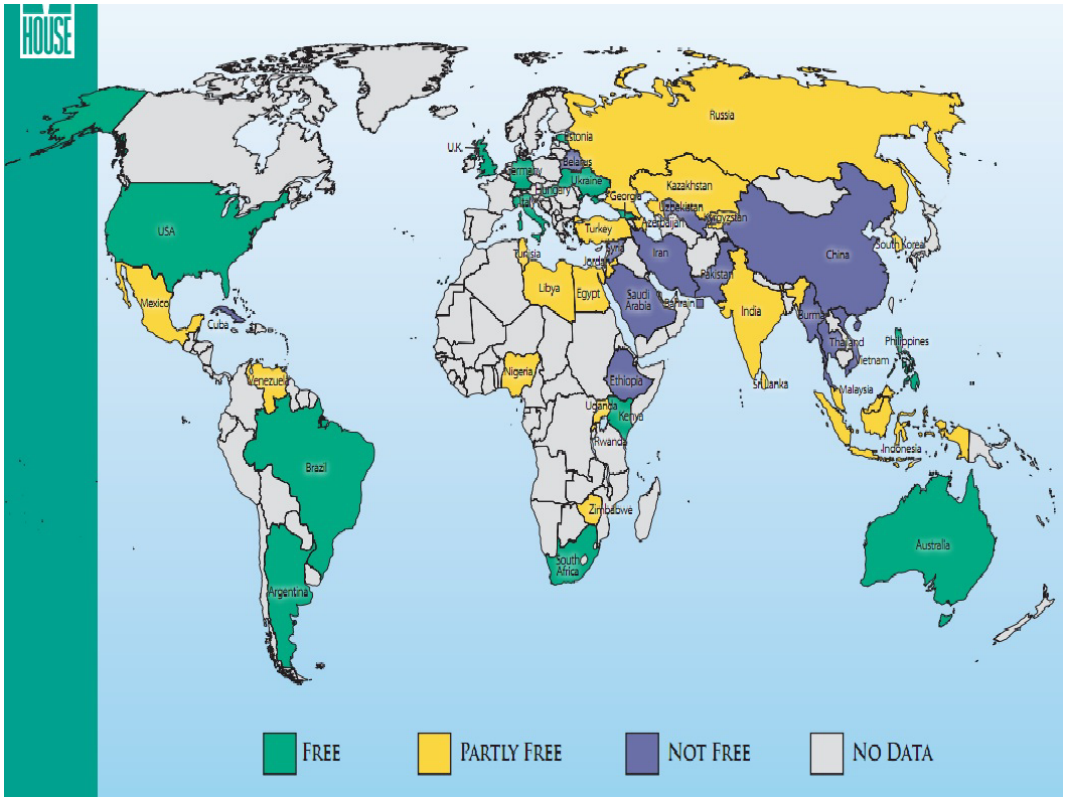

Figure 2.4
A global assessment of Internet and digital media [18]

also showed that when BitTorrent application was used by VoIP application in the same shared bandwidth, VoIP applications' performance would get affected. It made another problem for those real time applications such as VoIP applications. Hence, the ISPs would change the settings and give the prioritization to specific traffic. There are other ways ISPs use when they aim to manipulate traffic such as rate limiting, message dropping, or altering the content.

In order to compare performance of different ISPs, Srikanth Sundarean et al. [78] studied different ISPs and prepared the table depicted in Figure 2.5 [78]. Each of these metrics was important for specific applications. If different ISPs are compared based on these metrics, then users will be able

to select ISPs based on their requirements and the application they have used more.

Metric	Why it matters
Sustainable throughput	Throughput for long transfers
Short-term throughput	Throughput for short transfers
Minimum throughput	Captures network load
Baseline last-mile latency	Interactive applications
Maximum last-mile latency	Captures network load
Maximum jitter	Real-time/multimedia
Loss rate	TCP throughput/multimedia
Loss burst length	TCP throughput/multimedia
Availability	Basic reachability

Figure 2.5
Performance metrics for the network nutrition label [78]

Srikanth Sundarean et al. [77] checked the performance from home gateway devices. They tested the throughput and latency from 4000 gateway devices. They looked for reasons affecting the users' performance such as users' modem and ISPs' different traffic shaping. The difference between their test with the previous tests was that in those studies the experimenters just checked the upload and download. Another shortcoming of those studies was that they measured the speed once and not several times because the previous studies ran their tool through end host and did not take into account the effect of confounding factors. Some examples of confounding factors included home network cross-traffic wireless network, firewall at the user side, or end host configuration.

Previous studies reveal a gap in our understanding of what is being censored in New Zealand, who performs censorship in New Zealand, and what the reason for performing Internet censorship is. In the following sections we review filtering in New Zealand and free Internet providers in Wellington, New Zealand.

2.7 Filtering in New Zealand

Based on 2012 New Zealand national census, the population of New Zealand was 4,449,768 as on Thursday, 20 Dec 2012 at 10:03:54 am [46]. A study [55] conducted by the Ministry of Social Development showed that in 2006, 66 percent of households had access to the Internet. The percentage increased from 43 percent in 2001 to 66 percent in 2006. In December 2009, 75 percent of households had Internet access at home which is another increase from 66 percent in 2006. The Internet is provided by around 70 ISPs in New Zealand [30]. The most popular ISPs in New Zealand are TelstraClear, Vodafone, and Telecom.

In New Zealand the Film, Video, and Publication Classification Act of 1993 provides a mechanism for classifying publications based on whether or not their content is considered to be objectionable, or suited only to specific age groups. The Act covers films, printed publications, computer games, computer files, books, and videos/DVDs. Because the Act applies to computer files, including images and other types of documents available on the Internet, it is possible for the Office of Film and Literature Classification to classify websites, though this happens very rarely, usually at the request of a law enforcement agency [11].

To understand the ratio of sexual and violent material on the Internet, Zimmer et al. [49] studied the contents of web sites and showed that 3.8% of the materials on the Internet contained graphically sexual or violent content. But even these small portions of web sites are easy to access.

In order to prevent the use of these materials, the 1996 Act was amended. Based on the Act, if someone tries to access this content, he/she would be considered guilty of trading objectionable material. The Act is silent regarding ISPs and therefore they cannot be forced to censor. Hence, ISPs can provide these objectionable materials to users.

The New Zealand Department of Internal Affairs (DIA) [14] provides a voluntary filtering system called Netclean Whitebox [35] that blocks objec-

tionable content. While some ISPs have signed up for this service, others have not. Customers whose ISPs have signed up for this filtering service may not be aware that their Internet access is subject to this invisible form of censorship. DIA does not reveal the list of banned sites, and attempting to access URLs is the only way to check whether a particular website is filtered. Based on section 6 (c) of the Official Information Act of 1982, they are allowed to leave unnamed their blocked web sites. Introduction of the filter was controversial. Some ISPs were in favor of it, as shown in the following quotes [62]:

Our customers would be disappointed to hear if we were not participating. So participation for us has always been a no-brainer.(Maxnet), We informed our customers of the trial, received positive feedback from them and it is likely we will participate further. (TelstraClear), It's a no-brainer for us, it's free. (Xtreme Networks), Telecom has announced it joined the Department of Internal Affairs' Digital Child Exploitation Filtering System. (Telecom).

In New Zealand, filtering works through the following steps. First, the Department of Internal Affairs (DIA) [14] prepares a list of banned sites and their Internet addresses. Second, DIA will inform the participating ISPs that the best way to get the Internet address of the banned sites is through DIA's filtering server. Therefore, when a user needs to access a URL, their ISP diverts the request to the DIA's server. DIA's filtering server will either respond to the request or refuse it. The server will respond through DIA's Internet connection.

2.8 Free Internet in Wellington New Zealand

There are many places [54] in Wellington New Zealand where Internet access is offered free to their customers such as CBDfree Wi-Fi, National Library of New Zealand, Wellington Airport, Airport Flyer buses, Te Papa Museum, McDonald's, and Starbucks. In addition, there are many small and independent cafes and restaurants which provide free WiFi hotspots.

They offer free Internet but with conditions. There are different policies for different organizations. The problem, though is that these polices are not clear for users. For example, McDonald's [26] blocked gay websites in New Zealand. The fast food chain's Wi-Fi network was hacked by a member of New Zealand's gay community because it censored gay web sites.

Wellington City Council (WCC) [60] said that the free Central Business District Wireless Internet (CBDfree WiFi) would be censored and it was possible that free Wi-Fi would be restricted in the future to specific web sites. Wellington City Council would block some websites but, other than that, Internet is open and accessible to most Internet sites. With prioritization in the free Wi-Fi traffic for example, users can do web browsing and check email but they cannot use free Internet for streaming or downloading large files.

Philippa Bowron [61], Wellington City Council's strategy adviser stated that there were no plans to use content filtering to ban offensive Internet sites. They thought that it was not the organization's responsibility to do censoring of the content. Such being the case, they did not have any issues with Waterfront WiFi and in case any problem happened, the right in the agreement would reserve the implementation of filtering.

Internet censorship has its opponents as well. In New Zealand, there are some organizations which are against filtering [31] and they believe that central filtering does not respect human rights. Two of these organizations are TechLiberty and InternetNZ. Their opinions are discussed in the following sections.

Tech Liberty group [5] which is a non-government group was established to protect people's rights on the Internet. The group is responsible to educate users about their rights. If someone's rights are infringed, they teach them how to defend themselves. Tech Liberty's main focus is to stop filtering and it has published a number of articles regarding this matter.

Another organization which is dedicated to protecting and promot-

ing the Internet in New Zealand is InternetNZ (Internet New Zealand Inc.)[31]. The main goal of this charity-based organization is to provide high performance and unfiltered access for all New Zealanders. It has eight main policies to ensure the transparency and predictability of the approaches taken to Internet. InternetNZ's goal is to keep the Internet "open and uncaptureable" but cyber attacks, politically motivated censorship, and government control over Internet infrastructure are emerging as a threat to the Internet [40].

InternetNZ believes that a centrally monitored government filtering system is not the answer. Even though many parents assume the DIA filtering system provides a safe environment, the objectionable material may still be available, since filters are not always successful [28]. In January 2010, InternetNZ [29] asked the Department of Internal Affairs to undertake a study of the extent of access of child abuse material on the Internet and the most effective ways to address this problem. InternetNZ argues that filtering is not subject to all the checks and balances that apply to all other parts of New Zealand's censorship regime.

Jordan Carter [32] also believes that the solution is not centralized government filtering. On the contrary, it will be a better solution if individuals themselves choose filtering and apply that to their computers. For example, Netsafe [3] can be used to educate and support individuals by promoting cyber safety and champion digital citizenship.

ONI [7] compared the openness of Internet in New Zealand with a close neighbor Australia. Australia follows the same policies and regulations for Internet as many other western countries. However, the laws toward Internet in New Zealand are less severe. Generally speaking, Australia's censorship regime for Internet is harsher than that of New Zealand and other Western countries. Although Australia is going beyond the norms of filtering compared with the more democratic states, its filtering is not as much as or comparable with the repressive regimes that Opennet Initiative has investigated [16].

2.9 Different web censorship monitoring tools

This section reviews various tools proposed for doing blocking testing because we want to select the best one to carry out a New Zealand study. For this purpose, different censorship monitoring tools based on the following criteria were compared.
1) Is the tool an open source?
2) Is the data collected made public?
3) Is the data format used for publication easy to interact with?
4) Are the methodologies explained?
5) Is the tool able to be used by the general public?
6) Does the data collected by the tool include potentially sensitive information?
7) What kinds of tests does the tool perform?
8) How accurate are the tests?
9) Is confidentiality and integrity of reported data being maintained?
10) What are its strengths?
11) What are its weaknesses?

It was important to find out what these tools measured because the goal was to measure the prevalence of censorship in terms of content and service blocking. It was important whether the gathered data was publicly available and easy to interact with because the data needed to be opened for other researchers. It was important that the tool be publicly available online so that they could be easily used. Also it was important to know whether the data gathered by the tool included sensitive data because we did not want to make security issues for the users, so we also compared them in terms of considering confidentiality and integrity of the data.

A number of existing tools have been developed to either detect content blocking, or identify traffic manipulation. Table 2.1 and 2.2 shows a summary of the characteristics of these tools. They were categorized

based on whether they were detecting censorship or whether they were recognizing traffic manipulation.

Bismark, Netalyzer, NeuBot, Switzerland could be considered as traffic manipulation detection tools. Alkasir, CensMon, Herdict, ONI, 403 checker could be considered as censorship detection tools. OONI is considered in both categories.

2.9.1 Traffic manipulation detection tools

Traffic manipulation can be applied through a small manipulation in the header field, by completely changing the traffic, giving priority to certain type of traffic, preventing special traffic to pass, etc. Traffic manipulation tools, as it is clear from their names, look for manipulations in the traffic. They aim to detect the presence of some sort of tampering with the Internet traffic between the client and a server. For example, Switzerland [53] looks for the changes, inserted or missing packets, between the client and the server. NeuBot [36] focuses more on BitTorrent traffic and identifying the ISPs which block BitTorrent. BISmark [58] is a tool for measuring ISP performance. Netalyzr [25] is a relatively complete product for checking the services. OONI [38] is a product that considered both traffic manipulation and content blocking. A description on each of the tools has been given below.

BISmark: BISmark [58] is a tester for the Broadband Internet Service. This project was conducted by the cooperation of Georgia Tech and the University of Napoli Federico. BISmark provides insight about the ISP's low level operations and is written in C, Python, Perl, and bash script. It also has the ability to visualize and monitor traffic patterns from users and devices inside home network, and manage usage caps. It is provided for the public use. Its data is also online.

BISmark is a combination of network analytic tools. By using this tool we can get an insight about ISP's traffic manipulation, priori-

tization, and discrimination. It provides a safe communication for sending the user critical information. The designers used tunneling to the router to run the test as well as collect the information in order to maintain privacy.

Netalyzr: Netalyzr [25] is developed by an independent nonprofit research institute affiliated with the University of California, Berkeley. Netalyzr has been written in Java and it shows and tests various properties of the Internet connection. This project is not an open source.

This checking includes blocking important services, HTTP caching behavior and proxy correctness, the DNS server's abusing, NAT detection, as well as latency & bandwidth measurements, and it will report all the findings. It runs as a Java applet on the computer. The data is not secure at all. It will be sent to the wire without considering the integrity and confidentiality.

The problem is that this tool is not an open source but is available for public use. If someone participates in their test, then the user can access the collected data.

NeuBot: NeuBot [36] is another project on network neutrality. This project is also an open source. When the program is run, it will be tested with the server and all the details will be sent to the server. The servers are hosted on the Measurement Lab platform. The Neubot team is based at the NEXA Center for Internet & Society at Politecnico di Torino. The project manager is Simone Basso. He developed Neubot in collaboration with Dr. Antonio Servetti, Dr. Federico Morando, and prof. Juan Carlos De Martin.

This project is written in Python and is fully prepared from scratch for Windows, Debian, Ubuntu, and MacOs. The data that the tool has gathered is freely available on M-Lab. This tool also measures the performance of the Network. The speed is checked by using

HTTP protocol. It measures the round trip time, download, and upload data. It also identifies BitTorrent traffic and if the ISPs were blocked by BitTorrent traffic. Some side effects of using this tool are: because the tool keeps the IP of the user it will not be safe. Neither will the data be encrypted for transferring to the server which can be affected by attackers.

OONI: OONI [38] is a web censorship monitoring tool which collects high quality data from the network. This data includes information such as types, methods and amount of surveillance and censorship. It shows the reasons for filtering as well as the network interface of the filtered network. Tests in OONI are divided into two subcategories; traffic manipulation and content blocking. Developers Arturo Filasto and Jacob Appelbaum are the co-creators of OONI-probe who are from the Tor project. This application is an open source and written in Python. The items such as URLs, keywords, IP addresses should be given as input to the application.

Switzerland: Switzerland [53] was developed by Electronic Frontier Foundation. It is an open source tool. This tool is mostly written in Python but C also was used for speed improvements. It is implemented to check whether the data between the network has integrity or not. It will find the modified data packets over the IP network and inform the user. This software has a server and many clients. The server will check whether any packets are dropped, forged, or modified whenever the clients transfer the packets between each other. The data will be encrypted and then sent to the server but it does not send it as anonymous. The alpha version of the software is released for public to use. It runs through command line, so it needs a level of expertise. The data this project collects is not available because it is sensitive user information.

2.9.2 Censorship detection tools

Censorship detection tools, as their names suggest, try to find if there are contents which have been blocked by trying to access them. Censorship detection tests are aimed at counting the kind of content that is blocked. For example, ONI [4] and OONI [38] try to access the website from an experiment network and compare the result with a control network's result in order to find content blocking. 403 Checker [1] and Herdic [21] and Alkasir [6] try to access the URLs and if they fail, it will be considered as censorship. The problem with these tools is that it relies on users' feedback which can be false negative since users cannot identify network failure from network censorship. The benefit of using Alkasir is that this tool can also be used as a circumvention tool. CensMon [76] is another web censorship monitoring tool which operates automatically and it does not rely on users' feedback. A brief summary of each of them will be given below.

Alkasir: Alkasir [6] is a censorship detection tool plus a circumvention tool. Walid Al-Saqaf is the chief programmer of Alkasir. This tool is different from the other circumvention tools because other circumvention tools route all the traffic through a proxy. They do not make any exceptions for non-blocked websites. For example, in the traffic of web emails (Gmail, Yahoo mail, etc.), online bank accounts, or online shopping, all the details are transferred through a proxy server which is dangerous.

In contrast, Alkasir is used when you browse blocked websites. Hence, this solution is more secure because the data is not transported when there is no need for that. Furthermore, this tool does not save the IP addresses of its users. It also prepares the list of blocked website based on the user feedback. The tool is applicable for Windows users and it is not an open source.

CensMon: CensMon [76] is another web Censorship Monitor tool which

operates automatically. This tool was written by Andreas Sfakianakis, Elias Athanasopoulos, and Sotiris Ioannidis in the Institute of Computer Science, Foundation for Research and Technology, Hellas. It can differentiate access network failure from censorship. It identifies the filtering technique used by the censor. The writers evaluated their system and their evaluation showed that CensMon could successfully detect censored content.

Herdict: Herdict [21] is a tool developed by the Berkman Center for Internet & Society to identify the web blockages. The test includes identifying denial of service attacks, censorship, and other filtering. When a user can't access the content they want, a report will be sent to the Herdicour. This application relies on users' feedback. That is why sometimes the results are not accurate, because some users cannot identify the difference between network connectivity problem and blocked websites. This data will help them monitor the health of the Internet in real time, but it is not secure.

ONI: The OpenNet Initiative [4] is a project resulting from the cooperation of four institutions: The Citizen Lab at the Munk School of Global Affairs, University of Toronto; the Berkman Center for Internet & Society at Harvard University; and the SecDev Group (Ottawa).

This project tries to identify and analyze Internet filtering. This information will be helpful for informing better public policy. It is not an open source and is not available for users to test. This is just a research tool. Up until now, this tool has been run in 76 countries and the results are available for each country. Because ONI relies on users' feedback, it will cause false positive since users will not be able to differentiate network failure from censorship.

403 Checker: 403 Access Denied Checker [1] is implemented by the Tunisian blogger and activist Astrubal. This tool was designed to test local

blockages. It was not developed to circumvent censorship. This tool provides the ability for the user to check a huge number of URLs. Then the results will be published to the public. It does not save the IP address which makes it safe to use. The alpha version of this application has been released.

The problem is that there is no link to this software anymore. The first version of the software was released in 2007. But it was good for novice users to just check their list of URLs to understand whether they were being blocked or not.

Projects	Comprehensiveness	Supported platforms	Open source	Language	Level of Expertise	Open Data	Security
Bismark	Upload, download speed test, jitter, delay and packet loss	-	Yes	-	Needs knowledge	Yes	User privacy was considered
Netalyzr	blocking of important services HTTP, DNS	-	No	Java	Easy	after the tests are completed	Not at all
NeuBot	BitTorrent test - Speed test	MacOS, Ubuntu, Windows, Debian	Yes	Python	Yes	M-Lab	Not encrypted
OONI	Traffic manipulation-Content blocking	Windows, Mac/OSX, Linux, BSD	Yes	Python	Command Line	Yes	Yes
Switzerland	Identifies ISP interference at the packet level	Windows, Mac/OSX, Linux, BSD	Yes	C, Python	Command line-need knowledge	N/A	Encrypted but not anonymous

Table 2.1
Comparison between different traffic manipulation tools

Projects	Comprehensiveness	Supported platforms	Open source	Language	Level of Expertise	Open Data	Security
Alkasir	Censor circumvention plus listing URL block list	Windows	NO	.Net Framework	Easy	NO	NO
CensMon	Blocked URLs	-	No	-	-	Yes	-
Herdict	-	Not On Google Chrome	Not completely			Yes	No
ONI	Censorship detection	-	No	-	-	Yes	under license
OONI	Traffic manipulation-Content blocking	Windows, Mac/OSX, Linux, BSD	Yes	Python	Command Line	Yes	Yes
403 Checker	Checking blocked URLs	Windows, Mac/OSX, Linux	No	-	Easy	Yes	Safe

Table 2.2

Comparison between different censorship detection tools

2.10 Tool selection

Our comparison of different web censorship monitoring tools led us to use OONI since this tool met all of our criteria so we initially selected it. The architecture of OONI is explained in the following section.

2.10.1 OONI architecture

In this section, we will explain how OONI works and how different parts of OONI project are interacting with each other. In Figure 2.6, the schema of the OONI framework has been shown. There are two main components in their framework namely, *Oonib* and *Ooniprobe*.

Ooniprobe is the client side of the project. The tests are run by *Ooniprobe* on the experimental network. *Ooniprobe* runs the tests through test template. Test templates are responsible for preparing the settings for measurement and also providing error handlings. In addition, it provides a specific template for each specific test. *Oonib* acts as a backend component for OONI. All the results will be collected by *Oonib*.

Testhelpers are responsible for implementing server side protocol. This acts as assistance when *Ooniprobe* runs the tests.

There are some steps for *Ooniprobe* to run. First, it starts to connect to Tor. Tor is used to have a known good channel for reporting the results. Tor is also used as the control network because OONI's project developers assumes that by using Tor, they can bypass any blocking that is happening locally. Second, it gets IP address from Tor. Third, it will send the result as a report by getting a report ID for submitting the result. Fourth, based on the inputs, it will chunk the inputs and run the test and update the collector.

The methodology for OONI tests is based on comparing the results from experimental environment with those from control environment. The experiment environment refers to the network suspicious of surveillance. The control environment refers to the uncensored network. The results

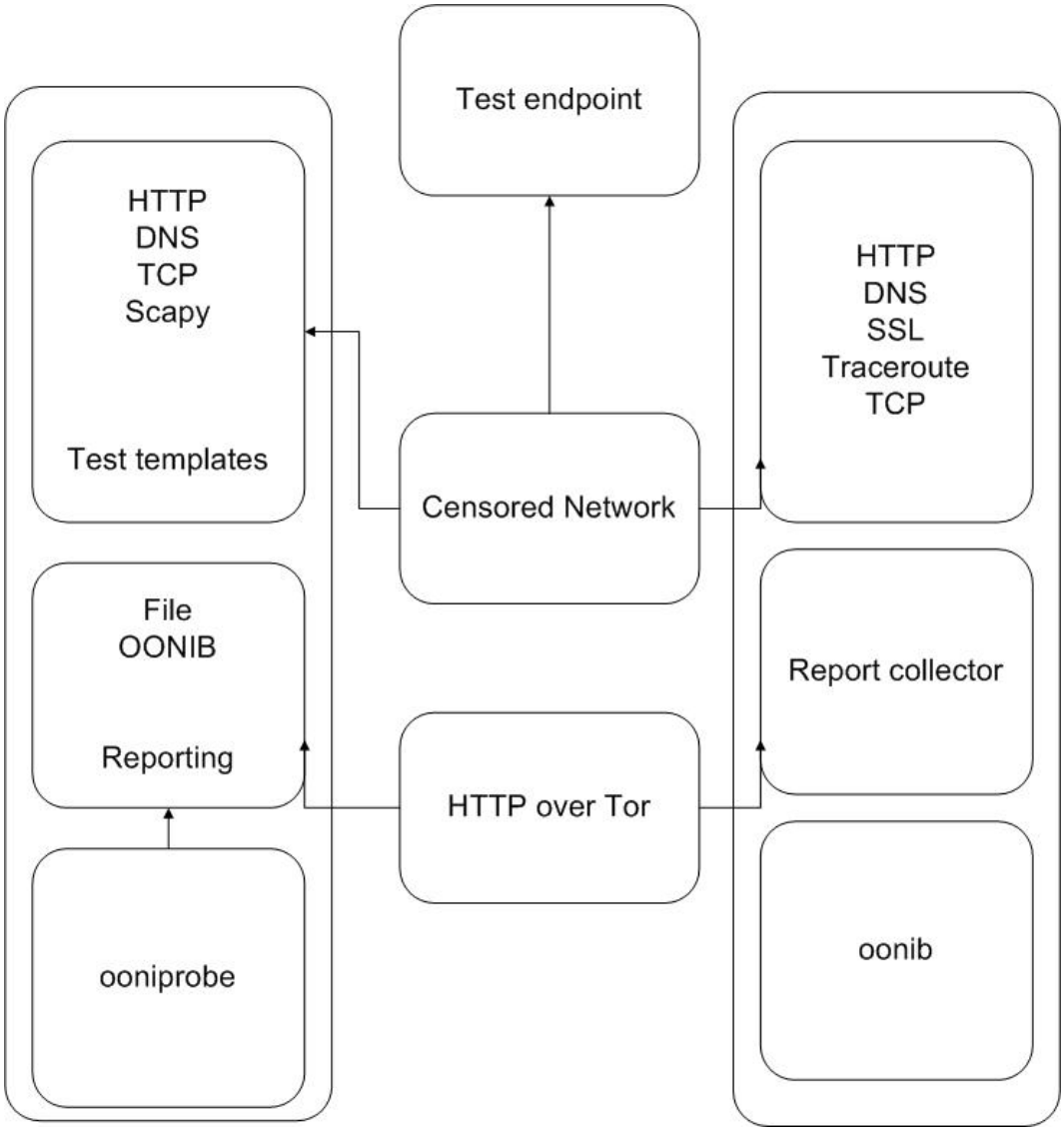

Figure 2.6
Overview of the OONI architecture

from these two experiments are compared in order to find mismatches. If there is a mismatch between the two results, censorship will be reported.

The differences between these two results indicate the presence of censorship. However, this may lead to a false positive due to reasons such as

dynamic content. As another example of giving false positive in the OONI methodology is when DNS results from two networks are compared. DNS results are specific and predictable for lots of records but there are lots of webpages which are geographically diverse. Comparing these two results will again lead to false positive.

2.11 Practical issues preventing use of OONI

Our preliminary work with OONI showed that we could not run OONI in our experimental networks, either. There were issues in running OONI in our experimental networks. In this section, we will summarize issues we have identified with OONI. Based on these issues, the requirements for a new tool have been identified.

2.11.1 Issue 1, using Tor

The first issue with OONI was that, in the OONI project, the results of the experimental network was compared with the Tor network. But Tor was blocked in most places. Hence the tests could not be run.

Solution 1

Initially, it was thought that it would be better to make changes to the OONI project to prevent using Tor. In this situation, offline comparing was used between the experimental network and a network which did not block Tor. But it was revealed that some of the content was changing very rapidly in less than an hour. Thus, the idea of comparing the results from a non-censored place with experimental network at a different time led to lots of false positive. Hence, the use of a web server was suggested.

2.11.2 Issue 2, using web server

Using a web server, the problem of finding censorship remained unsolved. Because by using web server we just passed censorship at an organization level but we were still affected by country level censorship.

Solution 2

To solve issue 2, using Tor at a web server was suggested. By using Tor at web server, we passed censorship at a country level and we would have a reliable uncensored network.

2.11.3 Issue 3, giving false positive

The third issue with OONI was that Requesthost test compared the header and body length of the two results from control network and experiment network with each other. If the test found any differences, it would report the manipulation in the result. This was not exactly the evidence of censorship. The difference between the lengths of bodies could be due to having different character encodings or having dynamic content. This solution which was used by OONI would give us a lot of false positives.

Solution 3

In order to decrease the number of false positive while calculating the body and header length, the content of the results should be compared with each other. It was also possible that the images may have been changed. Hence, it was essential to compare the images as well.

2.11.4 Issue 4, providing reliable endpoints for testing

In this section, there were few number of services which needed to be checked, some services such as BitTorrent was illegal to use but legal to

check. For this reason, there was a need to have a fake server with all the desired service on listening mode in order to check their reachability. Then from the client it was possible to connect to those ports at the server to understand whether it was possible to connect to those services or not. In New Zealand it might be true that most of the organizations did not block the websites but they had blocked lots of ports and services.

Solution 4

It is required to run our implemented server at the Victoria University's server and try to connect to different ports from those experimental networks. This test would give us the idea about reachability on different ports.

2.11.5 Issue 5, targeting bandwidth limitation

In OONI, the results would be sent back to the backend server. In our context since we were faced with the limitation in using organization's provided data, we could not use our limited data to send the result to the server.

Solution 5

In order to avoid using data, the result from control networks would be saved on the Victoria University of Wellington (VUW) servers and the result from experimental network would be saved on experimenter's computer. This solution prevented the creation of any security issues for the data.

2.12 Requirements

Based on the limitations of OONI, we set the following requirements for our new tool.

Requirement 1, Using web server: Since we did not have access to Tor directly at the organizations, we had to do proxy to access the content. Therefore, we had to write an oracle web server to handle our requests. The oracle web server had access to the network without censorship and the Tor network and provided two control networks for us.

Requirement 2, Reducing the number of false positives: In order to decrease the number of false positives in addition to what OONI relies on as evidence of censorship, we concentrated on the contents. In our methodology, the content and images of the experimental network's results should be compared with control network's result. Differences in the results would determine censorship.

Requirement 3, Checking for service blocking: There were places where URLs would not be blocked but they did block ports and services. It was essential to make them transparent for users. Therefore, we had to implement our own web server on listening mode on desired services and try to access its open services from organizations' network.

Requirement 4, Keeping the results separate: In order to decrease the usage of data provided for us in the organizations, we had to keep the result separate in our servers in VUW and in our computer in the experimental network and not send the data between client and servers.

2.13 Summary

In this chapter of the thesis, we started our discussion with an overview on Internet filtering and provided a discussion on background and related work. We also reviewed tools for investigating Internet filtering and used the results of this review to define the requirements for our new tool called Web Censorship Monitoring Tool. In the next chapter, the design and implementation of our tool based on the requirements defined in this chapter is given.

Chapter 3

Design and implementation

In this chapter, we provide an overview of the design and implementation of our new filtering detection tool called Web Censorship Monitoring Tool (WCMT), that addresses the problems identified in the previous chapter.

3.1 Design based on the requirements

The following requirements for our new tool were identified in the previous chapter.

Requirement 1, Using web server:

Requirement 2, Reducing the number of false positives:

Requirement 3, Checking for service blocking:

Requirement 4, Keeping the results separate: In order to satisfy the requirements identified in the previous chapter, the following design was proposed.

D1: In order to satisfy requirement (R1), our own implemented web server was used. The web server acts as a proxy and requests pages on behalf of the client from Victoria University open and Tor networks. This request

goes over a separate network which is not under the control of the organization doing the blocking. The result obtained from web server would be considered as the first control network. We also used Tor at the web server to have the second control network. It is clear that by using Tor at the web server, the request could go beyond the territory of the country. Getting result from Tor network gave us information which helped identify blockings by the government that affected the whole country.

D2: In order to satisfy requirement (R2), more investigation was needed.
It was proposed that in order to find content blocking in different organizations a focus not only on the length but also the content of the results were needed. Our methodology was based on comparing the overall length and content of three sources. It started to compare the headers and body length from those three networks. After that the body and header's contents of the results from Experiment, VUW, and Tor network were compared with one another in order to find differences. The result of comparing were saved in the third file. This third file was used in our manual checking. If the differences were not benign, then it would be considered as a censorship.

For more in-depth investigation, the images from three results were compared together in terms of name and size of the images. If these results showed any differences, this also showed evidence of censorship. In order to compare the images from our three different networks, MD5 was used. MD5 is the actual comparison of the binary data. Hence, if the pictures were in different formats, they would have different binary data. If hash values of two images were the same, it meant that they must have been identical.

With this solution, we were trying to reduce false positives where the content was different but the size and names were the same. Removing false positives completely was impossible but we could improve

OONI to remove some false positives and simplify usage.

D3: In order to satisfy requirement (R3), the following check was required.
In order to make sure that there was no problem with users' connectivity in the service reachability test the connection were performed on 22 well-known services [59]. In New Zealand, there may not be many content blocking by the ISPs and organizations but perhaps there are lots of port and service blockings. There are special ports which will be used by most people and we needed to check their connectivities. The list of these services was illustrated in Table 3.1.

In order to test these ports and services, we needed a server with all these services on listening mode. A server with all these services will use lots of resources from the server. Therefore, we implemented a server which we could set it to listen on those special ports. With this design, it was possible to try to connect to the server on these special ports from the client side. This test helped us measure the prevalence of service blocking.

Table 3.1
Ports and services to check

Service name	**Port number**
FTP	21
SSH	22
SMTP	25
DNS	53
HTTP	80
POP3	110
RPC	135
NetBIOS	139
IMAP	143

Continued on next page

Table 3.1 (*finished*)

Service name	**Port number**
SNMP	161
HTTPS	443
SMB	445
SMTP/SSL	465
IMAP	585
SMTP	587
IMAP/SSL	993
POP/SSL	995
open VPN	1194
PPTP	1723
SIP	5060
BitTorrent	6881
TOR	9001

D4: In order to satisfy requirement (R4), the following design was required. In order to make sure that our design would cover the limitation of our context, we saved the result from our experimental network in our local computer. We also kept the result we had collected from control networks in our servers at Victoria University of Wellington's open network. Not having to transmit web page data from the client to the server and instead manually transporting it on the laptop reduced bandwidth usage.

3.2 WCMT architecture

In order to satisfy the requirements, two components were proposed for content filtering detection and service blocking detection. In this section, we will present an overview of each of these components.

3.2.1 Content filtering detection

Figure 3.1 consists of three basic building blocks: The Oracle, the client and analysis engine. The Oracle is a program that will be run in the server which has my own implemented web server and Tor installed on that. The client is a program that will be run in the experimental network. The analysis engine is the program for automatically defining the censorship. The main components of the system and the process executed on each of the components have also been shown in Figure 3.1.

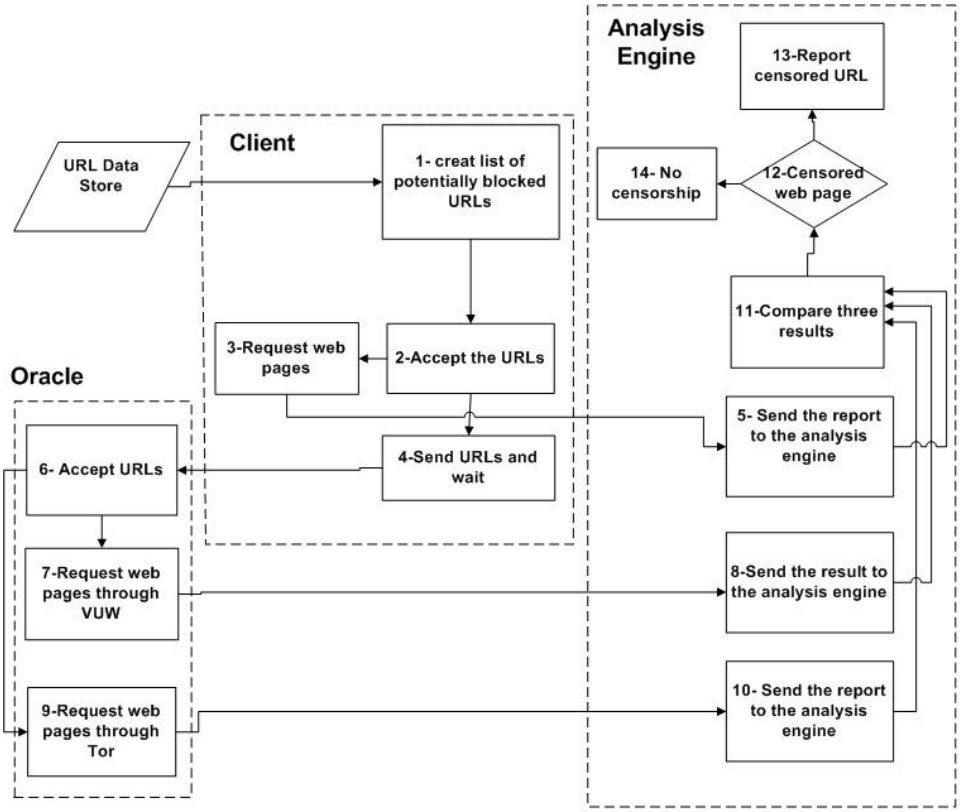

Figure 3.1
Overview of content filtering detection architecture

The Oracle acts as a proxy and will get the content from VUW and Tor

on behalf of the client. The Client is responsible to get the content from experimental network and send the URLs to the Oracle. The analysis engine is responsible to identify blocking based on the contents collected from Oracle and client.
1) A text file containing the URLs to check is created using an editor.
2) The client reads each URL from the text file and accepts the URLs.
3) The client will send the HTTP requests to fetch the URLs from the experimental environment. The result from the experimental network will be saved on the client.
4) Client will also send the URLs to the oracle for further investigation.
5) The client also sends whatever is returned from experimental network to analysis engine.
6) The Oracle will accept the URLs that are sent from the client.
7) The Oracle will fetch the result from the web server's network.
8) The Oracle will save whatever is returned from VUW on the Oracle server and will also send whatever is returned from VUW to the analysis engine.
9) The Oracle will fetch the result from the Tor networks.
10) The Oracle will save whatever is returned from Tor on the Oracle server and will also send the result to the analysis engine.
11) Whatever is returned from the three networks will be compared with each other at analysis engine.
12) Analysis engine will decide whether there is any censorship or not.
13) If there is any difference, the alarm of censorship will be set.
14) If there are not any differences, the alarm of no censorship will be set.

Analysis engine

Analysis engine is responsible to identify censorship. The methodology for determining censorship is based on comparing the results from experimental, VUW, and Tor network in terms of overall length, header length,

body length, header content, body content, number and size of images.

The engine tries to identify censorship by comparing HTML codes from local, web server, and Tor network. At the start, the overall length of the two text files are compared with each other. Second, the header length of the two results are compared with each other. Third, the body length of the two results are compared with each other. Fourth, the content of two headers are compared with each other. Fifth, the body content of the two results are compared with each other. Six, the two results in terms of size and number of images on those pages are compared with each other. Differences in each of the above tests are considered as evidence of censorship.

We did not use MD5 to compare the source HTML code received from the URLs because we also considered the differences in the contents in our analysis in the manual checking section in order to mitigate false positives. Therefore, we needed to compare the two text files line by line and saved the differences in the third file. To compare the images we used MD5 since getting the MD5 of the images clarified whether the pictures were identical or not.

3.2.2 Service blocking detection

Figure 3.2 consists of two basic building blocks; the server and a client. The server was on listening mode on the desired port numbers in Victoria University of Wellington' servers. The client was a program that was run in the experimental network and tried to connect to the server on those special ports. The main components of the system and the process executed on the components were also shown in Figure 3.2.

As depicted in this flow diagram, the service blocking component worked through the following steps.
1) A text file containing the ports to check is created using an editor.
2) The client will read the port numbers and accept it.

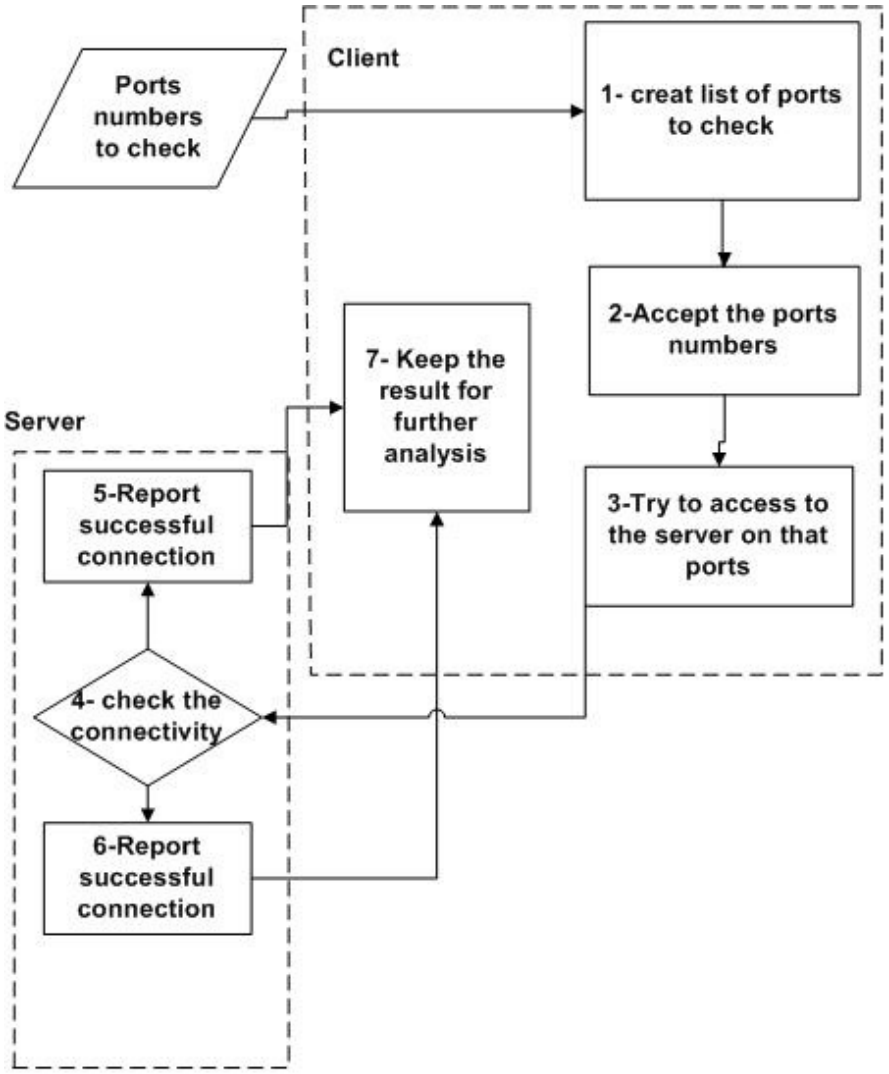

Figure 3.2
Overview of service blocking detection architecture

3) The client tries to connect to the server on those special port numbers.
4) The server is on listening mode in the control environment. When the request for the connection is received by the server on those special ports, the decision whether the connection is successful or not is made.
5,6) The result of connection reports whether or not it is a successful con-

nection.
7) The results are sent back to the client side and saved for further analysis.

3.2.3 Content filtering detection procedure

In this section, the procedure of content filtering detection will be described through the following four algorithms.

Algorithm 1 shows that a client will first read the URLs from the text file. When client gets the URL, its initial task is to try to connect to that URL at port 80 in order to detect if URL blocking takes place. Upon successfully connecting to the URL, the client tries to fetch the URL from the experimental networks. The client will save what it can fetch in text format in the client computer. In the next step, the client will forward the URL to the Oracle for further processing.

Data: Random URLs
while *not at end of the URLs* **do**
    Reading URL from text file;
    Fetching the URL from experimental environment;
    Saving the first result (ORG) to Text file 1;
    Sending URL to the Oracle;
end

Algorithm 1: Client operation

As shown in algorithm 2, the URL which is sent from a client will be received by the Oracle. The Oracle will try to fetch the URL from the web server as well as the Tor network. The Oracle will save the results which have collected from the web server and Tor networks for further analysis. All these three files will be saved in text format.

Finally in algorithm 3, analysis engine tries to identify censorship. Since, HTML codes from local, web server, and Tor network are saved in the computer, our tool tries to compare the HTML codes as well as the pic-

Data: Wait for request from the client
while *receiving URLs from client* **do**
    Fetching the URL from VUW network;
    Saving the result from oracle (VUW);
    Sending URL to the Tor;
    Fetching the URL from Tor;
    Saving the third result (Tor);
end

Algorithm 2: Oracle operation

tures of the same URL from each organization. It is the binary content of the pictures rather than HTML codes being compared here.The results are compared based on algorithm 4.

In algorithm 4, the hierarchy tests will be applied on the text file to determine the level of censorship. In this algorithm, the overall length, the header length of the two results, the body length of the two results, the content of the two headers, the body content of the two results, the two results in terms of size of the images on those pages are in that order compared with each other as proof of censorship.

In algorithm 4, we did not use MD5 to compare the source HTML code collected from the URLs because we also considered the differences in our manual analysis. Hence, we needed to compare two text files line by line in order to save the differences in the third file. It would help us in the last stage where we wanted to do manual checking in order to be sure that we were not giving false positives.

In this algorithm H1 and H2 referred to headers of the two files. B1 and B2 referred to bodies of the two files. P1 and P2 referred to two source HTML code collected from each URL. Img1 and img2 referred to the images downloaded from the URL which needed to be compared with each other.

3.2.4 Service blocking detection procedure

In this section, the procedure of detecting service blocking has been described through the following algorithms.

Algorithm 5 shows that the server will start its job by going through the list of ports it should listen on. Then the server will listen on those ports. When the request for the connection comes from a clients, a report will go to the client to make sure that the connection to that specific port is successful.

Algorithm 6 shows that client will read the port numbers from a file to test and then tries to connect to the server at those specific ports. When the connection is terminated, the report will be sent back to the client by the server.

3.3 Implementation

As mentioned before, the goal of this project was to make transparency in terms of content and service blockings for public users. The aim was to make this tool available for public users. Therefore, an open source programing language such as python was used. Python has many standard libraries for simple HTTP/Web Services and it is also easy to use other libraries from other programming languages.

In this project different libraries have been used for getting the source code, connecting to Tor, and analyzing the files. Since we had difficulties connecting to Tor and sending our request to Tor, we had to use a middle library to handle the request between us from experimental network and Tor network. Therefore, we used Polipo [45].

In terms of running the program, WCMT would be run through command line. Tor had to be started before starting WCMT. After building a circuit with Tor network, it was necessary to run Polipo as a middle library between experimental network and WCMT. After that WCMT were run.

WCMT started its job by reading the URLs through input file. Input file is a text file completed by the experimenter with the URLs we desired to check with WCMT in our experimental network.

3.4 Summary

In the previous chapter, issues with OONI were identified and based on OONI issues, the requirements for a new tool were clarified. In this chapter, to satisfy these requirements we presented the design and implementation of our new tool, WCMT. We also presented the architecture of the tool and our algorithms.

In the next chapter, we discuss how the preliminary test is done on our WCMT to make sure that our tool is ready to be used in organizations.

Data: Three txt files from organization, web server, and Tor
Result: Defining censorship level
Comparing result (ORG), result (VUW), and result (TOR);
while *not at end of the result* **do**
  if $Tor == Local$ **then**
    Return " No censorship ";
  else
    Return org censorship or country censorship **if**
    $Tor <> Localandtor == vuw$ **then**
      Return " Org censorship";
    else
      Return "Country censorship";
    end
  end
  if $Tor == VUW$ **then**
    Return "No censorship at country";
  else
    Return "Country censorship";
  end
  if $VUW = Local$ **then**
    Return "No censorship at org but maybe at country level so check the following equation" ;
    if $vuw == tor$ **then**
      Return "No censorship"
    else
      Return " Censorship at country level"
    end
  else
  end
  Return "Org censorship"
end

Algorithm 3: Analysis algorithm

Data: Http responses: H1, H2, B1, B2, P1, P2, Img1, Img2
Result: Having censorship or not
while *not at end of the result* **do**
    Reading two text files;
    Comparing the overall length of the two texts;
    if $length(p1) <> length(p2)$ **then**
        Return "Censorship due to having different lengths";
    end
    Comparing body length of two results;
    if $length(b1) <> length(b2)$ **then**
        Return "Censorship due to having different length in bodies";
    end
    Comparing header length of two results;
    if $length(h1) <> length(h2)$ **then**
        Return "Censorship due to having different lengths in headers";
    end
    Comparing body content of two results;
    if $content(b1) <> content(b2)$ **then**
        Return "Censorship due to having different contents in bodies";
    end
    Comparing header content of two results;
    if $content(h1) <> content(h2)$ **then**
        Return "Censorship due to having different contents in headers";
    end
end
while *not at end of the image result* **do**
    Compare images of the results;
    if $MD5(img1) <> MD5(img2)$ **then**
        Return " Censorship due to having images with different sizes ";
    end
end

Algorithm 4: Content blocking algorithm

```
while There are more ports to listen on do
    Reading the port number;
    Listening on that port number;
    while Receiving request from clients on special port do
        if Connection was successful then
            Return Connection was successful;
        else
            Return Connection was failed ;
        end
    end
end
```

Algorithm 5: Server operation in service blocking detection

```
while not at end of the port numbers do
    Reading port numbers from text file;
    Reading the server IP address;
    Trying to connect to the server at that specific port;
end
```

Algorithm 6: Client operation in service blocking detection

Chapter 4

Preliminary testing

The experiments were carried out by visiting free wifi access points where either time or using data might be limited. For example, CBDfree WiFi only allows connections for up to 30 minutes before having to log in again or Starbucks has a limit on the amount of data that can be used in one session. It is important to understand time performance of the tool to make it possible to plan the visit.

To overcome the limitation, the tool needs to be relatively fast in order to let us test more URLs. We had to minimize the amount of fetching time. Thus, we did it in the design stage considering our constraints. To identify how fast our tool is to overcome the limitation of the bandwidth and time, it is essential that we measure the fetching and analysis time through our experimental design.

This chapter describes the methods used for the experiments. Then it identifies the relevant hazards of our experimental design. Following that it discusses the significance of the results and the implications for the experiment that determine the total time taken for our experiment.

4.1 Experimental design

This section outlines the experimental setup of the performance results. This setup includes the description of the operating environment, procedure, and the test subjects.

4.1.1 Operating environment

The client machine through which the experiment ran was a HP Pavilion with Ubuntu version 12.10 operating system and 6 MB RAM and processor Intel(R) Core(TM) i7. The tool was installed on the HP machine. The tool needed a Python version 2.6 which was installed on the machine. The experiment was through domestic connection, the wireless network provided by Vodafone.

4.1.2 Procedure

The test was repeated for 30 times to control the variability in latency of the network. We also ran the test 30 times to calculate the average time taken to run 2400 URLs. We cleaned the cache in Internet Explorer each time before starting to run the test. We calculated the time taken to run the test in each round to have an average time for our formula. The experimenter carried out the following steps:
1) Experimenter recorded the details of time and day.
2) Experimenter started recording the time.
3) Experimenter started to run the test.
4) Experimenter saved the results to the external hard.
5) Experimenter wrote down the time when the test finished.

4.1.3 Subjects

URLs were subject to the system. The URLs sourced from five different categories from four country domains (see Section 4.2.1 for details).

4.2 Identifying hazards

There were risks to the validity of the experimental design. These risks could affect the intent of the experimental setting called internal validity as well as risks affecting beyond the experimental settings called external validity. An overview of the identified hazards has been given below.

4.2.1 URLs

URLs were inputs to the system. To decrease the risk of the URL hazards we created random URLs from larger population. It is worth mentioning that if we controlled the URLs, the results could be totally different. To mitigate the hazards, we decided to use Win Web Crawler[64] for randomly finding the URLs. This application has the ability to find the URLs from different search engines from different countries. The topics and domain were the other hazards which affected the experimental design as illustrated in the next section.

4.2.2 Topics

One important hazard to our system was the topic to be used to generate the URLs. We did not like to be specific by choosing special words to be used to generate the URLs. We have chosen our words based on the dmoz categorization [39, 47]. Five different categories of words were selected to generate the URLS, namely adults, music, news, user content, and warez.
Adults: sites related to adult entertainment and pornography.
Music: sites related to famous singers and artists.

News: sites related to different kinds of news such as sport, business, technology, and entertainment.
User content: sites related to forums and blogs.
Warez : sites related to hackers and their cyber attacks.

All the words used for submitting the queries were in English. At least the first 300 URLs returned by Win Web Crawler were used to build the final list of URLs for testing. The websites were double checked to ensure that the websites had the information regarding the specified category.

A summary on the number of URLs from each categorization has been shown in Table 4.1. Approximately 2400 URLs were sourced.

Source	Inspected URLs
Adult	700
Music	300
News	400
User Content	600
Warez	400

Table 4.1
Input URLs per topic

4.2.3 Domain

Domain was identified as another hazard to our system. We did not like to limit our experiment to the URLs from a special country such as New Zealand. Because free WiFi was mostly used by tourists, it was important even more to know where the tourists came from. Hence we considered domains from countries whose people mostly visit New Zealand. Most of the tourists came from Australia, United States, and China [27]. We

classified the URLs to be used in our experiment based on their domain name in Table 4.2.

Domain	Inspected URLs
New Zealand	900
Australia	700
United States	400
China	400

Table 4.2
Input URLs per domain

4.2.4 Stimuli

In the process of measuring the censorship, there were stimuli. This was the act of making a request to get the results. This process was a request for a URL through organization, VUW, and Tor networks. Based on the results collected from three networks, the decision was made by analysis engine. Analysis engine compared the results based on overall length, body length, header length, header content and body content. Moreover, the images from the results were compared with each other to get to final results.

A question raised here was whether the tool we had implemented performed what this tool was designed to do. This was the other hazard to our experimental design. Functional testing and monitoring functionality were used during the process to mitigate the hazard.

4.3 Total time taken

Formula 4.1 shows the total time taken for the experiment. The time to our experiment $\vec{T_i}(Total)$ was calculated based on the time for collecting the data $\vec{T_i}(C)$ and the time to analyse the data $\vec{T_i}(A)$. As depicted in formula 4.2, the time to collect $\vec{T_i}(Collection)$ extended to fetching time from three networks, local $\vec{T_i}(EN)$, VUW $\vec{T_i}(WN)$, and Tor $\vec{T_i}(TN)$. In Formula 4.3, 5880 was the average fetching time from local network, 6480 was the average fetching time from University network, and 7440 was the average fetching time from Tor networks.

Since the process of fetching from local, webserver, and Tor network takes place in parallel, the maximum of these three fetching times were considered as total fetching time. When the fetching finished the analysis engine compared the results based on the analysis algorithm explained in design and implementation chapter.

$$\vec{T_i}(Total) = \vec{T_i}(C) + \vec{T_i}(A); \tag{4.1}$$

$$\vec{T_i}(Collection) = Max(\vec{T_i}(EN), \vec{T_i}(TN); \vec{T_i}(WN)) \tag{4.2}$$

$$\vec{T_i}(Total) = Max(5880, 6480, 7440) + (30) \tag{4.3}$$

$$\vec{T_i}(Total) = (7440 + 30) = 7470Second \tag{4.4}$$

The results from 4.4 showed that the tool was fast enough to check 2400 URLs in the organization in each visit.

4.4 Summary

In this chapter we presented our performance evaluation of our developed tool by running our preliminary test. Our goal was to overcome the limi-

tation of our experimental networks. Our experimental networks had the limitation in using data and time. Based on our evaluation, our design successfully satisfied the limitation of our experimental design.

In the next chapter, we use our implemented tool to identify the number of content and service blockings in different organizations.

Chapter 5

Quantitative measurement of blocking

This chapter describes the methods used for identifying the number of content and service blockings in 8 different organizations in Wellington, New Zealand. This section explains how quantitative experiment was conducted using the implemented tool, describes the organizations chosen for the test, and lists the websites and services identified as blocked. This chapter also explains how the environment was set up to run the test. It further discusses experimental design, and hazards to experimental design. Then, it presents the results and discusses their significance and implications. In the next chapter the qualitative experiment determines why contents and services are blocked by providers of wireless internet access in Wellington.

5.1 Experimental design

This section outlines the experimental setup of the content and service blocking tests. This setup includes the description of the operating environment, procedure, and the test subjects.

5.1.1 Operating environment

This section outlines the environment in which the tests were run. The experiment required the use of the following computers: the client computer, the Oracle, the server, and the web censorship monitoring tool. The client machine was an HP Pavilion with Ubuntu operating system version 12.04 and 6 GB RAM memory and processor Intel(R) Core(TM) i7. The client side of the tool was installed on the HP machine. The tool needed a Python version 2.6 which was installed on the machine. The server side of the tool was run on one of the servers in open network at Victoria University of Wellington. Organizations were chosen based on whether they offered free Internet or we had access to their free WiFi.

5.1.2 Procedure

The experimenter carried out the following steps for each organization to find the blockings in URLs and services:

1. Experimenter recorded the details of time and day.
2. Experimenter recorded the details of the place that the connection went through.
3. Experimenter started to run the test.
4. Experimenter saved the results to the external hard drive.

Censorship is a sensitive subject, and the accuracy of the results regarding the number of content and service blocking was very important. Therefore, it was essential that we find out about the hazards on experimental design. There were hazards on each of the components of our Architectures 3.1 and 3.2. These hazards were risks to the validity of the experimental design [5]. Therefore, we had to consider these hazards and

try to mitigate them. Time, location, URLs, and services were important hazards which had an effect on measurement. A brief description of these hazards and how to mitigate them are given below.

5.1.3 Time

Time was the first hazard we identified in our experimental design. The web is dynamic, and it is possible that a specific web site is filtered at a specific time and not all the time. It is also possible that the designer sets a threshold and if the usage of the Internet reaches that threshold, the URL will be blocked. Administrators, by setting a threshold, actually set an exceed point. If users exceed that point, a given effect or result happens. To mitigate this kind of hazard, we ran our experiment in three rounds.

Initially we decided to compare the results at different times which was a time hazard to our experiment. Hence we changed the tool by implementing our web server so that we could fetch the content at the same time. This way, we could compare the results at the same time to mitigate the time hazard.

5.1.4 Location

Location in this study referred to the place where free Internet was offered or where we had access to it. Each location had its own terms and conditions for using the Internet which was another hazard to our system. For example, these places were different in terms of the amount of time and data that we could use. To mitigate the location hazard, we deliberately chose organizations that gave us reasonable amount of time for using their free WiFi. For example, using free Internet from airport buses were limited to the time that the experimenter was staying on the bus. Therefore, the number of URLs which we could test were limited. This was a hazard to our experimental design. Therefore, to mitigate the location hazard, we omitted these sorts of places from our experiment.

It was very important to get an idea about the number of URLs which we could test in each of the organizations. If we assume, for example, that each page takes around 2.45 seconds [8] to load and further assume that we can stay for 1 hour in each place, we can probably check 1470 URLs each time. Therefore, in order to mitigate the location hazard, we tried to run the experiment a couple of times.

5.1.5 URLs

URLs were inputs to the system and were considered as another hazard to our system. In order to find the data set of the URLs for running the test, we looked at other experiments. For example, CensMon used top 10 URLs returned by Google. Heredict web used reported URLs, and ONI published a list. ONI had two lists, namely global and local lists. The global list was constant for each country. The local list was different for each country. The local lists were designed individually for each country by regional experts.

In order to run the experiment with URLs, we were interested to find out if there was a data set for URLs and ports. We started our job with HerdictWeb. HerdictWeb was a web censorship monitoring tool which was available in different languages such as English, Persian, Arabic, and Russian. In HerdictWeb's methodology, whenever a user cannot access a website they will report it to the HerdictWeb server. This system relies on users' feedback which is not safe. Users cannot distinguish between different non-reachability reasons. For example, we tried to access the list of the URLs which was reported by users in New Zealand. Most websites were reachable. It showed that this data set was not reliable for this experiment.

We tried to make a data set based upon content blocked in other countries that had the same level of censorship. The closest country, Australia, was a good start. Wikileaks [33] published the list of the blocked web sites by the Australian government. This list contained 2395 web pages. The

list of blocked web sites was not limited to child pornography, extreme violence and even bestiality. On the list it showed that even the website of Queensland dentist and dog-boarding kennel were blocked. We did not use the data set because it was unclear of the legal ramifications and besides we were not testing that we can access a clearly reasonable thing to ban like a child pornography site. Also using Australian censored URLs published by wikileaks would lead to a comparison study between these two countries. It would be a hazard to our system. To decrease the risk of the URL hazards, we created random URLs using Win Web Crawler [65] because if we wanted to control the URLs, the results could be totally different. Then, we had to find dead URLs and omit them to mitigate the hazards. The categorization and domain selection were other hazards to our measurement studies.

One important hazard to our system was the topic we used to generate the URLs. We did not like to be so specific by choosing special words to be used to generate the URLs. Hence, we chose our words based on the categorization in [39]. Fifteen different categories of words were selected to generate the URLs such as racism, drugs, malicious sources, shopping, adult, proxy, Peer-to-Peer (P2P), jokes, dating, gambling, alcohol, hate speech, hacking, weapons, and sex education.

Domain was identified as another hazard to our system. We did not like to limit our experiment to the URLs from any especial country such as New Zealand. Hence, we considered domains of visitors who most frequently visit New Zealand.

5.1.6 Services

Services were the other hazard to our experimental design. To mitigate this hazard we chose the following list of services as they were very common for people to use.

FTP servers (port 21), SSH servers (port 22), SMTP servers (port 25),

DNS servers (port 53), HTTP servers (port 80), POP3 servers (port 110), MSSQL servers (port 1434), remote RPC servers (port 135), NetBIOS servers (port 139), IMAP servers (port 143), open VPN (port 1194), BitTorrent (port number 6881 and 4711), TOR (port 9001), SNMP servers (port 161), HTTPS servers (port 443), SMB servers (port 445), SMTP/SSL servers (port 465), IMAP servers (port 585), SMTP servers (port 587), IMAP/SSL servers (port 993), POP/SSL servers (port 995), SIP servers (port 5060), and BitTorrent servers (port 6881).

5.2 Experiment

We ran the Web Censorship Monitoring Tool (WCMT) publicly from July 2013 until September 2013. We ran WCMT to collect data from McDonald's, Starbucks Cafe, CBDfree WiFi, Te Papa Museum, Victoria University of Wellington, Vodafone, National Library of New Zealand, and Wellington Airport.

The first phase of our collected data included fetching URLs from different organizations. The data consisted of contents gathered from experimental and control networks.

The data was collected using 1075 URLs per site. These data were collected in three rounds of visits to the sites, each time for one hour. Contrary to what we thought, there were a lot of differences between the contents of the pages fetched from experimental network and control networks. Even for very ordinary websites we received false positives. Therefore, in the second tier of our analysis we used manual checking to identify what content was really filtered. Using manual checking was mandatory since, in the first tier of analysis, we faced many false positives as discussed in the following paragraphs. Getting different contents could have different reasons which are explained in the following paragraphs.

1. There were websites whose content could not be fetched through either Tor or experimental networks. It is possible that they could

identify humans from applications. These sorts of websites either did not respond or redirected the content which led to having different contents between control and experimental networks.

2. There were other websites with dynamic content. In spite of the fact that we sent the request for the content at the same time, the results were different. News websites were examples of these sorts of URLs. For example, CNN fetched through Tor brought totally different content from experimental and VUW's Networks.

3. Some other websites ran scripts that could lead to different contents being displayed.

4. There were websites which had different advertisements and image names based on the localization of the data which led to having differences in the contents.

5. Another reason was that since Tor brought the content from other countries for which the content had been localized, it was different from the content fetched from experimental and VUW networks. For example, Google fetched from New Zealand had a different content compared with the one brought from Australia.

In order to overcome the problems we faced, we used different solutions. At first, we tried to identify a pattern in order to accelerate the automatic analysis of the contents for each ISP but we could not identify a filtering pattern for all ISPs. Therefore, for each ISP we added different techniques of manual checking into our analysis methodology.

In the second tier of manual checking, we tried to find the contents with zero length and tried to find the reasons for that. This tier of analysis led to a list of suspicious URLs to censorship. Further analyses were done by fetching them again to find out if the time hazard had an effect, if the websites were dead, or if the URLs were blocked.

Next, we went through the content to find some special words to determine blocking. We wrote a program to check whether there were special contents such as "You don't have permission to access this server, access denied, page unavailable," etc., which could help us spot censorship. This solution was useful for places where blocking was transparent, and showed their blocked webpage, but not for all ISPs. This solution also gave us lots of false positives for places which did not have transparent blocking page.

For different ISPs it was not clear to us which way they had chosen to show censorship to their users. There were some ISPs which used a combination of methods. For example, Te Papa used a combination of transparent censorship messages, redirecting, and different kinds of error messages at the time of censorship in different situations. Therefore, some of the blocking could be found through following the pattern and the others had to be recognized through manual checking.

Manual checking was time-consuming as analysis engine generated the differences between contents fetched from experimental and control networks and saved them on a third file. It was 1075 multiplied by 2 (1075*2) files which were not possible to go through all for each site in our limited time. Our initial aim was to go through all the data but it was not feasible considering time. Therefore, we went throughas as much data as we could analyze to see if the changes were benign or not.

Our methodology was successful in determining censorship and it reduced the number of false positives in presenting the number of blocked URLs. The limitation of this methodology was that it was slow due to relying on real users for collecting and analyzing the data.

Port checking was the second phase in our methodology. With this test we could identify the reachability of different services. Port blocking is a policy control for site administrators. Port blocking means closing special applications and it is assigned TCP or UDP ports. Since most of the applications use well-known ports, port blocking is an effective way of im-

plementing censorship. Port blocking is a simple and cheap way of controlling traffic on the network. Unfortunately, many ISPs used these techniques which raised debates over network neutrality. Authorities had different motivations for port blocking such as stopping propagating worms, providing security for specific applications, or economic reasons for peer to peer applications.

In this chapter, we quantified the prevalence of port blocking in different organizations. We did not like to emphasize whether port blocking was legitimate or justifiable. We, however, just intended to make it transparent. In our methodology, we checked whether the ISP allowed us to connect through those special ports. If, for example, client C with IP address (IP1) from network N could connect to port P of our server, we would consider port P as open.

5.3 Data collection

In this section, an overview on the data we collected from 8 organizations will be given.

5.4 Blocked URLs

Different organizations have blocked different contents. In this section, the results of content blocking for different organizations are presented.

5.4.1 Content blocking in organization 1

At organization 1 we tested the free WiFi with our first round of URLs. In our first round of our data collection, out of 180 URLs, 114 URLs were identified as being censored. It was interesting how organization 1 treated the customers by providing filternet as free Internet. URLs from drug, racism, malicious sources, adult shopping, pornography, proxy, dating,

and anonymizer categories were blocked. Surprisingly, alcohol websites were open. Since at organization 1, there was a transparent blocked page shown at time of censorship, we investigated the fetched content for that specific pattern and prepared a list of blocked URLs which is depicted in Table A.1 in Appendix A.

Organization 1 was extreme in terms of content blocking. After using Internet for half an hour, I received an error message stating that I used excessive amount of data. In the message, it is mentioned that I can use Internet again in 30 minutes. Exceeding that 30 minutes, I tried a couple of times, but I was unable to use their free WiFi. Next, I tried a different branch of organization 1 center. At this location, URLs from different categories were checked in that limited time.

It was interesting that all gay websites were blocked but the gay websites which were not based in NZ were not filtered such as `http://www.gay.ru/`. We checked websites from lesbian websites but they were open. It was not clear to us why they did not treat similar category of URLs the same. Even dead websites such as `http://www.casinogamblingexposed.com/` were blocked.

Most hacking websites we checked were blocked but not `http://www.hackinthebox.org/`. We were eager to understand how they categorized the URLs for blocking and if they updated their category regularly?

5.4.2 Content blocking in organization 2

We tried to access 1075 URLs through organization 2 network. Almost all of the websites were open except a couple of URLs documented in Appendix A. Our request to the URLs we identified, was redirected or skipped to the desired content of the site administrator.

Organization 2 was considered a paid service. Therefore, it was expected to provide more freedom to the customers. The results also showed

that only two pornography websites were blocked. In both cases, we were redirected.

5.4.3 Content blocking at organization 3

Table A.2 in Appendix A showed the results of content blocking test at Vorganization 3's wireless network. As it is shown in the result table, the organization 3 extensively blocks students from accessing substantial number of web pages.

Organization 3 uses a template with different content blocking categorization at the time of blocking. Therefore, it was easier to identify the blocking through processing the content. organization 3 is more transparent to their users compared to other places which redirect the page to other websites or use network errors or lengthy timeout.

As it is shown in Table A.2 in Appendix A, the websites which had contents related to proxy, malicious websites, adults, file sharing, anonymizer websites, racism, pornography, gambling, religious, illegal drug, and dating were blocked. It was interesting to know how they used different categories for implementing censorship.

Since a list of blocked websites was not available and was not transparent, implementing censorship led to categorizing the URLs incorrectly. For example, `http://www.jihadonline.org/` was blocked at the organization 3 in the "Search Engines/Portals; Malicious Sources; Malicious Outbound Data/Botnets," category which is not right. This is a website for Muslims.

Websites such as `http://ww41.ourworldkids.info/` was about children and their toys, which was mistakenly filtered at the organization 3 as a "Pornography; Extreme," categorization.

URL `http://Torproject.org` was blocked as well in the "Proxy Avoidance" category. Like many other web pages, this web page which could provide students with useful information was blocked. Some of

the websites are informative. So why does the organization 3 not allow students to have access to the information related to Tor and just prevent them from using it, if this is objectionable for them.

URL: `http://www.megago.com/l/` was blocked at the organization 3 in the category of "Search Engines/Portals; Malicious Outbound Data/Botnets". This URL was an informative website; they prevented users from getting information.

URL: `http://dirtyjokesinc.com/` was blocked at the organization 3 but the URL was not actually alive.

URL: `http://www.dirtyjokesinc.com/` was dead but it was blocked at organization 3 in "Adult/Mature Content; Newsgroups/Forums; Humor/Jokes" category.

URL: `http://3apa3a.tomsk.tw/c/cfg.bin` was dead but was blocked in "Malicious Sources; Malicious Outbound Data/Botnets" category.

URL: `http://sxetc.org/` was dead but categorized at "Adult/Mature Content; Abortion; Sex Education".

It was interesting to know why websites from the same category were not all blocked. For example, URL: `http://friendfinder.com/` was blocked but URL: `http://www.findsomeone.co.nz/` was not blocked.

It was interesting to know that organization 3 had so many sub categories as we had just categorized the websites to fifteen different topics.

5.4.4 Content blocking in organization 4

1075 URLs were fetched. Among the URLs we tested, there were URLs which were spotted as blocked by either receiving a transparent blocking page message, an error message, not seeing the content without any error message, or being redirected. The list of blocking in organization 4 is presented in Table A.3 in Appendix A.

We identified the list of blocked content in organization 4 by trying to find the blocking pattern. Other than that, there were also websites which

were blocked but the consistent blocked page was not shown to the users. For example, the response to the request for accessing h33tunblock.info which is a "Peer-to-Peer (P2P)" was empty.

Surprisingly, in the second round of data collection the h33tunblock.info was shown to the experimenter. It is clear that their polices change over time. It is also clear that they do not want to be transparent to their users to show blocking page at the time of blocking.

The other way used by organization 4 at the time of censorship was showing "504 gateway time out" error message instead of showing transparent block page. "504 gateway time out" error message was shown to us when we tried to access URLs such as http://youngtop.info/ and http://www.torrentbytes.net/.

In our second round of data collection http://www.torrentbytes.net/ was shown to the user but http://youngtop.info/ was redirected.

There were situations in organization 4 wireless network where the request was redirected to special pages. Requesting URLs such as http://vi5search.com/, http://xtra.co.nz, and http://pretty-pretty.info/ led to redirecting to the administrator's desired websites.

Requesting all gay websites was responded by transparent blocked message but request to http://www.gaynz.com/MYSA/mysa_redir.php?a_id=202&t=1&c=0 as an example of gay website was redirected to the desired websites of authorities.

All the gay and lesbian websites we checked were blocked but interestingly URL: http://www.gay.ru/ was not blocked as the source of this web site was not from New Zealand.

There were some websites the content of which we could not see but the error message was not shown either. URLs such as http://www.marijuanareform.org/ and http://www.hackinthebox.org/ were examples of that.

5.4.5 Content blocking in organization 5

1075 URLs were fetched. Almost all the web pages were open in organization 5. Alcohol, drugs, gambling, gay and lesbian, hacking, pornography, hate speech, and proxy websites were open. They did not use any transparent way of showing blocked webpages to their users. There were situations in which the request were redirected such as `http://vi5search.com`, `http://xtra.co.nz` , `http://www.feminista.com/`, and `http://youngtop.info/`.

5.4.6 Content blocking in organization 6

1075 URLs were fetched in three rounds of data collection. Most of the websites were open. We could not find any transparent blocked pages there. Therefore, we tried to find the content with the words clarified in section 5.2 which led to lots of false positives.

The problem with organization 6 was that the request for objectionable content was redirected but to a variety of websites and not to a specific website. If the request were redirect to the specific website, it would be easier to find out the blockings. URLs from alcohol, hackings, VPN, proxy, pornography, gay and lesbian, BitTorrent, gambling, drugs, and malicious sources categorizations were open.

In organization 6 WiFi, a request to certain websites was redirected to the administrator's desired contents. The sample of blocked URLs which led to redirections is presented in Appendix A. Interestingly, most of the pornography websites were open except for a few of them which were redirected.

`http://kickassunblock.info/search/.kickasstorrents/` was not redirected but the error message mentioning that "gatwway error 504" was shown to the user. URLs such as `http://www.jokes.com/` and `http://www.the-jokes.com/` which were from "Adult/Mature Content; Humor/Jokes" category could not be seen.

http://saudieng.net/,http://www.igc.apc.org/Womensnet/dworkin/,http://saudieng.net/,Http://www.jokes.com/,and http://www.the-jokes.com/ were not blocked in the second round of data collection.

The response to the request for www.prettynudists.com was the error message "This Domain is PARKED". Upon my second try on a different day of data collection, the URL could be seen. It showed that their policy changed over time.

5.4.7 Content blocking at organization 7

Similar to other places, 1075 URLs were tested using organization 7 free WiFi. Most of the URLs except for a couple of them were accessible. The websites from gay, pornography, jokes, hacking, shopping underwear, humor, drugs, malware, gambling, drugs, alcohol, proxies, nationalist and racism categories were open. Organization 7 WiFi used redirection for the requests to their blocked lists. The redirection was to random URLs. This prevented us from easily finding blocked URLs. Nine URLs from pornography and peer to peer were redirected to the desired websites of the administrators. The list of the URLs which we requested, led to redirection is shown in Appendix A.

Political website such as http://hizbollah.tv/ was redirected. Some of the websites could not be seen using organization 7 free WiFi. It was sometimes difficult to tell whether a site was blocked or simply inaccessible for technical reasons. It will be helpful if we can ask in the interview if it is due to excessive usage or not.

http://teenpregnancy.org/ could not be seen either, but without any error message. There were websites that were blocked but were not redirected to other websites either. They were just blocked. The other example of not showing the content and not even redirecting were requests

to URLs such as Peer to Peer. The list of Peer to peer URLs we could not open at organization 7 is illustrated in Appendix A.

URL http://www.yahooka.com/ from drug category could not be seen there. URL http://friendfinder.com/ from dating category could not be seen either.

5.4.8 Content blocking at organization 8

1075 URLs were checked at the organization 8. All the racist, drug, malicious, adult, pornography, gambling, adults shopping, proxy, P2P, joke, dating, and alcohol websites were open except 14 URLs.

The request to access URL: http://3apa3a.tomsk.tw/c/cfg.bin led to "page not found".

There were requests to some URLs which led to redirection such as: http://sexual.vipzax.com/,http://risk.vipzax.com/,http://nude.vipzax.com/, and http://pretty-pretty.info/.

Request to http://thepiratebay.org/, http://xxx.com/ led to an error message depicted in Figoure **??** which is a transparent blocking page.

There were situations that the requests were redirected to other desired web pages such as a request for http://www.avizoon.com/.

The request for URL: http://tiptopteens.net/ had different responses. Upon the first try, it redirected the request to the desired web pages of that place's site administrator. Upon other trials, we faced errors such as "403 forbidden" and "Not Found".

The request for URL: http://teens-models.org/ was interesting as it showed an message saying, " You should update your media player". Upon the second try on the other day it was redirected.

At the time of requesting URLs such as http://ourworldkids.info/, http://youngtop.info/, and http://www.feminista.com/, the

requests were redirected to the Mozilla error message.

5.4.9 Summary of content blocking

In this section, we compared different organizations in terms of the number of blockings from each category. When we wanted to generate the URLs we used topic such as racism, drugs, malicious sources, shopping, adult, proxy, Peer-to-Peer (P2P), jokes, dating, gambling, alcohol, hate speech, hacking, weapons, and sex education.

It was interesting that the categorizations were different in those 8 organizations. All the organizations had blocking but they were different in terms of the number and the category of blockings. It was interesting that some of the URLs were blocked in the first round of data collection but not in the second round.

Out of these 1075 URLS in each of the organizations, the number of blocked websites from each organization is shown in Table 5.1.

Table 5.1
Comparison between content blocking among 8 different organizations

Category	**Org 1**	**Org 2**	**Org 3**	**Org 4**	**Org 5**	**Org 7**	**Org 6**	Org 8
Racism	0	0	1	0	0	0	0	0
Drugs	24	0	1	2	0	0	1	0
Malicious Sources	11	0	9	1	0	3	0	1
Shopping	16	0	3	6	0	0	0	0
Adult	41	2	37	32	5	10	5	8
Proxy	20	0	20	0	0	0	0	0
Peer-to-Peer (P2P)	8	0	8	0	0	0	7	0
Jokes	5	0	5	5	0	3	0	0
Dating	9	0	5	0	0	0	1	0
Gambling	30	0	1	1	0	0	0	0
Alcohol	1	0	1	3	0	0	0	0
Hate speech	15	0	0	0	0	0	0	0

Hacking	4	0	0	1	0	0	0	0
Weapons	6	0	0	0	0	0	0	0
Sex education	11	0	2	9	0	0	0	0

Table 5.2 shows the number of content blocking per organizations.

Table 5.3 shows the percentage of content blocking in each of the organizations. It is shown that different levels of blocking existed for each organizations. It was shown that different levels of blocking existed for organizations. For example, Organization 1 by 18.88% was extreme in terms of content blocking followed by Organization 3 with 9.12%, and Organization 4 with 5.58% blocking. These values were not absolute values as these websites were sampled for a specific time. It could change every now and then.

5.5 Blocked services

In the second phase of our data collection, we ran service blocking test in our 8 different experimental networks to identify the ports and services which were blocked. The result obtained from each of these organizations is given below.

5.5.1 Service blocking in Organization 1

We ran our service checking client through Organization 1 network. The result of service blocking test in Organization 1 is illustrated in Appendix B. Ports FTP, SSH, SMTP, DNS, POP3, RPC, NetBIOS, IMAP, SNMP, SMB, MTP/SSL were not open. However, ports HTTP, HTTPS, secure IMAP, servers, IMAP/SSL, POP/SSL , VPN, PPTP , SIP, and BitTorrent were open.

5.5.2 Service blocking in Organization 2

We tried the service reachability test with Organization 2 broadband network. All the services were open. The result of the test is illustrated in Appendix B.

Organization	Number of content Blocking per organization
Organization 1	203
Organization 2	3
Organization 3	98
Organization 4	60
Organization 5	4
Organization 6	16
Organization 7	14
Organization 8	9

Table 5.2
Snapshot view of the number of URL blocking

5.5.3 Service blocking at Organization 3

The result of service blocking test at Organization 3 network is displayed in Appendix B. As it is shown, most of the ports were open to use except ports SMTP (port 25), RPC (port 135), NetBIOS (port 139), and SMB (port 445).

5.5.4 Service blocking in Organization 4

The result of service blocking test in Organization 4 was illustrated in Appendix B. As it is shown most of the ports we tested were open to use except DNS (port 53), RPC (port 135), NetBIOS (port 139), and SNMP (port 161).

Organization	Percentage of content Blocking per organization
Organization 1	18.88%
Organization 2	0.279%
Organization 3	9.12%
Organization 4	5.58%
Organization 5	0.37%
Organization 6	1.49%
Organization 7	1.30%
Organization 8	0.84%

Table 5.3
Snapshot view of percentage of URL blocking in each organization

5.5.5 Services blocked in Organization 5

The details of service blocking test in Organization 5 is illustrated in Appendix B. All the ports were open even those that were mostly blocked in other organizations due to security reasons. It could be because they relied on their customers or because they did not consider security issues. These are questions expected to be answered in the interview sessions.

5.5.6 Services blocked in Organization 6

The details of service blocking test in Organization 6 is illustrated in Appendix B. In Organization 6 Wi-Fi all the services we tested were open except TCP access to remote SMTP servers port 25 which was prohibited. Because of this blocking, it was not possible to send email via SMTP. This sort of blocking was very common because this port could be used by

hackers for generating spam.

5.5.7 Services blocked at Organization 7

The result of port checking at Organization 7 is presented in Appendix B. As it is shown most of the ports were blocked. It was true that Airport did not block a lot of URLs but it blocked a lot of services. Ports 25, 53, 80, 110, 143, 443, and 993 were open and all others were closed.

5.5.8 Services blocked in Organization 8

All the services were open at Organization 8. The details of service blocking test in Organization 8 is illustrated in Appendix B.

5.5.9 Summary of service blocking

As mentioned before, investigating port blocking is important as it will affect network neutrality. We need to mention that blocking is a way of implementing censorship. Free Internet does not necessarily mean that access to everything should be free. Rather, we propose that access to Internet whether contents or services must be transparent to users. The incident of port blocking in 8 different organizations is shown in Table 5.4.

Table 5.4
Comparison between service blocking among 8 different organizations

Category	**Org 1**	**Org 2**	**Org 3**	**Org 4**	**Org 5**	**Org 7**	**Org 6**	Org 8
FTP	No	Yes	Yes	Yes	Yes	No	Yes	Yes
SSH	No	Yes	Yes	Yes	Yes	No	Yes	Yes
SMTP	No	Yes	No	Yes	Yes	Yes	No	Yes
DNS	No	Yes	No	No	Yes	Yes	Yes	Yes
HTTP	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
POP3	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
RPC	No	Yes	No	No	Yes	No	Yes	Yes
NetBIOS	No	Yes	No	No	Yes	No	Yes	Yes
IMAP	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SNMP	No	Yes	Yes	No	Yes	No	Yes	Yes
HTTPS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SMB	No	Yes	No	Yes	Yes	No	Yes	Yes
SMTP/SSL	No	Yes	Yes	Yes	Yes	No	Yes	Yes
IMAP	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
SMTP	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
IMAP/SSL	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Finished on next page

Table 5.4 (*finished*)

Category	**Org 1**	**Org 2**	**Org 3**	**Org 4**	**Org 5**	**Org 7**	**Org 6**	Org 8
POP/SSL	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Open VPN	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
PPTP	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
SIP	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
BitTorrent	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
TOR	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes

Table 5.5 shows the number of blocked services in each of the organizations. Wellington Airport with 15 port blockings was extreme followed by McDonald's with 12 blockings. Vodafone, the National Library of New Zealand, and Starbucks cafe did not block any services for their customers.

Table 5.6 shows the percentage of service blocking in each of our experimental networks.

Table 5.7 shows the number of organizations where each services is blocked. This table also shows how many organizations did not block a certain service. The most frequently blocked port were RPC servers (port 135) and NetBIOS servers (port 139) which were blocked by 4 organizations. Ports HTTPS servers (port 443), IMAP/SSL servers (port 993), and HTTP servers (port 80) were not blocked in any organizations.

Table 5.7
Number of organization blocking one special service

Category	**Number of organizations blocked this service**	**Number of organizations opened this service**
FTP	2	6
SSH	2	6
SMTP	3	5
DNS	3	5
HTTP	0	8
POP3	1	7
RPC	4	4
NetBIOS	4	4
IMAP	1	7
SNMP	3	5
	Finished on next page	

Table 5.7 (*finished*)

Services	**Number of organizations blocked this service**	**Number of organizations opened this service**
HTTPS	0	8
SMB	3	5
SMTP/SSL	2	6
IMAP	1	7
SMTP	1	7
IMAP/SSL	0	8
POP/SSL	1	7
open VPN	1	7
PPTP	1	7
SIP	1	7
BitTorrent	1	7
TOR	1	7

5.6 Findings

In all our experimental locations, we tried to fetch 1075 URLs, and our data showed that the practice of censorship was common among the Internet Service Providers (ISPs). It also showed how differently network administrators looked at their customers.

If we divide our experimental networks into commercial and educational groups, it would be expected that educational places such as organization 3, organization 4, and organization 8 had to have more blockings compared to commercial ones such as organization 7, organization 6, organization 5, organization 1, and organization 2. But even a consistent pattern was not observed among educational and commercial places. For

Organization	Number of Blocked services per organization	Number of service opened per organization
Organization 1	12	10
Organization 2	0	22
Organization 3	5	17
Organization 4	5	17
Organization 5	0	22
Organization 6	1	21
Organization 7	15	7
Organization 8	0	22

Table 5.5
Snapshot view of number of service blockings

example among educational places, organization 8 was an exception with fewer content blockings and not having service blockings. As for commercial places organization 1 was extreme in the number of content blocking and organization 7 extreme in the number of service blocking.

If we want to divide the experimental networks into two different groups, namely paid and free services, organization 2, organization 3, and organization 5 will be considered as paid ones. organization 7, organization 6, organization 1, organization 4, and organization 8 will be considered as free services. Even among paid-service providers such as organization 3, we did not have extreme freedom and also in free services not everything was blocked. As an example, neither organization 7, organization 6, organization 4, nor organization 8 were extreme in terms of content blocking except organization 1 which was extreme in terms of content blocking and organization 7 which was extreme in terms of service blocking.

Organization	percentage of service Blocking per organization
Organization 1	54.45%
Organization 2	0%
Organization 3	22.73%
Organization 4	22.73%
Organization 5	0%
Organization 6	4.54 %
Organization 7	68.18%
Organization 8	0%

Table 5.6
Snapshot view of percentage of service blockings in different organizations

As we saw in our two ways of categorization we did not see the same pattern in implementing blockings neither in commercial and educational nor in paid and free services. Each of these organizations had its own idea regarding blocking.

After studying blocking in these organizations, we identified problems regarding practice of blocking in these studied organizations. The list of these problems could be summarized as follows:
A: Inconsistency in classifying the blocked URLs.
B: Inconsistency in reporting blockings.
C: Changing policies over time.

The reason for this problem will be expected to be clarified in our next phase of our data collection which is interviewing authorities in charge of implementing blockings for experimental networks.

5.7 Summary

In this chapter, we gave an overview on our quantitative research and the methodology for running our experiment. We discussed our experimental design and the hazards on our experimental design. We also identified how to mitigate the hazards in order to increase the validity of our experiment.

In this Chapter, we also presented our results regarding content and service blocking in each of the organizations. We made a profile for each of these organizations and at the last stage we compared them based on the number of blockings in content and services.

In all the study areas, we tried to fetch 1075 URLs. Our data showed the practice of censorship was common among the Internet Service Providers. It also showed how differently network administrators looked at their customers, how they changed their policies over time and how they were different in classifying and reporting blockings.

In the next Chapter we present our methodology for qualitative re-

search and how we use this methodology for our interviews with IT professionals.

Chapter 6

Qualitative investigation of policies

In the previous chapter, we presented our first phase of our data collection. We provided our methodology for quantitative research and the results we obtained from our content and service blocking tests.

In this chapter we present our methodology for second phase of our data collection on how we interviewed the people and we also explain in detail how we used the methodology to conduct our qualitative research. We explain in detail our research design, method of data collection, our participants, how we recruited people for the interview, what we asked, and how we discussed the polices with them.

The goal of our qualitative investigation was to find out the reasons for implementing filtering and the policy behind this decision. Through our qualitative research, we hoped to clarify if blocking was applied, and if it was because of censorship, network efficiency, or security. We also sought to clarify whether or not it had been a proper decision to block a certain website.

6.1 Methodology for interviews

In the second phase of our methodology, we wanted to interview the people who were responsible for implementing the censorship. After running the test in different organizations and finding the blocked URLs and services, we had to interview the people in charge of those organizations in order to find the reasons for implementing censorship. The details of the interview methodology is described below.

6.1.1 Research design

The research design for this part of study will be qualitative. Qualitative research seeks to explore the nature of the phenomenon without any preconceived hypotheses. Unlike quantitative research where the researcher remains objective, the qualitative researcher is a human observer who observes a human condition and is historically positioned and locally situated [68].

6.1.2 Objectives of the study

Most organizations monitor and/or control the use of the Internet through the use of logs and/or Internet filters. This monitoring/control is usually invisible, and users are only aware of it if they try to view a website or use a service that is blocked. However, there is little information about what is being blocked and the reasons for that. The goal of this section of our experiment was to identify the reasons for the types of sites and services which were being blocked in different types of organizations. This section intends to have a more informed policy debate about the need to block access to Internet websites and services and its impact on people's ability to access the information they need.

6.1.3 Benefit and scientific value of the project

To date no one in New Zealand has researched the restrictions in places which users have access to either paid or free WiFi services. Moreover, to date related policies have not been debated.

6.1.4 Ethics of this research study

Our review of literature shows that studies about blocking and censorship were not conducted in countries where they implement excessive censorship. It is highly likely that in most countries it is an inappropriate topic to even discuss the issue. In order to make the reason for implementing censorship transparent for the user, we needed to interview the people in charge of those organizations. Since this section of the study involved humans, it needed to get approval from Human Ethics Committee (HEC).

In the process of Human Ethics approval, the application for Human Ethics was filled out by the experimenter. It was also needed to prepare the consent form and an information sheet in order to give enough information to the user before they could make up their mind regarding participating in the research. In the application form, the purpose of the study, the way we questioned the interviewee and the data that we were to gather were explained. We have attached Human ethics application form in Appendix C.

Consent forms and information sheets were required to make the reason for an interview transparent for the participants. In the information sheet submitted to the Human Ethics Committee, we wrote a description of the project, the nature of the data collection, and the intended use of this data. In the consent form we explained that this data would be used by the experimenter only for the purpose of this study and it would not be used for any other purposes. It was also mentioned that the participants' identities were confidential and that their role and their organization would not be revealed. We also mentioned that they could withdraw from the

study up to two weeks after finishing data collection. The consent form and information sheet for managers are illustrated in Appendix D and E.

6.1.5 Participants

The focal participants in this study were the people in charge of implementing content and service blocking in those specific organizations in which we ran our test.

6.1.6 Recruitment

In the first part of our data collection, we ran our own Web Censorship Monitoring Tool (WCMT) on-sites. The sites were selected by compiling a list of free and paid WiFi providers by consulting sources such as the Yellow Pages and directories of WiFi providers in Wellington. We ran the project by ourselves. The main benefits of running the experiment by ourselves were that we did not expose any third parties to risks associated with accessing potentially blocked content and we could monitor the running of the tests directly to determine if there were any security or other risks that could occur during the testing.

In the second part, we contacted the people in charge of management of the service. This had been done in person at the businesses by providing a letter containing an information sheet, and, if there was no customer representative, we used resources such as the email contact detail provided by the WiFi provider and/or used resources such as the Companies Directory to source contacts details for the owners of the companies providing the service. A draft of the email is provided in Appendix F.

6.1.7 Interview questions

In the interviews, we asked the people in charge about their ideas regarding blocking and censorship, whether they were the pros or cons of blocking, to what extend they implemented censorship, from which category they had chosen the URLs and services for blocking, if they made the blocked list by themselves or they just purchased it, how often they updated their lists and so on. The participants in our study reflected their views of their organizations. In Appendix G, the list of interview questions is presented.

6.1.8 Hazards or inconvenience for the participants

There were three main risks/issues to consider:

1. We carried out the testing first and informed the organizations afterwards. The rationale was that we:

(a) simply carried out activities that an ordinary user might carry out; and, (b) if organizations were aware of our experiments ahead of time we might not be able to recruit participants and we were also interested in the difference between the perceptions of organizations with respect to their beliefs about what their restrictions and the realities were.

2. The URLs used to test for blocking had been sourced from search engines where the URLs were generated for different categories news, banking, adult sites, hackings sites, and entertainment, etc. We sourced these by using search terms associated with these categories and using URLs returned by the search engine (with safe filtering disabled). These URLs were auto generated and there was a possibility that the URLs may contain links currently blocked or monitored by the Department of Internal Affairs (DIA). It was impossible to check this ahead of time because they did not allow individuals or organizations to check which URLs may or may not be monitored at a national level.

3. After running the quantitative experiment, I provided managers

with the information sheets and ask for their participation.

6.1.9 Data collection method

The data collection was performed in one phase. After sending emails to ask them for participation, we met the IT professionals in their office. We prepared a list of questions to ask managers. They were all asked the same questions. We had been collecting two types of data: observing and interviewing. The interview data was a recorded audio for 30 to 50 minutes in one meeting. We had been wearing a microphone in the meeting in order to record the interviews.

There were different places in New Zealand which offered wireless Internet access (WiFi) either free or paid to their customers. They claimed that they were offering free Internet. By using our implemented tool that we ran at their site, these networks were investigated and the list of block URLs and services were prepared for all of these locations in order to make it transparent for users.

After observation, managers of those organizations which offered free Internet were asked to be interviewed for the reasons behind content and service blocking. That is why we had to obtain their consent in order to record their voice.

6.1.10 Data analysis

The interviews were transcribed and reviewed. Reviewing was an iterative process, which made visible themes from the data that related to the censorship polices. The inferences and conclusions were about the reasons for implementing censorship.

6.1.11 Implications of the proposed research

This study will have an impact on the transparency for the users. Because the people who implemented censorship did not like to make it clear for the users, the users could not understand the reason for not having access to certain content and services. This research sheds light on the activity of the censors.

6.2 Data collection

In this section, detailed information regarding our interviews with the IT professionals for implementing censorship and the reasons for that have been presented.

Unfortunately, three out of eight organizations replied to our request for participating in the interviews. And two out of three individuals in those organizations agreed to participate in our interviews. One possible reason for the low response is a lack of interest in the topic, though it is also possible that people did not have time to be interviewed.

The interviews were conducted between September 2013 and October 2013. Table 6.1 shows different organizations missions, region activity, and whether they had censorship or not.

The order of interview was based on the order they responded to our request emails. In order to get the signature from our participants, we took an information sheet and a consent form to all the places we went for the interview.

We also asked them to write their email address if they liked to have a copy of the results. Both Organization 1 and Organization 3 asked to have a copy of the blocked results from all the organizations.

Organization	Activity type	mission	censorship idea
Organization 1	Education	Promoting critical thinking	Having censorship
Organization 2	commercial	Contributing to economy	Having censorship
Organization 3	Education	Heritage and natural environment	Having censorship

Table 6.1
Organizations' information

People	Roles
Organization 1	Architecture and Security Manager
Organization 2	IT support
Organization 3	Manager Information Technology Services

Table 6.2
IT professionals' role in the organizations

The people we interviewed and their roles are presented in Table 6.2.

6.3 Managers' perspectives on blocking

After sending requests for an interview to all the IT professionals from our experimental networks, we interviewed those people who replied to us. A description of the interviews that had been conducted is followed below. We used pseudonyms to protect the identity of interviewees and their organizations.

6.3.1 Organization 1

Our first interview was with Jack who is an architecture and security manager.

Jack mentioned that Organization 1 used Blue Coat [43] for implementing blocking system. Blue Coat Devices are capable of doing filtering, censorship, and surveillance. These devices are used around the world. All the responsibility regarding blocking is with Blue Coat team. He also pointed out that Organization 1 purchased the list and they did not prepare the list of blocking. Then they decided which URL should or should

not be blocked.

In response to why the categorization was wrong. He said that if the website needed to be blocked, they would not care about the wrong categorization.

We asked about the benign websites which had been blocked. He maintained that this was all about Blue Coat as they did not prepare the list. He also stated that if websites were hacked they could be considered as malicious. In those cases, the websites would be blocked.

We asked why they did not treat equally all the URLs from the same category. He mentioned the decision had been made by Blue Coat.

Organization 1 just provided the service for the users but someone else was responsible for blocking. At least at Organization 1, he mentioned that if someone had a problem with any sort of blocking, he/ she could request for unblocking and the decision would not take more than one week to make in the worst case scenario.

According to Jack, they had decreased the level of blocking since three weeks ago. Organization 1 used to be more restricted but they reduced the number of blockings. This was a positive point at Organization1. When we asked him why they implement filtering, he said that the organization did not have any sort of filtering six or seven years ago but they had received complaints from parents that they did not like their children to have access to objectionable materials. It was one of the reasons to start implementing censorship. The other reason was to protect their network from bad usage. He also stated that it was also something related to the history of Organization 1 and its policy and the loyalty of Organization 1 to special communities they were registered with. These were the reasons that pushed them to start implementing blocking. In this regard he said, ”they very rarely received calls for unblocking meaning that they achieved a balance in their blockings”.

6.3.2 Organization 2

When we sent an email to Joe, the IT professional in Organization 2, to ask for an interview, we received an email from him stating that he could not comment on what sites were blocked through WiFi as he did not manage that and it was provided externally. He iterated that we would be best to talk to business internet providers such as Gen-I."

It was interesting that these places offered services to customers but they were not responsible for that. We were under the impression that if people provided services for users, they had to be responsible for that. But what we saw regarding censorship was different. We observed that third parties made decision for all these places and that the organizations just followed their decisions.

It is also interesting that they all had filtering but they were different in terms of the number of filtering and the category they blocked which showed that they did not have similar ideas regarding censorship.

6.3.3 Organization 3

We spoke With Alex who was the manager of Information Technology Services at Organization 3.

He pointed out that 18 months ago, Organization 3 started to implement blocking with Telstraclear. They wanted to have a little censorship and they wanted to have blocking for pornography websites. Organization 3 knew that Telstraclear had joined the project of blocking from DIA but since they wanted to apply more blocking, they purchased another blocking system called DansGuardian [13]. This system was good for educational systems and since Organization 3 was a family facility, DansGuardian software was a good choice for them. Alex mentioned that because the organization was a strong brand and had its own good reputation, it was not appropriate for them to provide these sorts of materials online for their users.

He said they were happy with their light level of censorship and they received a balance in their level of censorship. The only complaint they received was from a German couple when they got a blocked page for the political websites.

He also mentioned that they did not want to be active in censorship but at the same time they wanted to remain as a family environment.

When we asked him why they did blocking for certain topics such as alcohol, joke, gambling, and drugs, he mentioned he did not know that alcohol, drug, gambling, or joke websites had been blocked in his organization.

We also asked him why they did not behave the same for the same category of URLs. He mentioned it was all about DansGuardian. He also stated that as it was clear, DansGuardian was not a proper software for them to use since through using DansGuardian they had blocked web sites that they did not want to block. He said the organization had not tested their systems yet and it was great that we tested their blocking system and informed them about the results of blockings there.

He mentioned that they did not want to block gambling, alcohol, and drug. They just wanted to block pornography. He also stated that maybe these websites had links to high ranked censorship websites, and that was why they had been blocked by the software they used.

We asked him about peer to peer applications. He said that they did not allow users to use file sharing because he believed that the organization was not a proper place for file sharing activity.

We asked them whether they had different level of censorship at a different time. He said that he did not think so. He said it was possible that Telstraclear or the DansGuardian software got updates at that time and they had got different blockings at different times. He also pointed out that they wanted to keep mostly open rather than mostly closed.

We asked him why they did not use a same pattern for all their blockings, he mentioned that since Telstraclear had its own blocking system and

DansGuardian had its own, there was not a consistent pattern for blocking. We mentioned that it was not an ordinary Telstraclear service as we had the same service at home but the number of blockings was different. Then he said that Telstraclear might have added more blocking for us. It was interesting that they had very little information about how their system worked. They also did not know which categories of websites had been blocked.

We asked him why they did not prepare the list of blockings and the reasons for that. He said it was not feasible for them considering time and cost.

We asked him that people had access to 3G, so even if he blocked special contents, they could access those objectionable materials. He replied that for cellphone parents were responsible and Organization 3 was responsible for their free WiFi but not for cellphones. He also stated that they did not want people to have access to objectionable material through their network. He said the organization had its good reputation and they did want to remain as a family environment. He emphasized that most of the people would be horrified if they said they would provide all the material without any blocking. He also mentioned that a majority of people wanted blocking.

In the end, he said cost, the type of people they had, and education were important factors in providing the service. He believed that because they provided a free service, people could not ask for more. People who needed more freedom on the Internet could pay for it and have more freedom.

He said to us that if we could conclude our study with the best solution or a reasonable approach that they could apply, that would be great as there were lots of pros and cons regarding blocking.

6.4 Findings

Our data showed that there was a variety of perspectives regarding Internet censorship. Different free Internet Service Providers had different ideas regarding implementing censorship. But it was clear that they all had blocking and they were just different in terms of the number and the categories of blockings.

For example, Organization 1 and Organization 3 blocked file sharing protocols to reduce Internet traffic. Organization 1 believed that it was illegal to use file sharing application and Organization 3 said that it was not a proper place for file sharing activates. At the same time, they could not use the same reasons for blocking alcohol or joke websites. But they both mentioned because there were third parties who would make the blocked list, it was out of their control. It is interesting how unthoughtfully they purchased the list and offered the service based on that list.

In Internet censorship, motivation was the most important element we faced. Before this study we assumed that censors were motivated to block content and they just had a vague instruction to alter anything they believed was inappropriate even if it was against the law. When we interviewed the IT managers, for example, Organization 1 emphasized that half of the reasons for implementing blocking was to provide security and half for policy and history of Organization 1. Organization 1, as Jack said, had a loyalty to special committees they were registered with, and that was why they had to implement blockings.

As in Organization 3 the motivation for implementing censorship came from the idea that they did not want to damage to the reputation of Organization 3 and its famous brand. They wanted to keep the family environment of Organization 3 by implementing blocking for pornography websites. They also pointed out that Organization 3 was not a proper place for doing file sharing and they did not let people use that because it was against the law. Moreover, these applications used their bandwidth. But

he did not know that they had blocked URLs from alcohol, drug, gambling, and jokes categories. He said that as it was free service, it was customers' decision to use free service with limitation or paid services without any limitations.

Both these two organizations believed that censorship was essential for their security. They also emphasized that it would keep the society and organization safe, and that was good for protecting individuals as well as preventing abuse.

They mentioned when they blocked peer-to-peer content, they wanted to prevent the network resources from abuse, and that when they blocked the pornography, gambling, and content related to drugs and alcohol, they wanted to keep the society safe.

One interviewee stated that those organizations such as Organization 1 which had branches in different countries had different censorship criteria to match that country's culture, and these branches were controlled by external places. But unfortunately we did not manage to convince any of these organizations to take part in our research.

6.5 Summary

In this chapter, we presented our methodology for qualitative research. We defined how we conducted our research and collected the data, whom we wanted to interview, how we recruited people for an interview, and what we asked interviewees.

The purpose of conducting this phase of our data collection was to make transparency for the users by informing users regarding reasons for implementing blocking. In this chapter, we also presented our qualitative interviews. We presented the opinions of IT professionals from those organizations in which we ran our experiment. It was clear that different organizations had different implementations, software, and blocked lists in order to implement their blockings. We also saw that motivation

was different. Motivation could be for security issues, keeping the society and organization safe, network efficiency, or preventing abuse of their networks. But it was clarified that all had a level of filtering.

In the next chapter we discuss the issues we have identified regarding blocking and the problems it has raised. Based on the issues we identify, we offer our solutions.

Chapter 7

Discussion

In this chapter, we review what we presented in the previous chapters. Then we propose that because censorship is a fast growing area, researchers should investigate it more profoundly.

We believe that there are issues with implementing censorship, especially with how the system is working now. Some of these issues include: blindly purchasing a blocked list by authorities and implementing censorship based on that, filtering many benign contents as objectionable materials, not treating many similar websites in the same way, not being transparent to the users at the time of blocking, and not being responsible for wrongly implementing censorship.

Our findings show that the system of implementing censorship which is widely used had many problems. Thus, it is our duty to address this problem and ask for support in this area. If we do not address this practice and fight against it, blocking will be implemented more and more and consequently will affect more people accessing their desired material.

In what follows, we summarize our discussion into three sections, namely what is happening around the world, what is happening in New Zealand, and what we can do about it.

7.1 What is happening around the world?

Web censorship is a phenomenon across the globe. Governments monitor and censor the Internet. The policy for implementing censorship is similar to a black box for the users. Users do not know about the traffic which has been monitored and classified as censored. Everyday more and more users find the Internet has been under surveillance, controlled, and fragmented. Users believe that Internet access means accessing whatever Internet offers and not having access to the approved application and content.

Saltzer who is one of the key players in the development of the Internet in 1981 mentioned about principles of the End-to-End: "Application-specific functions ought to reside in the end hosts of a network" [75]. This principle is not being considered nowadays when the data is captured through control of either side of the connection. These activities make the open Internet under threat. By implementing censorship, also known as filtering, users are prevented from accessing the desired content considered unsuitable for them by governments.

Motivation of governments to implement censorship and take control of the Internet has increased due to huge use of the Internet. Governments easily shape the Internet based on the norms and culture of the society. Therefore, censorship has become political, social, religious, and child pornographic lookalikes in different countries.

Different countries have different scenarios and degrees of censorship for their citizens. For example, China has the strongest censorship in the world by blocking social websites such as Facebook, Twitter, Tumblr, and political websites related to political leadership, etc. Some countries, such as Saudi Arabia, consider religious morals in implementing censorship. Iran, as another example, talks about fragmenting the Internet and is going to have "Iranian Internet". It allows the flow of information within the country but not beyond the country.

Filtering the Internet has profit for authorities. Considering the econ-

omy, making limitation on Internet is more profitable. Considering social and political aspects, authorities are capable of taking control of the society and preventing it from harmful activities which are against the law and not suitable for the government. To put everything in a nutshell when governments close the Internet, it is much easier to control it.

While some people believe that censorship is not self-regulated and the governments are responsible for implementing the censorship, there are others who believe that censorship is not a great idea and it affects the users' needs and trust.

People who oppose censorship believe that there are different issues with the practice of blocking. For example, people will not be able to access their desired material which has an effect on trust in society, its knowledge, and democracy.

Censorship regarding films, books, and games is clear and transparent for the users but not for the websites and services. This lack of transparency will leave people confused as to whether the website is blocked or offline. The other impact of lack of transparency in implementing censorship is that benign content is sometimes blocked and classified as offensive content by authorities. If censorship is transparent the benign websites which are considered offensive will be clarified and the beneficial content will not be restricted for people. Moreover, secretly implementing the censorship would make citizens lose trust in governments.

The idea of filtering comforts parents and authorities that their children or staff are prevented from accessing the unwanted content. This may give a false sense of security to parents which is not appealing. The government, Internet Service Providers, and families are responsible to teach parents how to prevent their children from accessing inappropriate content. It is also beneficial if parents can teach their children to be responsible for their safety instead of waiting for their parents to provide safety for them. Most of the people in general and children in particular access the Internet through their phones which increases the concern of

their families.

Censorship or filtering is an offense to democracy. It is similar to the government holding the users' hands to prevent them from doing things which, it thinks, can offend them. The problem is that government thinks that by censoring websites they could prevent users from accessing the inappropriate content. This is true for just a portion of the society, as there are many ways for the motivated people to bypass censorship.

7.2 What is happening in New Zealand?

New Zealand is a digital country since Internet is used in 4 out of 5 New Zealand homes [22]. In terms of Internet access, New Zealand is one of the countries with the highest Internet access rate. There are different reasons which have led the number of Internet users to increase such as a decrease in the price of broadband, mobile access, ADSL, and motivation for applying for jobs online, etc. More than 93% of the Internet in New Zealand is provided by ISPs such as TelestraClear, Telecom, and Vodafone. These Internet Service Providers have implemented filtering in conjunction with DIA.

DIA states that the URLs which are mostly related to child abuse materials are restricted and no one knows exactly what they have blocked. Once the government starts blocking, they can start to filter websites which are not convenient for them. All requests will be routed to the government servers. The user's request will be compared to the blocked list. If it is matched with their black list, the request will be denied. This blocked list is revised by staff each month to have an updated blocked list every time [15].

In [15], it is published that Child Exploitation Filtering System costs $150000 which is given freely to the ISPs to block around 7000 objectionable sites. It is also published that the number of blocked websites is 5 times more than the ones in the UK list and twice as much as those in

Australia. It is in contrast to what is heard from the public and published by ONI that censorship in Australia is more than it is in New Zealand.

Techliberty [63] announced, The Government has no Mandate to Filter the Internet. They mentioned that censorship was not covered by law and still no laws were passed in the Parliament. They believe that implementing censorship to mass websites is against the Bill of Rights.

Even if censorship is legitimate, implementing it secretly does not give good sense to people. As a response to this, DIA claims that publishing list of web sites is a pointer to the crime and DIA uses its power not to publish it. Starting to implement censorship gives power to governments to implement more censorship whenever and whatever they like.

A survey commissioned by InternetNZ [52] about public thoughts about the government's Internet filter has shown interesting results about this study. Only 9% of the people knew whether or not their ISP used government filter. The ISPs which provide more than 90% of the NZ Internet market use government's censorship program. Only 23% of the people wanted the government to filter their Internet connection.

It is also worth mentioning that authorities know that even implementing censorship is not effective and motivated people will access the desired content. Given that, what is the reason for breaching the privacy and freedom of the citizens? Prior to this action, the citizens cannot trust the government.

DIA has clarified that censorship has been applied to child pornography sites. The question remains why they do not consider other ways such as requesting the servers hosting these sorts of websites to delete them. Child pornography is illegal in almost all countries. Thus, it is a better idea to fight against this issue globally by removing it from Internet and not implementing censorship on what brings dishonesty for governments. Even if the government believes that child pornography is blocked, it is still there. There are lots of websites with the same content.

7.2.1 Specifying the problem

There are problems with the level of filtering such as:

1. In New Zealand censorship is applied secretly and is not transparent for people. One of the effects of implementing censorship secretly is that some of the websites are incorrectly blocked. For example, German couples could not access a political site in Germany through Te Papa Museum free WiFi because it was categorized as Japan's porn websites [54].

2. It is mentioned that only child abuse web sites are blocked which is not true and there are more websites and services which are blocked. Based on the work we have done, depending on the organizations, there are other categories of websites and services which are blocked as well.

3. If filtering is applied for children in order to keep them safe on the Internet, it is hard for adults to bypass it. At the same time those who were the target of censorship could bypass it.

4. If organizations are against child pornography, they should fight against this issue globally not breaching privacy of people.

5. Censorship or filtering operates as an offense to democracy. It is similar to the government holding the users' hands to prevent them from doing things which, it thinks, can offend them. The problem is that government thinks by censoring websites, they could prevent users from accessing the inappropriate content. This is true for just a portion of the society, as there are many ways for the motivated people to bypass censorship.

6. Authorities who implement censorship are not responsive. Even the authorities mention that users should inform in case things are wrongly blocked, they are not easy to convince.

7.2.2 Problems arising by implementing censorship

There are different problems which arise in society by implementing censorship. Firstly, implementing censorship affects the economy of the businesses because customers cannot get the direction or information from the website. Secondly, implementing censorship will give good sense of safety to parents that their children are safe on the Internet but actually the content is still there. Thirdly, filtering will affect the knowledge of society. For example, filtering sexual material will prevent young people from accessing the healthy information, making young people blind in terms of their future safety. The other effect is the performance of the Internet. Since all the traffic needs to go through DIA, it may cause a performance issue and will make a single point of failure. Fifthly, distrust will come to the society, and citizens of the country will not trust the government as they know they are censoring more than child pornography websites.

7.3 So, what should be done?

We were motivated to conduct this study to find out about the scale of the problem to find the blocked websites and services in order to make transparency for people. It is clear that blocking has been applied to websites other than child abuse, but finding all the blocked content has not been easy. Although the list of blocked websites is not published by DIA, the users of the Internet know about the probability of censoring some websites. Child abuse websites are blocked by government but different sites and services are blocked at different times by ISPs and organizations. All the filtering affects principles of human rights organizations. But because Internet filtering is so widespread and supported by strong opposition, it is so hard to debate.

Finding a unique and reliable way of finding censorship was not easy. We had limitation in terms of using different ports as most of the desired

ports for us were blocked in different organizations. Therefore, we had to make lots of changes in our implemented tool for each organization. Also due to the importance of censorship subject, we had to use manual analyses as well as automatic ones.

Our experience in places such as Organization 3, Organization 4, Organization 1, Organization 5, Organization 7, Organization 8, organization 2, and organization 6 showed how these places treated their WiFi users differently and how they restricted the use of Internet to specific traffic and websites through specific ports.

We tried to access a variety of URLs from different categorizations using our implemented tool to find out about the prevalence of censorship in different organizations. It was thought that there might not be much content blocking in organizations in New Zealand but there were of course ports and services which were blocked. Our results showed that not only lots of ports and services had been blocked but also there were lots of URLs from each category which had been blocked.

For example, in a short distance from organization 5, organization 1 offers free Internet to their customers but with more restrictions in accessing the websites and using ports. Organization 1 restricts access to most services by blocking ports such as FTP, SSH, SMTP, DNS, POP3, RPC, NetBIOS, IMAP, SNMP, SMB, MTP/SSL. In organization 1 we also tested the free WiFi with our first round of URLs. Out of 180 URLs, 114 URLs were censored. These include categories from proxy, gambling, malicious sources, adults, file sharing, anonymizer websites, racism, drug, religions, and games. If they were eager not to allow access to objectionable material, it was possible for users to access it through their mobile phones or other free WiFis close to them.

In contrast, organization 5 provides open access to all the services. Almost all the tested URLs were also open access except four URLs.

In organization 4 most of the ports we tested were open to use except for DNS (port 53), RPC (port 135), NetBIOS (port 139), and SNMP (port

161). These ports were expected to be blocked as they were not used generally in local networks. But it was interesting how organization 4 used different ways of implementing censorship for different contents. For example, for some of the websites from jokes and adult entertainment category they were clear by providing transparent blocked messages. For some of the websites from Peer-to-Peer (P2P) or adult entertainments category, they sent an error message. The problem was that a consistent pattern had not been used for all blockings. organization 4 was one of the organizations with high number of content blocking.

At organization 3 access to ports SMB (port 445), SMTP (port 25), DNS (port 53), RPC (port 135), and NetBIOS (port 139) were blocked. We expected to have these port blockings because of security purposes. But it is interesting how organization 3 did over-blocking by implementing benign websites as offensive ones and how they categorized these blocked websites wrongly. For example, organization 3 blocked access to some online shops selling children's toys and they categorized it under "Pornography; Extreme". Our experiment in organization 3 identified that the URLs had been blocked from different categories such as proxy, gambling, malicious sources, adults, file sharing, anonymizer, racism, and drug.

In organization 6, all the services we test were open except TCP access to remote SMTP server's port 25 which was prohibited. With this blocking it was not possible to send email via SMTP. This sort of blocking was very common because this port could be used by hackers for generating spam. In terms of content blocking, organization 6 blocked some of the content from adult entertainment by redirecting them to administrator's desired content.

Organization 2 provided access to all services. They had a couple of blockings which were not transparent for users. In their strategy, they redirected the request for objectionable content from adult entertainment to administrator's desired content.

The organization 8 provided access to all services. In terms of content

blocking they redirected the request to desired content of authorities. In some cases a transparent error message was shown to the users.

Organization 7 was extreme in terms of service blocking. Ports 25, 53, 80, 110, 143, 443, and 993 were open and all others were closed. In terms of content blocking they blocked content from pornography and P2P websites.

Our interview with Organization 1 and Organization 3 showed that for example Organization 1 applied blocking due to two important reasons: security and network efficiency. Organization 1 also mentioned that this was all about the policy and history of the organization and its loyalty to the organizations it was registered with.

Organization 3 mentioned that their organization was a great brand and they wanted to keep the great brand of their organization and its family environment. But they had less information regarding their blocking system. For example, they did know that they had blocking from alcohol, drug, and gambling websites.

We also obtained interesting feedback from Organization 2. The IT professional from this organization pointed out that they were not responsible for their free WiFi. It is interesting that they offered free WiFi but they were not responsible for that. It shows how blindly they implemented the system based on what they purchased.

The variety of the results of what was blocked showed that ISPs and blocking software did not have a set of agreed approaches for implementing blockings. More open discussion about what it is appropriate to block and what should be available is needed.

We would like to argue that implementing censorship at a national level is not a good idea and implementing censorship by individuals is a better decision. There is another solution to identify adults from young children. Then we can ask adults whether or not they want censorship and give them all the information about censorship, how they implement it and what the categorization is. We also need to teach them how to report

if the censorship is incorrectly applied.

It is also possible that parents who need to implement the blocking for their children get information through their ISPs when they subscribe for the services. This solution will be helpful for children who need safety on the Internet.

It has to be mentioned to the authorities and parents who like to prevent users and children from watching these objectionable materials in their sites that at the same time there are different sources of getting Internet for people. Therefore, if one of these free WiFi's blocks some websites, the users still have access to 3G. So free Internet Service Providers cannot worry about downloading banned materials. And, it is not their responsibility to control their Internet usage. Therefore, blocking content could not be effective when people access different sources to get information. It is, then, better to implement blocking on individual computers and cellphone devices through parents.

Educating people is very important. Government could educate parents how to keep their children safe on the Internet. At the same time government and parents can teach children how to keep themselves safe and not wait for their parents and government to keep them safe in the Internet.

Thus, we believe we could consider other solutions as suggested below:

1. We should not apply blocking at national level, and let individuals implement blocking.
2. We should educate children about how to be safe on the Internet.
3. We should fight against the issue of child pornography globally.
4. We should be transparent for the users by providing the list of blockings in terms of content and services.
5. We should be helpful with providing reasons for blockings.

6. We should be more cautious in choosing the software for implementing blocking.

7. We should not rely on the available blocked lists and being more cautious on categorizing blocked content.

7.4 Summary

This chapter discussed the prevalence of Internet blocking around the world and particularly in New Zealand. We presented the problems of implementing central censorship and the issues raised by these problems. We emphasized that nobody knew what was happening on the Internet and what would happen to the Internet in future. There was a need for a system to collect, analyze, provide visibility to manage the Internet better. We mentioned that censoring free movies and music prevented illegal downloading of files which were against the copyright agreement. Censoring child pornography kept children safe. But when it came to censoring adults' jokes, and political websites, it was annoying for people. We also discussed that there was a regulation in adults and circumvention tools. It was not ethical to restrict adult, entertainment and social networks for people.

Chapter 8

Conclusions

This thesis has addressed the prevalence of content and service blockings in different organizations in Wellington, New Zealand. Due to huge use of the Internet, governments and organizations have been motivated to implement censorship and take control of the Internet. This blocking, whether for security or for network efficiency, has significant effects on people's access to services and information. Also, this sort of restriction will affect the society in terms of economy, communication, knowledge, expressing ideas, etc. These issues are often not considered when authorities implement blocking.

There are lots of studies conducted around the world to prove Internet blocking at a governmental level. But there was no such study conducted in New Zealand. This study bridges the gap by focusing on censorship at an organizational level. The responses to what is being blocked in New Zealand, who implements blocking in New Zealand, and what the reasons for these blockings are, motivated us to conduct our study. And, we did our study in Wellington New Zealand in places we had access to their WiFi or where they provided free WiFi for their customers in order to make transparency for the users.

For this reason, in Chapter 2 of our thesis, we gave a discussion on Internet filtering around the world and particularly in New Zealand and

provided a discussion on background and related work. We also reviewed tools for investigating Internet filtering. We reviewed different monitoring tools since we did not want to implement a new tool. Unfortunately, we could not use available tools for conducting our study since they just focused on a special traffic such as HTTP or they did not consider the limitation of our context in terms of service blocking. Comparing available web censorship monitoring tools led us to choose OONI. Our preliminary work showed that OONI did not work in our context due to different issues. Firstly, OONI used Tor and Tor was blocked in almost all organizations we focused on. Secondly, using this tool gave us lots of false positives since OONI compared just the length of content fetched from experimental with Tor control network. It was possible that there were lots of contents whose lengths were similar but their contents were different. Thirdly, OONI did not consider service blocking. Due to these issues with OONI we set the requirements for our own tool.

In Chapter 3 we presented the design and implementation of our new tool (WCMT) based on the issues we identified with OONI. We implemented our own web server which acted as a proxy and request the page on behalf of the client from VUW open network and Tor network. This enabled us to compare the contents at the same time. We compared the results retrieved from our control and experimental networks in terms of header length, body length, header and body content and images to decrease the number of false positives. Our tool became capable of finding service blocking because we assumed we would not have much content blocking but we would have service blocking. However, the results showed that we had both content and service blockings.

In Chapter 4 we performed a preliminary test on our tool in order to see how our new tool could satisfy our requirements. Our performance evaluation showed that WCMT successfully covered the bandwidth limitation of our context. The tool was fast enough to run 2400 URLs.

In Chapter 5, we set our methodology for running quantitative re-

search. We ran our first phase of experiment using our implemented tool (WCMT). We prepared a profile for each of the organizations by identifying the list of blockings in terms of contents and services. It was shown that different levels of blocking existed for organizations. For example, organization 1 by 18.88% was extreme in terms of content blocking followed by organization 3 with 9.12% and organization 4 with 5.58% blocking. These values were not absolute ones as these websites were sampled and the test was conducted at a specific time and it could change now and then.

In terms of service blocking, different organizations blocked different number of services. Organization 7 with 15 port blockings was extreme followed by organization 1 with 12 blockings and organization 3 and organization 4 with 5 port blocking. organization 2, organization 8, and organization 5 did not block any services for their customers. Our evaluation of content and service blocking at organizational level showed that our methodology was successful in finding content and service blocking.

In Chapter 6, we set our methodology for our qualitative research to interview authorities about implementing censorship. We wanted to make the reason for blocking transparent for the users. Since our study involved people, we had to apply for Human Ethics approval. We filled the Human Ethics application form and prepared an information sheet and consent form for the focal participants. After getting approval from the Committee, we started the second phase of our data collection.

We started our study by requesting for an interview from IT managers of organizations in which we ran our tool through their network. We liked to have a more informed policy debate about the need to block access to Internet websites and services, and its impact on people's ability to access information which they needed. Three out of eight organizations replied to our emails requesting for an interview. Two out of those three accepted to take part in an interview. Our data showed that there were different reasons for implementing blocking such as performing security, protecting

individuals, and network efficiency.

Organization 1 mentioned that they applied blocking due to two important reasons: security and network efficiency. He also mentioned that the reason for implementing blocking came from the policy and history of the organization and its loyalty to the organizations they were registered with.

The IT professional in Organization 3 mentioned that this organization was a great brand and they wanted to keep their great brand and its family environment. But they did not have enough information regarding their blocking system. For example, they did not know that they had blocking from categories such as alcohol, drug, and gambling websites.

But unfortunately, Organization 2 mentioned they were not responsible for their free WiFi. It was interesting that they offered free WiFi but they were not responsible for that. It showed how blindly they implemented the system they purchased.

In Chapter 7, we presented a discussion based on the issues we had identified in our study. Our study concluded that central blocking, itself was not a good idea for preventing access to the content because there were always different ways to bypass the blocking for a motivated person. We believe that there are issues with implementing central blocking, especially with how the system is working now.

Our quantitative study showed that there were blockings in all the organizations. and the blockings were different just in terms of category and their number. It was interesting how different organizations looked at their customers differently, how they were not transparent to their customers with not using the same pattern in all blockings, and how their polices changed over time. We have to emphasize that the variety of the results of what was blocked showed that ISPs and blocking software did not have a set of agreed approaches for implementing blockings. More open discussion about what it is appropriate to block and what should be available is needed.

Our qualitative study showed that either the topic of investigating blocking was not an interesting topic for the authorities or that they were very busy since we received a few replies to our email regarding request for the interview. The results showed that they blindly purchased their blocking list from third parties and they did not have enough information regarding their blocking system.

In the following section, I will discuss specific contributions of this study.

8.1 Contributions

Our study represented first content and service blocking measurements in New Zealand. We hoped that our findings could give enough information regarding censorship and network neutrality debates. The goal of this study was to raise people's awareness and ask them to contribute to this area. Specific contributions of my thesis are described below:

1. Designing and building a new tool for detecting content and service blocking that does not rely upon access to an oracle network (Tor) at the testing sites. This new tool was called WCMT.

2. Designing and conducting an empirical study to quantify what was and was not blocked at a representative sample of organizations providing free wireless access to the Internet.

3. Designing and conducting a qualitative study to follow up with the organizations investigated quantitatively in order to look into the motivation for and understanding of blocking polices adopted by organizations.

8.2 Further study

This thesis presented the design, development, and analysis of WCMT, but there were areas that needed improvement. Here, I offer further suggestions:

Firstly, in this thesis we concentrated on content and service blocking while there are other ways for ISPs to do traffic manipulations. It would be useful to find out about different ways ISPs used to manipulate traffic.

Secondly, we checked special ports or services, while there are other services and ports which may be blocked. It is beneficial to identify them.

Thirdly, we have run our experiment in 8 Internet service providers while there are other places which have equal or more censorship. It is useful if we could prepare a list of blockings in other places as well.

Last but not least, during the qualitative part of our study, we understood that there was variation in terms of number and category of blockings in different places. When we asked IT professionals for interview, we also noticed variation in their responses or lack of interest in doing so. Hence, in a future work, we recommend further investigation into the underlying reasons why IT professionals are apparently reluctant in dealing with censorship in their relevant organizations in New Zealand.

Appendices

Appendix A

Content blockings' results

In this appendix, the results of content blocking tests in different organizations are given.

A.1 Content blocking in organization 1

The results of content blocking tests in organization 1 are given in Table A.1.

Table A.1
Organization 1 test results

Category	Blocked Sites

Continued on next page

Table A.1 (*continued*)

Category	**Blocked Sites**
Adult/Mature Content; Illegal Drugs	http://www.pass-any-drug-test.com/, http://www.the-hive.ws/, http://www.weedtalk.com/, http://www.hightimes.com/, http://www.the-head-shop.co.uk/, http://www.cannabis.com/, http://www.marijuana.com/, http://amphetamines.com/, http://heroin.org/, http://www.neonjoint.com/, http://www.erowid.org/, http://www.mpp.org/, http://www.overgrow.com/, http://www.yahooka.com/, http://www.project420.com/, http://www.thegooddrugsguide.com/, http://www.shroomery.org/, http://bong.com/, http://www.420.com/, http://www.everyonedoesit.com/, http://www.marijuanareform.org/, http://www.420auction.com/, http://pass-any-drug-test.com/, http://hightimes.com.
Malicious Sources	http://warco.pl/, http://kremlinhotel.ru/js/jshttp/pha/seiko.php, http://368500.cn/vm/to.htm, http://saudieng.net/, http://gmpg.org/xfn/11.
Malicious Sources; Malicious Outbound Data/Botnets	http://3apa3a.tomsk.tw/c/cfg.bin, http://panazan.ro/online/libraries/pa http://panazan.ro/online/libraries/pattemplate/patTemplate/Modifier/I
Malicious Sources; Malicious Outbound Data/Botnets; Newsgroups/Forums; Society/Daily Living	http://floranimal.ru/articles/mashrooms/zh/cfg.bin, http://www.hotspotshield.com/, http://twitter-badges.s3.amazonaws.com/.

Continued on next page

Table A.1 (*continued*)

Category	Blocked Sites
Adult/Mature Content; Shopping	http://www.underworldfashions.com/, http://lingerie.com/, http://exoticfashionmall.com/, http://www.ignitethepassion.co.nz/, http://lingeriebowl.com/, http://bodylingerie.com/, http://venusswimwear.com/, http://victoriassecret.com/, http://abcunderwear.com/, http://trashy.com/, http://blueskyswimwear.com/, http://figleaves.com/, http://spikybras.com/, http://freshpair.com/, http://panties.com/, http://mehzavod.ru/.
Adult/Mature Content	http://www.wolfmarksden.com/, http://www.fanpix.net/picture-gallery/0705289/csi-miami-pictures.html, http://www.prettynudists.com/
Pornography	http://udfn.com/, http://www.k-k-k.com/, http://crazyshit.com/, http://desijammers.com/, http://nudes.hegre-art.com/, http://www.nudistnudes.com/, http://www.met-art.com/, http://www.nztop100.co.nz/, http://tgpme.com/, http://vi5search.com/, http://7chan.org/, http://www.avizoon.com/, http://smsmovies.net/, http://ww42.proscribed.com/, http://bangbus.com/, http://www.nudes-nudes.com/, http://www.nudesfromdownunder.com/, http://www.mc-nudes.com/, http://www.gmbill.com/, http://xxx.com/, http://www.adultauctions.co.nz/, http://www.retropornarchive.com/.
Adult/Mature Content; Sex Education	http://www.nzpersonals.com/, http://adultshop.nzpersonals.com/.

Continued on next page

Table A.1 (*continued*)

Category	**Blocked Sites**
Pornography; Extreme; Scam/ Questionable/ Illegal	http://tiptopteens.net/.
Child Pornography	http://sexual.vipzax.com/, http://risk.vipzax.com/.
Adult/Mature Content; Placeholders	http://nude.vipzax.com/, http://teens-models.org/.
Pornography; Scam/ Questionable/ Illegal; Suspicious	http://smalltopsite.com/.
Proxy Avoidance; Suspicious	http://www.vtunnel.com/.
Proxy Avoidance	http://getmearound.net/, http://ooni.nu/, http://www.hotspotshield.com/, http://www.ad-free-proxy-site.info/, http://www.ad-free-proxy-site.info/, http://www.anonymizer.ru/, http://tornadoproxy.com/, http://www.fsurf.com/, http://getmearound.net/, http://torproject.org/ http://www.vpnbook.com/, http://www.justfreevpn.com/, http://www.hotspotshield.com/, http://www.proxy4free.com/, http://proxy.org/, http://proxyserver.asia/, http://exitb.net/, http://psiphon.ca/, http://ultrasurf.us/.

Continued on next page

Table A.1 (*continued*)

Category	**Blocked Sites**
Peer-to-Peer (P2P)	http://songbox.pk/, http://thepiratebay.org/, http://kickassunblock.info/search/kickasstorrents/, http://h33tunblock.info/, http://www.bittorrent.com/, http://eztv.it/, http://www.torrentbytes.net/, http://tracker.istole.it/.
Adult/Mature Content; Malicious Outbound Data/ Botnets; Personals/ Dating; Chat/ Instant Messaging	http://friendfinder.com/
Pornography; Extreme	http://bbs12.mail15.su/, http://ourworldkids.info/, http://pretty-pretty.info/, http://eroticaexpo.co.nz/, http://youngtop.info/.
Adult/ Mature Content; Humor/ Jokes	http://www.lotsofjokes.com/, http://www.jokesgalore.com/, http://www.jokesgallery.com/, http://dirtyjokesinc.com/, http://collegehumor.com/.
Adult/ Mature Content; Personals/ Dating	http://broonline.co.nz/, http://sxetc.org/ http://findsomeone.co.nz/, http://www.nzdating.com/, http://w3.nzdating.com/, http://broonline.co.nz/.

Continued on next page

Table A.1 (*continued*)

Category	**Blocked Sites**
Gambling; Malicious Sources; Malicious Outbound Data/ Botnets	http://carsands.com/, http://www.poker.net/, http://www.ildado.com/, http://sportsgambling.about.com/, http://www.casinogamblingexposed.com/, http://www.bjmath.com/, http://www.simslots.com/, http://online-keno.com/, http://onlinegamblingtips.com/, http://blackjackinfo.com/, http://keno-info.com/, http://blackjackplaza.com/, http://bjmath.com/, http://blackjack-gambler.com/, http://sportsgambling.about.com/%20h, ttp://www.casinogamblingexposed.com/, http://ildado.com/, http://roulette.sh/, http://gamblingnewsletter.com/, http://allcraps.com/, http://pokerroom.com/, http://planetpoker.com/, http://onlinecasino.com/, http://poker.com/, http://homepoker.com/, http://gamingday.com/, http://poker.net/, http://4online-gambling.com/, http://gambling.com/, http://www.nzlotteries.co.nz/.
Adult/ Mature Content; Alcohol	http://moderndrunkardmagazine.com/.
Search Engines/ Portals; Malicious Outbound Data/ Botnets	http://jihadonline.org/, http://www.megago.com/1/.
Adult/Mature Content; Entertainment	http://photos.lucywho.com/csi-miami-photos-t685904.html, http://www.acephotos.org/t685904/csi-miami-photos.html.

Finished on next page

Table A.1 (*finished*)

Category	Blocked Sites
Entertainment; Pornography	http://www.iheartchaos.com/.
hate speech	http://americannaziparty.com/, http://stormfront.org/, http://godhatesfags.com/, http://nsm88.com/, http://nationalvanguard.org/, http://nationalist.org/, http://jewwatch.com/, http://thebirdman.org/, http://bhbulgaria.com/, http://martinlutherking.org/, http://skrewdriver.net/, http://resistance-radio.com/, http://armyofgod.com/, http://www.front14.org/, www.sedoparking.com/feminista.com/.
Hacking	http://hacktivismo.com/, http://nmrc.org/, http://hackcanada.com/, http://cultdeadcow.com/.
Weapons	http://collectiblefirearms.com/, http://guns.ru/pvo/, http://uws.com/, http://aum-shinrikyo.com/, http://zmweapons.com/, http://hecklerkoch-usa.com/.
Sex education	http://condoms.getiton.co.nz/, http://www.sextherapy.co.nz/, http://sieccan.org/, http://premaritalsex.info/%20, http://www.ultimatebirthcontrol.com/, http://positive.org/, http://plannedparenthood.org/, http://sfsi.org/, http://scarleteen.com/, http://siecus.org/, http://teensource.org/.

A.2 Content blocking in organization 2

A sample of websites which lead to redirections is:

`http://xtra.co.nz`, and `http://pretty-pretty.info/images/js\_preloader.gif`.

A.3 Content blocking in organization 3

Table A.2 shows the results of content blocking tests at organization 3's wireless network.

Table A.2
Organization 3's test results

Category	**Blocked Sites**
Category	blocked sites
Adult/Mature Content; Violence/Hate/Racism; Extreme	http://www.rotten.com/.
Adult/Mature Content; Illegal Drugs	http://www.shroomery.org/.
Malicious Sources	http://warco.pl/, http://kremlinhotel.ru/js/jshttp/pha/seiko.php, http://368500.cn/vm/to.htm, http://saudieng.net/.
Malicious Sources; Malicious Outbound Data/Botnets	http://3apa3a.tomsk.tw/c/cfg.bin, http://panazan.ro/online/libraries/pa http://panazan.ro/online/libraries/pattemplate/patTemplate/Modifier/I

Continued on next page

Table A.2 (*continued*)

Category	**Blocked Sites**
Malicious Sources; Malicious Outbound Data/Botnets; Newsgroups/Forums; Society/Daily Living	http://floranimal.ru/articles/mashrooms/zh/cfg.bin, http://www.hotspotshield.com/, http://twitter-badges.s3.amazonaws.com/
Adult/Mature Content; Shopping	http://www.underworldfashions.com/, http://lingerie.com/, http://exoticfashionmall.com/.
Adult/Mature Content	http://www.wolfmarksden.com/, http://www.fanpix.net/picture-gallery/0705289/csi-miami-pictures.html, http://www.prettynudists.com/.
Pornography	http://udfn.com/, http://www.k-k-k.com/, http://crazyshit.com/, http://desijammers.com/, http://nudes.hegre-art.com/, http://www.nudistnudes.com/, http://www.met-art.com/, http://www.nztop100.co.nz/, http://tgpme.com/, http://vi5search.com/, http://7chan.org/, http://www.avizoon.com/, http://smsmovies.net/, http://ww42.proscribed.com/, http://bangbus.com/,0 http://www.nudes-nudes.com/, http://www.nudesfromdownunder.com/, http://www.mc-nudes.com/, http://www.gmbill.com/

Continued on next page

Table A.2 (*continued*)

Category	**Blocked Sites**
Adult/Mature Content; Sex Education	http://www.nzpersonals.com/, http://adultshop.nzpersonals.com/.
Pornography; Extreme; Scam/Questionable/Illegal	http://tiptopteens.net/.
Child Pornography	http://sexual.vipzax.com/, http://risk.vipzax.com/.
Adult/Mature Content; Placeholders	http://nude.vipzax.com/, http://teens-models.org/.
Pornography; Scam/Questionable/Illegal; Suspicious	http://smalltopsite.com/.
Proxy Avoidance;Suspicious	http://www.vtunnel.com/.
Proxy Avoidance	http://getmearound.net/, http://ooni.nu/, http://www.hotspotshield.com/, http://www.ad-free-proxy-site.info/, http://www.ad-free-proxy-site.info/, http://www.anonymizer.ru/, http://tornadoproxy.com/, http://www.fsurf.com/, http://getmearound.net/, http://torproject.org/ http://www.vpnbook.com/, http://www.justfreevpn.com/, http://www.hotspotshield.com/, http://www.proxy4free.com/, http://proxy.org/, http://proxyserver.asia/, http://exitb.net/, http://psiphon.ca/, http://ultrasurf.us/.

Continued on next page

Table A.2 (*continued*)

Category	**Blocked Sites**
Peer-to-Peer (P2P)	http://songbox.pk/, http://thepiratebay.org/, http://kickassunblock.info/search/.kickasstorrents/, http://h33tunblock.info/, http://www.bittorrent.com/, http://eztv.it/, http://www.torrentbytes.net/, http://tracker.istole.it/.
Adult/Mature Content; Malicious Outbound Data/Botnets; Personals/Dating; Chat/Instant Messaging	http://friendfinder.com/.
Pornography; Extreme	http://bbs12.mail15.su/, http://ourworldkids.info/, http://pretty-pretty.info/, http://eroticaexpo.co.nz/, http://youngtop.info/
Adult/Mature Content; Humor/Jokes	http://www.lotsofjokes.com/, http://www.jokesgalore.com/, http://www.jokesgallery.com/, http://dirtyjokesinc.com/, no problem with this site:, http://collegehumor.com/.
Search Engines/Portals; Malicious Sources; Malicious Outbound Data/Botnets	http://jihadonline.org/, http://www.megago.com/l/?.

Finished on next page

Table A.2 (*finished*)

Category	Blocked Sites
Adult/Mature Content; Personals/Dating	http://broonline.co.nz/, http://sxetc.org/.
Gambling; Malicious Sources; Malicious Outbound Data/Botnets	http://carsands.com/,
Adult/Mature Content; Alcohol	http://moderndrunkardmagazine.com/.
Pornography	http://www.udfn.com/, http://xxx.com/.
Adult/Mature Content; Entertainment	http://www.acephotos.org/t685904/csi-miami-photos.html.
Entertainment; Pornography	http://www.iheartchaos.com/.

A.4 Content blocking in organization 4

The list of blocked URLs in organization 4 is illustrated in Table A.3.

Table A.3
Organization 4's test results

Category	Blocked Sites
	Continued on next page

Table A.3 (*continued*)

Category	**Blocked Sites**
Pornography; Extreme; Scam/Questionable/Illegal	http://tiptopteens.net/.
Pornography	http://bangbus.com/, http://ww42.proscribed.com/, http://smsmovies.net/, http://www.avizoon.com, http://7chan.org/, http://www.nudistnudes.com/, http://www.met-art.com/, http://eroticaexpo.co.nz/, http://nudes.hegre-art.com/,
Adult/Mature Content; Placeholders	http://teens-models.org/, http://www.mc-nudes.com/, http://www.nudes-nudes.com/, http://www.nudesfromdownunder.com/, http://www.udfn.com/, http://eroticaexpo.co.nz/, http://pretty-pretty.info, http://crazyshit.com/, http://tgpme.com/.
Adult/Mature Content	http://www.prettynudists.com/.
Adult/Mature Content; Shopping	http://exoticfashionmall.com/, http://lingerie.com/, http://www.panties.com/, http://www.trashy.com/, http://www.bodylingerie.com/, http://www.lingeriebowl.com/.
Adult/Mature Content; Humor/Jokes	http://www.jokesgallery.com/, http://www.crazyshit.com/, http://www.the-jokes.com/, http://justjoking.com/, http://www.funnyjokes.com/.
Adult/Mature Content; Alcohol	http://www.moderndrunkardmagazine.com/, http://www.skyy.com/, http://www.beer.com/.
Pornography; Extreme	http://bbs12.mail15.su/, http://ourworldkids.info/.

Finished on next page

Table A.3 (*finished*)

Category	Blocked Sites
Malicious Sources; Malicious Outbound Data/Botnets	http://3apa3a.tomsk.tw/c/cfg.bin.
Pornography; Extreme; Scam/Questionable/Illegal	http://tiptopteens.net/.
Drug	http://www.hightimes.com/, http://www.marijuanareform.org/.
Gambling	http://www.casino.net/en/default.htm.
gay and lesbian	http://www.tsroadmap.com/, http://transsexual.org/, http://www.queernet.org/, http://www.gaycenter.org/, http://www.gayegypt.com/, http://www.lesbian.org/, http://www.bisexual.org/, http://lesbians-against-violence.com/, http://gaytoday.badpuppy.com/.
hacking	http://www.hackinthebox.org/
Adult/Mature Content; Abortion; Sex Education	http://www.sxetc.org/, http://www.scarleteen.com/, http://www.sfsi.org/, http://www.positive.org/, http://www.sieccan.org/, http://www.teenpregnancy.org/, http://www.nostatusquo.com/ACLU/dworkin/, http://www.feminista.com/, http://www.agender.org.nz/.

A.5 Content blocking in organization 5

`http://vi5search.com`,
`http://xtra.co.nz`,
`http://www.feminista.com/`,
`http://youngtop.info/`.

A.6 Content blocking in organization 6

The URls blocked in CBDfree WiFi are as follows:
http://youngtop.info,
http://prety-prety.info,
http://xtra.co.nz/,
http://feminista.com/,
http://www.igc.apc.org/Womensnet/dworkin/,
http://saudieng.net/,
http://www.avizoon.com/,
http://tiptopteens.com,
http://teens-models.org/,
http://www.feminista.com/,
http://sexual.vipzax.com/,
http://risk.vipzax.com/,
http://nude.vipzax.com/.

A.7 Content blocking in organization 7

The list of URLs which were redirected when we requested them is shown below:
http://youngtop.info/,
http://xtra.co.nz/,
http://pretty-pretty.info/,
http://tiptopteens.net/.

The following Peer to peer URLs could not be seen at organization 7:
http://songbox.pk/,
http://thepiratebay.org/,
http://kickassunblock.info/search/.kickasstorrents/,
http://h33tunblock.info/,

http://www.bittorrent.com/,
http://www.torrentbytes.net/,
http://tracker.istole.it/.

A.8 Content blocking in organization 8

The list of blocked URLs in organization 8 is shown below:
http://3apa3a.tomsk.tw/c/cfg.bin,
http://sexual.vipzax.com/,
http://risk.vipzax.com/,
http://nude.vipzax.com/,
http://pretty-pretty.info/,
http://thepiratebay.org/,
http://xxx.com/,
http://www.avizoon.com/,
http://tiptopteens.net/,
http://teens-models.org/,
http://ourworldkids.info/,
http://youngtop.info/,
http://www.feminista.com/.

Appendix B

Service blocking test results

In this appendix the detailed information regarding service blocking in each of these organizations is given.

B.1 Service blocking test in organization 1

The results of port checking in organization 1 are given below.
It was not possible to connect to remote FTP servers (port 21).
It was not possible to connect to remote SSH server (port 22).
It was not possible to connect to remote SMTP servers (port 25).
It was not possible to connect to remote DNS servers (port 53).
It was possible to connect to remote HTTP servers (port 80).
It was not possible to connect to remote POP3 servers (port 110).
It was not possible to connect to remote RPC servers (port 135).
It was not possible to connect to remote NetBIOS servers (port 139).
It was not possible to connect to remote IMAP servers (port 143).
It was not possible to connect to remote SNMP servers (port 161).
It was possible to connect to remote HTTPS servers (port 443).
It was not possible to connect to remote SMB servers (port 445).
It was not possible to connect to remote SMTP/SSL servers (port 465).
It was possible to connect to remote secure IMAP servers (port 585).

It was possible to connect to remote authenticated SMTP servers (port 587).
It was possible to connect to remote IMAP/SSL servers (port 993).
It was possible to connect to remote POP/SSL servers (port 995).
It was possible to connect to remote OpenVPN servers (port 1194).
It was possible to connect to remote PPTP Control servers (port 1723).
It was possible to connect to remote SIP servers (port 5060).
It was possible to connect to remote BitTorrent servers (port 6881).
It was possible to connect to remote TOR server (port 9001).

B.2 Service blocking test in organization 2

The results of service blocking in organization 2 are illustrated as follows:
It was possible to connect to remote FTP servers (port 21).
It was possible to connect to remote SSH servers (port 22).
It was possible to connect to remote SMTP servers (port 25).
It was possible to connect to remote DNS servers (port 53).
It was possible to connect to remote HTTP servers (port 80).
It was possible to connect to remote POP3 servers (port 110).
It was possible to connect to remote RPC servers (port 135).
It was possible to connect to remote NetBIOS servers (port 139).
It was possible to connect to remote IMAP servers (port 143).
It was possible to connect to remote SNMP servers (port 161).
It was possible to connect to remote HTTPS servers (port 443).
It was possible to connect to remote SMB servers (port 445).
It was possible to connect to remote SMTP/SSL servers (port 465).
It was possible to connect to remote secure IMAP servers (port 585).
It was possible to connect to remote authenticated SMTP servers (port 587).
It was possible to connect to remote IMAP/SSL servers (port 993).
It was possible to connect to remote POP/SSL servers (port 995).

It was possible to connect to remote OpenVPN servers (port 1194).
It was possible to connect to remote PPTP Control servers (port 1723).
It was possible to connect to remote SIP servers (port 5060).
It was possible to connect to remote BitTorrent servers (port 6881).
It was possible to connect to remote TOR servers (port 9001).

B.3 Service blocking test in organization 3

The list of blocked and open ports in organization 3 is given below:
It was possible to connect to remote FTP servers (port 21).
It was possible to connect to remote SSH servers (port 22).
It was not possible to connect to remote Direct TCP access to remote SMTP servers (port 25).
It was not possible to connect to remotee DNS servers (port 53).
It was possible to connect to remote HTTP servers (port 80).
It was possible to connect to remote POP3 servers (port 110).
It was not possible to connect to remote RPC servers (port 135).
It was not possible to connect to remote NetBIOS servers (port 139).
It was possible to connect to remote IMAP servers (port 143).
It was possible to connect to remote SNMP servers (port 161).
It was possible to connect to remote HTTPS servers (port 443).
It was not possible to connect to remote Direct TCP access to remote SMB servers (port 445).
It was possible to connect to remote SMTP/SSL servers (port 465).
It was possible to connect to remote secure IMAP servers (port 585).
It was possible to connect to remote authenticated SMTP servers (port 587).
It was possible to connect to remote IMAP/SSL servers (port 993).
It was possible to connect to remote POP/SSL servers (port 995).
It was possible to connect to remote OpenVPN servers (port 1194).

It was possible to connect to remote PPTP Control servers (port 1723).
It was possible to connect to remote SIP servers (port 5060).
It was possible to connect to remote BitTorrent servers (port 6881).
It was possible to connect to remote TOR servers (port 9001).

B.4 Service blocking test in organization 4

The result of service blocking test in organization 4 is given below.
It was possible to connect to remote FTP servers (port 21).
It was possible to connect to remote SSH servers (port 22).
It was possible to connect to remote SMTP servers (port 25).
It was not possible to connect to remote DNS servers (port 53).
It was possible to connect to remote HTTP servers (port 80).
It was possible to connect to remote POP3 servers (port 110).
It was not possible to connect to remote RPC servers (port 135).
It was not possible to connect to remote NetBIOS servers (port 139).
It was possible to connect to remote IMAP servers (port 143).
It was not possible to connect to remote SNMP servers (port 161).
It was possible to connect to remote HTTPS servers (port 443).
It was possible to connect to remote SMB servers (port 445).
It was possible to connect to remote SMTP/SSL servers (port 465).
It was possible to connect to remote secure IMAP servers (port 585).
It was possible to connect to remote authenticated SMTP servers (port 587).
It was possible to connect to remote IMAP/SSL servers (port 993).
It was possible to connect to remote POP/SSL servers (port 995).
It was possible to connect to remote OpenVPN servers (port 1194).
It was possible to connect to remote PPTP Control servers (port 1723).
It was possible to connect to remote SIP servers (port 5060).
It was possible to connect to remote BitTorrent servers (port 6881).

It was possible to connect to remote TOR servers (port 9001).

B.5 Service blocking test in organization 5

The result of port checking in organization 5 is given below.
It was possible to connect to remote FTP servers (port 21).
It was possible to connect to remote SSH servers (port 22).
It was possible to connect to remote SMTP servers (port 25).
It was possible to connect to remote DNS servers (port 53).
It was possible to connect to remote HTTP servers (port 80).
It was possible to connect to remote POP3 servers (port 110).
It was possible to connect to remote RPC servers (port 135).
It was possible to connect to remote NetBIOS servers (port 139).
It was possible to connect to remote IMAP servers (port 143).
It was possible to connect to remote SNMP servers (port 161).
It was possible to connect to remote HTTPS servers (port 443).
It was possible to connect to remote SMB servers (port 445).
It was possible to connect to remote SMTP/SSL servers (port 465).
It was possible to connect to remote secure IMAP servers (port 585).
It was possible to connect to remote authenticated SMTP servers (port 587).
It was possible to connect to remote IMAP/SSL servers (port 993).
It was possible to connect to remote POP/SSL servers (port 995).
It was possible to connect to remote OpenVPN servers (port 1194).
It was possible to connect to remote PPTP Control servers (port 1723).
It was possible to connect to remote SIP servers (port 5060).
It was possible to connect to remote BitTorrent servers (port 6881).
It was possible to connect to remote TOR servers (port 9001).

B.6 Service blocking test in organization 6

The result of port checking in organization 6 is given below.
It was possible to connect to remote FTP servers (port 21).
It was possible to connect to remote SSH servers (port 22).
It was not possible to connect to remote SMTP servers (port 25).
It was possible to connect to remote DNS servers (port 53).
It was possible to connect to remote HTTP servers (port 80).
It was possible to connect to remote POP3 servers (port 110).
It was possible to connect to remote RPC servers (port 135).
It was possible to connect to remote NetBIOS servers (port 139).
It was possible to connect to remote IMAP servers (port 143).
It was possible to connect to remote SNMP servers (port 161).
It was possible to connect to remote HTTPS servers (port 443).
It was possible to connect to remote SMB servers (port 445).
It was possible to connect to remote SMTP/SSL servers (port 465) .
It was possible to connect to remote secure IMAP servers (port 585).
It was possible to connect to remote authenticated SMTP servers (port 587).
It was possible to connect to remote IMAP/SSL servers (port 993).
It was possible to connect to remote POP/SSL servers (port 995).
It was possible to connect to remote OpenVPN servers (port 1194).
It was possible to connect to remote PPTP Control servers (port 1723).
It was possible to connect to remote SIP servers (port 5060).
It was possible to connect to remote BitTorrent servers (port 6881).
It was possible to connect to remote TOR servers (port 9001).

B.7 Service blocking test at organization 7

The result of port checking at organization 7 is given below.
It was not possible to connect to remote FTP servers (port 21).
It was not possible to connect to remote SSH servers (port 22).
It was possible to connect to remote SMTP servers (port 25).
It was possible to connect to remote DNS servers (port 53).
It was possible to connect to remote HTTP servers (port 80).
It was possible to connectto remote POP3 servers (port 110).
It was not possible to connect to remote RPC servers (port 135).
It was not possible to connect to remote NetBIOS servers (port 139).
It was possible to connect to remote IMAP servers (port 143).
It was not possible to connect to remote SNMP servers (port 161).
It was possible to connect to remote HTTPS servers (port 443).
It was not possible to connect to remote SMB servers (port 445).
It was not possible to connect to remote SMTP/SSL servers (port 465) .
It was not possible to connect to remote secure IMAP servers (port 585).
It was not possible to connect to remote authenticated SMTP servers (port 587).
It was possible to connect to remote IMAP/SSL servers (port 993).
It was not possible to connect to remote POP/SSL servers (port 995).
It was not possible to connect to remote OpenVPN servers (port 1194).
It was not possible to connect to remote PPTP Control servers (port 1723).
It was not possible to connect to remote SIP servers (port 5060).
It was not possible to connect to remote BitTorrent servers (port 6881).
It was not possible to connect to remote TOR servers (port 9001).

B.8 Service blocking test in organization 8

The results of service blocking in organization 8 are given below:
It was possible to connect to remote FTP servers (port 21).
It was possible to connect to remote SSH servers (port 22).
It was possible to connect to remote SMTP servers (port 25).
It was possible to connect to remote DNS servers (port 53).
It was possible to connect to remote HTTP servers (port 80).
It was possible to connect to remote POP3 servers (port 110).
It was possible to connect to remote RPC servers (port 135).
It was possible to connect to remote NetBIOS servers (port 139).
It was possible to connect to remote IMAP servers (port 143).
It was possible to connect to remote SNMP servers (port 161).
It was possible to connect to remote HTTPS servers (port 443).
It was possible to connect to remote SMB servers (port 445).
It was possible to connect to remote SMTP/SSL servers (port 465).
It was possible to connect to remote secure IMAP servers (port 585).
It was possible to connect to remote authenticated SMTP servers (port 587).
It was possible to connect to remote IMAP/SSL servers (port 993).
It was possible to connect to remote POP/SSL servers (port 995).
It was possible to connect to remote OpenVPN servers (port 1194).
It was possible to connect to remote PPTP Control servers (port 1723).
It was possible to connect to remote SIP servers (port 5060).
It was possible to connect to remote BitTorrent servers (port 6881).
It was possible to connect to remote TOR servers (port 9001).

Appendix C

Human ethics form

VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wananga o te Upoko o te Ika a Maui

HUMAN ETHICS COMMITTEE
Application for Approval of Research Projects

Please write legibly or type if possible. Applications must be signed by supervisor (for student projects) and Head of School

Note: The Human Ethics Committee attempts to have all applications approved within three weeks but a longer period may be necessary if applications require substantial revision.

1 NATURE OF PROPOSED RESEARCH:

(a) Staff Research ☐	Student Research ☒	(tick one)
(b) If Student Research	Degree Master of Computer Science	Course Code COMP591
(c) Project Title: Invisible barriers		

2 INVESTIGATORS:

(a) Principal Investigator	
Name	Shadi Esnaashari
e-mail address	esnaasshad@myvuw.ac.nz
School/Dept/Group	School of Engineering and Computer Science

(b) Other Researchers Name	Position

(c) Supervisor (in the case of student research projects)	
Ian Welch and Brenda Chawner	

3 DURATION OF RESEARCH

(a) Proposed starting date for data collection May. 2013
(Note: that NO part of the research requiring ethical approval may commence prior to approval being given)
(b) Proposed date of completion of project as a whole SEP.2013

4 PROPOSED SOURCE/ S OF FUNDING AND OTHER ETHICAL CONSIDERATIONS

(a) Sources of funding for the project

Please indicate any ethical issues or conflicts of interest that may arise because of sources of funding e.g. restrictions on publication of results

NA

(b) Is any professional code of ethics to be followed **Y☒ N☐**

If yes, name ACM code of Ethics

(c) Is ethical approval required from any other body **Y☐ N☒**

If yes, name and indicate when/if approval will be given

5 DETAILS OF PROJECT

Briefly Outline:

(a) The objectives of the project

Most organisations monitor and/or control use of the Internet through the use of logs and/or Internet filters. This monitoring/control is usually invisible, and staff are only aware of it if they try to view a website or use a service that is blocked. However, there is little information about what is being blocked, and the reasons for that. The goal of this project is to identify the types of sites and services that are being blocked, in different types of organisations, the reasons for the blocking, and the effects of this blocking. The project will also identify the extent to which Internet traffic is monitored and the ways in which these records are used. The project's goal is to have a more informed policy debate about the need to block access to Internet websites and services, and its impact on people's ability to access information they need.

b) Method of data collection

There are different places in New Zealand which offer wireless internet access (Wifi) either free or paid to their customers such as CBD free Internet which covers different places such as Newtown Library Zoo, Central Library, Water Front Zone, Wellington Airport and the Airport Flyer buses, Te Papa Museum, McDonnalds, Starbucks, and many small and independent cafes and restaurants are also providing free WiFi hotspots.
They claim that they are offering free Internet. By using a modified version of the Open Observatory for Network Interference (OONI) tool that I will run at their sitel, these networks will be investigated and the list of block URLs and services will be prepared for all of these locations in order to make it transparent for users.
After observation, managers of those specific organizations which offer free Internet will be interviwed for the reasons behind content blocking.
I have a plan to collect data for 3 months. I will interview the managers of the organizations for 1

month. After gaining their consent I will audio record their voice.

(c) The benefits and scientific value of the project

No-one in NewZealand has researched the restrictions in place on users of either paid or free Wifi services nor on the policies used by organizations.

(d) Characteristics of the participants

The focal participants in this study will be the responsible people for implementing the content blocking in those specific organizations.

(e) Method of recruitment

In the first part of my data collection, I will run the OONI software on-site. The sites will be selected by compiling a list of free and paid Wifi providers by consulting sources such as the YellowPages and directories of Wifi providers in Wellington. I will run the project by myself. The main benefits of running the experiment myself is that I do not expose any third-parties to risks associated with accessing potentially blocked content and I can monitor the running of the tests directly to determine if there are any security or other risks that occur during the testing.
In the second part, I will contact the people responsible for the management of the service. This will be done in person at the businesses by providing a letter containing an information sheet and if there is no customer representative, I will use resources such as the email contact details provided by the Wifi provider and/or use resources such as the Companies Directory to source contacts details for the owners of the companies providing the service.

A draft of the email is as follows.
Dear Sir/Madam,
I am a Masters student in Computer Science at Victoria University of Wellington. I am investigating the prevalance of content blocking on the Internet in some NZ organizations including yours. I have found a few blocked contents in your organization. I would like to invite you to participate in an interview and answer a few questions. The interview will take 30 minutes.
Thank you very much in advance. I will be looking forward to hearing from you soon.
Regards,
Shadi Esnaashari
Postgraduate student
School of Engineering and Computer Science
Victoria University of Wellington,
NewZealand

(f) Payments that are to be made/expenses to be reimbursed to participants

N/A

(g) Other assistance (e.g. meals, transport) that is to be given to participants

N/A

(h) Any special hazards and/or inconvenience (including deception) that participants will encounter

There are two main risks/issues to consider:

1. We intend to carry out the testing first and inform the organisations afterwards. The rationale is that we are: (a) simply carrying out activities that an ordinary user might carry out; and, (b) if organisations are aware of our experiments ahead of time we may not be able to recruit participants and we are also interested in the difference between the perceptions of organisations with respect to their beliefs about what they restrict and the realities.

2. The URLs used to test for blocking will be sourced from search engines where the URLs are generated for different categories - News, Banking, Adult sites, Hackings sites, Entertainment etc. We will source these by using search terms associated with these categories and using URLs returned by the search engine (with safe filtering disabled). These URLs are auto-generated and there is a possibility that the URLs may contain links currently blocked or monitored by the Department of Internal Affairs (DIA). It is impossible to check this ahead of time because do not allow individuals or organisations to check which URLs may or may not be monitored at a national level.

3. After running the testing, I will provide the information sheet for the managers to ask for participation

(i) State whether consent is for:

(i)	the collection of data	Y☒	N☐
(ii)	attribution of opinions or information	Y☐	N☒
(iii)	release of data to others	Y☐	N☒
(iv)	use for a conference report or a publication	Y☒	N☐
(v)	use for some particular purpose (specify)	Y☐	N☒

Attach a copy of any questionnaire or interview schedule to the application

(j) How is informed consent to be obtained (see sections 4.1, 4.5(d) and 4.8(g) of the Human Ethics Policy)

(i) the research is strictly anonymous, an information sheet is supplied and informed consent is implied by voluntary participation in filling out a questionnaire for example (include a copy of the information sheet) Y☐ N☒

(ii) the research is not anonymous but is confidential and informed consent will be obtained through a signed consent form (include a copy of the consent form and information sheet) Y☒ N☐

(iii) the research is neither anonymous or confidential and informed consent will be obtained through a signed consent form (include a copy of the consent form and information sheet) Y☐ N☒

(iv) informed consent will be obtained by some other method (please specify and provide details) Y☐ N☒

n/a

With the exception of anonymous research as in (i), if it is proposed that written consent will not be obtained, please explain why

n/a

(k) If the research will not be conducted on a strictly anonymous basis state how issues of confidentiality of participants are to be ensured if this is intended. (See section 4.1(e) of the Human Ethics Policy). (e.g. who will listen to tapes, see questionnaires or have access to data). Please ensure that you distinguish clearly between anonymity and confidentiality. Indicate which of these are applicable.

(i) access to the research data will be restricted to the investigator Y☐ N☒

(ii) access to the research data will be restricted to the investigator and their supervisor (student research) Y☒ N☐

(iii) all opinions and data will be reported in aggregated form in such a way that individual persons or organisations are not identifiable Y☒ N☐

(iv) Other (please specify)

(l) Procedure for the storage of, access to and disposal of data, both during and at the conclusion of the research. (see section 4.12 of the Human Ethics Policy). Indicate which are applicable:

(i) all written material (questionnaires, interview notes, etc) will be kept in a locked file and access is restricted to the investigator Y☒ N☐
(ii) all electronic information will be kept in a password-protected file and access will be restricted to the investigator Y☒ N☐
(iii) all questionnaires, interview notes and similar materials will be destroyed:
(a) at the conclusion of the research Y☐ N☒
or (b) 3 years after the conclusion of the research Y☒ N☐
(iv) any audio or video recordings will be returned to participants and/or electronically wiped Y☒ N☐
(v) other procedures (please specify):

n/a

If data and material are not to be destroyed please indicate why and the procedures envisaged for ongoing storage and security

n/a

(m) Feedback procedures (See section 7 of Appendix 1 of the Human Ethics Policy). You should indicate whether feedback will be provided to participants and in what form If feedback will not be given, indicate the reasons why.

The participants will be given an option to receive the results of the study,

(n) Reporting and publication of results. Please indicate which of the following are appropriate. The proposed form of publications should be indicated on the information sheet and/or consent form

(i) publication in academic or professional journals Y☒ N☐
(ii) dissemination at academic or professional conferences Y☒ N☐
(iii) deposit of the research paper or thesis in the University Library (student research) Y☒ N☐
(iv) other (please specify)

7

Signature of investigators as listed on page 1 (including supervisors) and Head of School.

NB: <u>All investigators and the Head of School must sign before an application is submitted for approval</u>

	Date	
	Date	
	Date	

Head of School:

	Date	

Appendix D

Information for managers

School of Engineering and Computer Science

Invisible barriers

Information sheet for managers

Researcher: Shadi Esnaashari

I am a Master's student in the School of Engineering and Computer Science at Victoria University of Wellington. As a requirement of this degree, I am doing a research study which leads to a Master's thesis. In my research, I am exploring what restrictions are placed on accessing websites and Internet services, and how these affect users. I would like to invite you to participate in this research study.

I have used your service twice this month and found that a number of URLs and/ or services are blocked. I would like to interview you and ask you to discuss the reasons for implementing censorship on those specific URLs and services. The interview will take no more than 30 minutes.

Participation in this study is voluntary. If you agree to participate, you can ask me to remove from the study any of your recorded talk for any reason. Also, you may choose to withdraw from the study at any stage until two weeks after the data collection is complete. If you decide to do so, you can contact me and any data you have provided will be removed from the study.

In my reports on this research, I will use pseudonyms for you in the study in order to keep your identities and responses confidential. Audio will be destroyed or returned to you after the research is finished. Written materials such as field notes and transcripts will be destroyed three years after the completion of my research. This allows for publication of the research in academic journals.

I will report on the research in a thesis which will be submitted to the University and deposited in the University Library.

This research has been approved by the Human Ethics Committee of Victoria University of Wellington.

If you have any further enquiries regarding this research, please contact me or my two supervisors Dr. Ian Welch (email: **ian.welch@ vuw.ac.nz**) and Dr. **Brenda Chawner** (email **brenda.chawner@ vuw.ac.nz**) at the School of Engineering and Computer Science and School of Information Management, Victoria University of Wellington.

....

Shadi Esnaashari Date

(**email:** esnaasshad@myvuw.ac.nz)

Appendix E

Consent form

School of Engineering and Computer Science

Consent Form for managers

Name of person: ……..

I have read the information sheet regarding this research and have had an opportunity to ask any questions about the research and have them answered to my satisfaction.

I understand that:

- taking part in this research is voluntary.
- I can withdraw from the study at any time until two weeks after the data collection is complete i.e. August 30, 2013.
- Shadi will keep the information I give confidential.
- no one except Shadi and her two supervisors will have access to the data.
- pseudonyms will be chosen to ensure confidentiality.
- the information I give will be used by Shadi to investigate the reasons for blocking content in different organizations.
- **participants will have access to the recordings and could remove any part they deem inappropriate.**
- **participants will be informed about the finding of my research via email.**

  Please write down your email, if you would like to receive a copy of the findings of this research.
  Email: ……

- the recordings will be destroyed by Shadi three years after this research is finished.

If I have any further concerns which require more explanation, I can contact Shadi or her supervisors Dr. Ian Welch (email: **ian.welch@ vuw.ac.nz**) and Dr. **Brenda Chawner** (email **brenda.chawner@ vuw.ac.nz**) at the School of Engineering and Computer Science and School of Information Management, Victoria University of Wellington.

I agree to take part in this research.

Signature of Manager……
Date……

Appendix F

Email to ask for interview

The following email has been used to recruit the interviewees:
Dear Sir/Madam,
I am a Masters student in Computer Science at Victoria University of Wellington. I am investigating the prevalence of content blocking on the Internet in some NZ organizations including yours. I have found a few blocked contents in your organization. I would like to invite you to participate in an interview and answer a few questions. The interview will take 30 minutes.
Regards,
Shadi Esnaashari
Postgraduate student
School of Engineering and Computer Science
Victoria University of Wellington, New Zealand

Appendix G

Interview questions

The following questions have been used in the interviwes:
1. Please tell me about your main reasons for blocking access.
2. Would you maintain a list of blocked sites/services by your organizations?
3. How often would you look at them and revise them?
4. Could you tell me a little about the categorizations? Please tell me who has chosen these categorizations ? How have they chosen these categories?
5. Are you aware that whether you have over-blocking or not?
6. Why are some dating websites open and some others closed? Are there any people to check the lists regularly?
7. Have you had much feedback about blocking access to these sites/services?
8. To what extent does your organization block access to Web sites or online services such as ftp, ssh, and irc, etc?
9. Would you consider if people asked for unblocking some URLs? How much time would it take to make decisions for unblocking?
10. Are staff in your organisation/users of your free WiFi aware that their access to these sites/services is blocked? If not, why dont you make this information available to them?
11. Have you considered any other types of solutions rather than imple-

menting blocking?
12. Do you know that with blocking you prevent users from accessing useful information?
13. Considering that people have access to 3G and the fact that they can access the desired content, what is the reason for blocking?
14. Why would you not publish the list? If someone tried to access these blocked websites, would there be any punishment for that?
15. Would you like to talk about anything else?

Bibliography

[1] 403 checker - global voices advocacy. http://advocacy.globalvoicesonline.org/projects/403checker/, [ONLINE], [Accessed 12 November 2012].

[2] About filtering, opennet initiative. http://opennet.net/about-filtering, [ONLINE], [Accessed 11 November 2012].

[3] About netsafe. http://www.netsafe.org.nz/about-netsafe/, [ONLINE], [Accessed 04 November 2012].

[4] About oni, opennet initiative. http://opennet.net/about-oni., [ONLINE], [Accessed 11 November 2012].

[5] About tech liberty, tech liberty nz. http://techliberty.org.nz/about-tech-liberty/, [ONLINE], [Accessed 04 November 2012].

[6] Alkasir, for mapping and circumventing cyber-censorship. https://alkasir.com/, [ONLINE], [Accessed 12 November 2012].

[7] Australia and new zealand. https://opennet.net/research/australia-and-new-zealand, [ONLINE], [Accessed 11 November 2012].

[8] The average web page loads in 2.45 seconds google reveals. http://news.softpedia.com/news, [ONLINE], [Accessed 02 May 2013].

[9] Blue coat. http://www.packeteer.com/., [ONLINE], [Accessed 18 December 2012].

[10] [chapter 6] packet filtering. http://www.diablotin.com/librairie/networking/firewall/ch06_01.htm., [ONLINE], [Accessed 07 November 2012].

[11] Classification and the internet: Classification in nz: Office of film and literature classification. http://www.censorship.govt.nz/about-censorship/classification-and-the-internet.html., [ONLINE], [Accessed 13 August 2013].

[12] Content blocking — electronic frontier foundation. https://www.eff.org/issues/content-blocking./, [ONLINE], [Accessed 09 March 2014].

[13] Dansguardian - true web content filtering for all. http://dansguardian.org/, [ONLINE], [Accessed 13 August 2013].

[14] The department of internal affairs te tari taiwhenua. http://www.dia.govt.nz/, [ONLINE], [Accessed 04 November 2012].

[15] Euthanasia activist wants new zealand website blacklist released - wikileaks. 2013. https://wikileaks.org/wiki/Euthanasia_activist_wants_New_Zealand_website_blacklist_released., [ONLINE], [Accessed 01 July 2013].

[16] Faq, tech liberty nz. http://techliberty.org.nz/issues/internet-filtering/filtering-faq/., [ONLINE], [Accessed 04 November 2012].

[17] Firewalls, internet security, corporate firewall - vicomsoft. http://www.vicomsoft.com/learning-center/firewalls/, [ONLINE], [Accessed 12 November 2012].

[18] Freedom on the net 2012, freedom house. http://www.freedomhouse.org/report/freedom-net/freedom-net-2012, [ONLINE], [Accessed 04 November 2012].

[19] Freegate - cnet download.com. http://download.cnet.com/Freegate/3000-2085_4-10415391.html., [ONLINE], [Accessed 20 December 2012].

[20] Google trends - hot searches. http://google.com/trends/, [ONLINE], [Accessed 20 June 2013].

[21] Herdict : Home. http://www.herdict.org/, [ONLINE], [Accessed 12 November 2012].

[22] Household access to the internet - statistics new zealand. http://www.stats.govt.nz/browse_for_stats/people_and_communities/households/household-access-to-the-internet.aspx., [ONLINE], [Accessed 01 July 2013].

[23] How to bypass internet censorship. http://en.flossmanuals.net/bypassing-censorship/ch010_simple-tricks/, [ONLINE], [Accessed 20 December 2012].

[24] How to bypass internet censorship with vpn. http://techandscience.com/techblog/ShowArticle.aspx?ID=1610., [ONLINE], [Accessed 20 December 2012].

[25] Icsi netalyzr. http://netalyzr.icsi.berkeley.edu/, [ONLINE], [Accessed 11 November 2012].

[26] Inferring mechanics of web censorship around the world, usenix. https://www.usenix.org/conference/foci12/inferring-mechanics-web-censorship-around-world., [ONLINE], [Accessed 04 November 2012].

[27] International visitor arrivals to new zealand - statistics new zealand. 2014. http://www.stats.govt.nz/browse_for_stats/population/Migration/iva.aspx., [ONLINE], [Accessed 13 August 2013].

[28] Internet filtering, internetnz. https://internetnz.net.nz/our-work/openness/internet-filtering., [ONLINE], [Accessed 13 August 2013].

[29] Internet filtering, internetnz. http://internetnz.net.nz/our-work/openness/internet-filtering., [ONLINE], [Accessed 04 November 2012].

[30] Internet service provider survey - statistics new zealand. http://www.stats.govt.nz/surveys_and_methods/completing-a-survey/faqs-about-our-surveys/internet-service-provider-survey.aspx., [ONLINE], [Accessed 19 December 2012].

[31] Internetnz. 2012. home. http://internetnz.net.nz/, [ONLINE], [Accessed 04 November 2012].

[32] Internetnz rejects centrally operated filtering for new zealand, internetnz. http://internetnz.net.nz/media-releases-2010, [ONLINE], [Accessed 04 November 2012].

[33] Issues that should be made understandable to the general public. http://www.wikileaks.org/Press.html, [ONLINE], [Accessed 02 March 2013].

[34] Net neutrality - issues - cisco systems . 2014. http://www.cisco.com/web/about/gov/issues/net_neutrality.html., [ONLINE], [Accessed 10 March 2014].

[35] Netclean whitebox trial in new zealand. http://www.watchdoginternational.net/index.php/watchdog-news/watchdog-news/104-netclean-whitebox-trial-in-new-zealand., [ONLINE], [Accessed 04 November 2012].

[36] Neubot, the network neutrality bot. http://www.neubot.org/, [ONLINE], [Accessed 11 November 2012].

[37] Onion routing. http://www.onion-router.net/, [ONLINE], [Accessed 11 November 2012].

[38] Ooni. http://ooni.nu/, [ONLINE], [Accessed 11 November 2012].

[39] Open directory project. http://www.dmoz.org/, [ONLINE], [Accessed 02 June 2013].

[40] Open internet in new zealand: Status, internetnz. http://internetnz.net.nz/news/blog/2011/Open-Internet-New-Zealand-Status, [ONLINE], [Accessed 04 November 2012].

[41] The osi model's seven layers defined and functions explained. http://support.microsoft.com/kb/103884., [ONLINE], [Accessed 1 December 2012].

[42] Parental control & monitoring - products. http://www.pcmag.com/category2/0,2806,1639158,00.asp., [ONLINE], [Accessed 2 December 2012].

[43] Planet blue coat: Mapping global censorship and surveillance tools. https://citizenlab.org/2013/01/, [ONLINE], [Accessed 13 August 2013].

[44] Planetlab — an open platform for developing, deploying, and accessing planetary-scale services. 2013. https://www.planet-lab.org/., [ONLINE], [Accessed 20 December 2012].

[45] Polipo a caching web proxy. http://www.pps.univ-paris-diderot.fr/~jch/software/polipo/, [ONLINE], [Accessed 13 August 2013].

[46] Population clock - statistics new zealand. http://www.stats.govt.nz/tools_and_services/tools/population_clock.aspx., [ONLINE], [Accessed 19 December 2012].

[47] Portal: Contents/categories - wikipedia, the free encyclopedia. 2013. Http://en.wikipedia.org/wiki/Portal:Contents/Categories, [ONLINE], [Accessed 02 March 2013].

[48] Psiphon: Analysis and estimation. http://www.cdf.toronto.edu/~csc494h/reports/2004-fall/psiphon_ae.html., [ONLINE], [Accessed 20 December 2012].

[49] Risk and the internet: Perception and reality,2000. http://www.copacommission.org/papers/webriskanalysis.pdf., [ONLINE], [Accessed 02 December 2012].

[50] Sandvine : Intelligent broadband networks - welcome! http://www.sandvine.com/., [ONLINE], [Accessed 18 December 2012].

[51] Skype surveillance — electronic frontier foundation. 2014. https://www.eff.org/foia/foia-skype-surveillance., [ONLINE], [Accessed 10 March 2014].

[52] Survey on internet filtering, internetnz. https://internetnz.net.nz/news/blog/2012/Survey-Internet-Filtering, [ONLINE], [Accessed 08 August 2013].

[53] Switzerland network testing tool, electronic frontier foundation. https://www.eff.org/pages/switzerland-network-testing-tool., [ONLINE], [Accessed 11 November 2012].

[54] Te papa does not know why it is censoring the internet, tech liberty nz. http://techliberty.org.nz/te-papa-internet-censorship/, [ONLINE], [Accessed 02 June 2013].

[55] Telephone and internet access in the home - social report 2010. http://socialreport.msd.govt.nz/social-connectedness/telephone-internet-access.html., [ONLINE], [Accessed 19 December 2012].

[56] Twitter. 2013. twitter. https://twitter.com/, [ONLINE], [Accessed 20 June 2013].

[57] Ultrasurf - free proxy-based internet privacy and security tools. http://ultrasurf.us/, [ONLINE], [Accessed 20 December 2012].

[58] Welcome to project bismark. http://projectbismark.net/, [ONLINE], [Accessed 11 November 2012].

[59] Well known tcp and udp ports used by apple software products. http://support.apple.com/kb/ts1629., [ONLINE], [Accessed 20 June 2013].

[60] Wellington council wi-fi free - but with conditions - computerworld new zealand. 2013. http://www.computerworld.co.nz/article/488721/wellington_council_wi-fi_free_-_conditions/, [ONLINE], [Accessed 04 November 2012].

[61] Wellington first for free wi-fi. http://www.stuff.co.nz/technology/digital-living/5055915/Wellington-first-for-free-wi-fi., [ONLINE], [Accessed 04 November 2012].

[62] Which isps will filter?, tech liberty nz. http://techliberty.org.nz/issues/internet-filtering/which-isps-will-filter/, [ONLINE], [Accessed 13 August 2013].

[63] Why we oppose internet filtering, tech liberty nz. http://techliberty.org.nz/why-we-oppose-internet-filtering/, [ONLINE], [Accessed 08 July 2013].

[64] Win web crawler - powerful webcrawler, web spider, website extractor. http://www.winwebcrawler.com/, [ONLINE], [Accessed 13 August 2013].

[65] Win web crawler - powerful webcrawler, web spider, website extractor. http://www.winwebcrawler.com/, [ONLINE], [Accessed 12 November 2012].

[66] ALISON POWELL, A. C. Net neutrality discourses: Comparing advocacy and regulatory arguments in the united states and the united kingdom. http://www.cs.siue.edu/~wwhite/IS376/Assignments/NetNeutralityDiscourses.pdf, [ONLINE], [Accessed 19 December 2012].

[67] BACKES, M., HAMERLIK, M., LINARI, A., MAFFEI, M., TRYFONOPOULOS, C., AND WEIKUM, G. Anonymous and censorship resistant content sharing in unstructured overlays. In *Proceedings of the twenty-seventh ACM symposium on Principles of distributed computing* (New York, NY, USA, 2008), PODC '08, ACM, pp. 429–429.

[68] BRUNER, E. M. The qualitative researcher is a human observer who observes a human condition and is historically positioned and locally situated, 1993. Introduction: The ethnographic self and the personal self. Anthropology and literature, 1-26.

[69] CHESWICK, W. R., AND BELLOVIN, S. M. Firewalls and internet security: Repelling the wily hacker. In *Firewalls and Internet Security: Repelling the Wily Hacker* (1994), AT&T and Lumeta Corporation.

[70] DAINOTTI, A., SQUARCELLA, C., ABEN, E., CLAFFY, K. C., CHIESA, M., RUSSO, M., AND PESCAPÉ, A. Analysis of country-wide internet outages caused by censorship. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference* (New York, NY, USA, 2011), IMC '11, ACM, pp. 1–18.

[71] DISCHINGER, M., HAEBERLEN, A., GUMMADI, K. P., AND SAROIU, S. Characterizing residential broadband networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement* (New York, NY, USA, 2007), IMC '07, ACM, pp. 43–56.

[72] DISCHINGER, M., MISLOVE, A., HAEBERLEN, A., AND GUMMADI, K. P. Detecting bittorrent blocking. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement* (New York, NY, USA, 2008), IMC '08, ACM, pp. 3–8.

[73] DORNSEIF, M. Government mandated blocking of foreign web content. In *In von* (2003), pp. 617–648.

[74] JOHN-PAUL VERKAMP, M. G. Inferring mechanics of web censorship around the world. *In USENIX Workshop on Free and Open Communication on the Internet FOCI '12* (2012).

[75] SALTZER, J. H., REED, D. P., AND CLARK, D. D. End-to-end arguments in system design. *ACM Trans. Comput. Syst. 2*, 4 (Nov. 1984), 277–288.

[76] SFAKIANAKIS, A., ATHANASOPOULOS, E., AND IOANNIDIS, E. Censmon: A web censorship monitor. *In USENIX Workshop on Free and Open Communication on the Internet (FOCI)* (2011).

[77] SUNDARESAN, S., DE DONATO, W., FEAMSTER, N., TEIXEIRA, R., CRAWFORD, S., AND PESCAPÈ, A. Broadband internet performance: a view from the gateway. *SIGCOMM Comput. Commun. Rev. 41*, 4 (Aug. 2011), 134–145.

[78] SUNDARESAN, S., FEAMSTER, N., TEIXEIRA, R., TANG, A., EDWARDS, W. K., GRINTER, R. E., CHETTY, M., AND DE DONATO, W. Helping users shop for isps with internet nutrition labels. In *Proceedings of the 2nd ACM SIGCOMM workshop on Home networks* (New York, NY, USA, 2011), HomeNets '11, ACM, pp. 13–18.

[79] UKIC, Z., WEBER, M., SVEDEK, V., VUKOVIC, M., KATUSIC, D., AND JEZIC, G. Technical aspects of network neutrality,. *Proceedings of the 2011 11th International Conference Telecommunications (ConTEL)* (15-17 Juner 2011), 405–410.