2021

An investigation into the efficacy of URL content filtering systems

Brett Ronald Turner

Edith Cowan University

An Investigation into the Efficacy of URL Content Filtering Systems

A thesis for a dissertation submitted in partial fulfilment of the requirements for the degree of
Master of Science (Computer Security)

Brett Ronald Turner

Supervisor(s):
Assoc. Prof Mike Johnstone
Assoc. Prof Paul Haskell-Dowland
Dr Patryk Szewczyk

School of Science
Edith Cowan University
2021

Abstract

Content filters are used to restrict to restrict minors from accessing to online content deemed inappropriate. While much research and evaluation has been done on the efficiency of content filters, there is little in the way of empirical research as to their efficacy. The accessing of inappropriate material by minors, and the role content filtering systems can play in preventing the accessing of inappropriate material, is largely assumed with little or no evidence.

This thesis investigates if a content filter implemented with the stated aim of restricting specific Internet content from high school students achieved the goal of stopping students from accessing the identified material. The case is of a high school in Western Australia where the logs of a proxy content filter that included all Internet traffic requested by students were examined to determine the efficacy of the content filter.

Using text extraction and pattern matching techniques to look for evidence of access to restricted content within this study, the results demonstrate that the belief that content filtering systems reliably prevent access to restricted content is misplaced. in this study there is direct evidence of circumvention of the content filter.

This is single case study in one school and as such, the results are not generalisable to all schools or even through subsequent systems that replaced the content filter examined in this study, but it does raise the issue of the ability of these content filter systems to restrict content from high school students. Further studies across multiple schools and more complex circumvention methods would be required to identify if circumvention of content filters is a widespread issue.

I certify that this thesis does not, to the best of my knowledge and belief:
i) incorporate without acknowledgement any material previously submitted for a degree or diploma in any institution of higher education;
ii) contain any material previously published or written by another person except where due reference is made in the text of this thesis;
iii) contain any defamatory material;
Signed:

[redacted]

Brett Turner

Dated: 5th November 2020

Acknowledgements

There are many people who contributed to the culmination that is this thesis.

For those external to the process I thank the Department of Education of Western Australia for engaging with this research and allowing the data for this thesis to be collected.

From a professional level I would like to express my gratitude to my supervisors.

To Mike Johnstone, my primary supervisor, for agreeing to be my supervisor, picking me up as an orphaned Masters candidate and helping move forward towards the finish line.

Paul Haskell-Dowland for providing me useful technical feedback and pointing out the word salad I usually wrote at 4am.

Patryk Szewczyk for hammering out of me my overuse of the word “this”. I can’t thank you enough for this.

While Trish Williams wasn’t my supervisor when the end arrived and she abandoned me for greener pastures, I do need to thank her for getting me moving again after I had well and truly stalled. If it wasn’t for Trish getting the writing wheels turning and her hatred of long run-on sentences, I don’t think this thesis would ever seen the light of day.

Finally, I need to thank my wife, Brenda. For putting up with my om grumpy periods, irritation and completely random outbursts. It was all the thesis, honest!

Contents

Tables and Figures

Chapter 1 Introduction

Between 2006 to 2012, the Australian Federal Government discussed Internet content filtering as a technological means to protect minors online (Beazley, 2006; Conroy, 2007a; Coonan, 2007; Australian Government, 2008a). These discussions on Internet content filtering formed the foundation of the Australian Labor party's election platform in 2007 and upon winning office was later proposed as legislation. The media has also discussed the need to stop material considered inappropriate, this being a moral rather than a legal distinction, from being available to minors (Hamilton, 2009; McMenamin, 2009; The Detail, 2020). From 2008 until 2012, the Australian Government announced a number of plans to block access to illegal and other unwanted content in an effort to protect minors (Conroy, 2007a; Australian Government, 2008b; Conroy, 2012). Unwanted content is content that while not illegal is placed on the list as inappropriate to access by those agencies permitted to add to the list.

While the filter proposal, known as Australian Labor Party's Plan for Cyber-safety (Conroy, 2007b), generated a significant amount of information regarding the effectiveness of content filtering systems in the form of Australian Communications and Media Authority reports (ACMA, 2008). Little has been said about how effective these content filters are at preventing minors, particularly young adults, from accessing legally restricted content. Rather than looking at the efficacy of content filtering systems, a system is measured by its efficiency. That is, how well the system utilises resources while performing a specific task or meets a technical benchmark such as the number of intercepts in a given dataset (efficiency) rather than the system's ability to achieve a desired goal or outcome (efficacy). An effective system should be measured not only on these technical metrics but also on how well actual users are prevented from accessing restricted content. These types of systems have existed in public institutions since the mid-1990's and there are many ways to bypass these types of systems (Greenfield, Rickwood, & Tran, 2001; The Citizen Lab, 2007; Mou, Wu, & Atkin, 2016; Stem, 2017). This research aims to provide empirical evidence as to how effective a common low end and easily accessible content filtering, Uniform Resource Locator (URL) proxy system is at stopping students, who are also minors, from accessing restricted material.

1.1 Background

Internet content filtering as a means of protecting minors is not a new policy and has been implemented by past Australian governments in other, less invasive, ways such as the NetAlert filter which was client side and opt-in (Coonan, 2007; NAIRN, 2007). The then Minister for the Department of Broadband, Communications, and the Digital Economy (DBCDE) , Senator Stephen Conroy proposed the use of Internet Service Provider (ISP) content filtering systems to prevent access to illegal and other unwanted content as a means of protecting minors online (Australian Government, 2008a). The Australian Government commissioned a report into the feasibility of ISP level content filtering in September 2007 (Collins, Love, Landfeldt, & Coroneos, 2008). The feasibility report provided a strong emphasis on mandatory filtering. Mandatory filtering features heavily in the report which in turn reflected the then Australian Labor Government's mentioned cyber-safety election platform of mandatory ISP 'Cleanfeed' filtering (Conroy, 2007a). The Australian Government relented on the mandatory filter plan in 2012 when Australian Internet providers agreed to block the sites on the already existing blacklist maintained by the Australian Communications and Media Authority (ACMA) (Conroy, 2012). Even though the national political debate has faded, the Western Australian Department of Education still maintains central, category-based, blacklist content filtering, encourages individual schools to run their own content filtering and encourages parents to run their own content filters at home (Department of Education, 2019).

One of the discussions around content filtering was a result of the Australian Government's proposal to filter the Internet of unwanted content at the ISP level. The effectiveness and performance of URL filtering has featured in public discussions surrounding the United Kingdom's Cleenfeed proposal and how this could be used for a model for Australian mandatory content filtering (Hamilton, 2009; Malone, 2009; McMenamin, 2009; Newton, 2009). Australian Labor Party's Cleanfeed policy was originally announced in 2006 (Beazley, 2006). The proposed policy would have required all ISP's to offer an opt-out filtering service that would block content deemed prohibited by the Australian Communications and Media Authority. The stated goal of this policy was to protect minors from prohibited material. Beazley (2006) claimed that prohibited material can lead to aggression against women, child abuse and other forms of unwanted behaviour.

In 2007, a new Cyber-Safety policy was released (Conroy, 2007a), which differed significantly from the Australian Labor Party's original policy. The opt-out clause was replaced with a mandatory clause in reference to "all homes, schools and public computers that are used by Australian children". The original filter proposal was intended to filter only illegal content, as defined by the Australian Federal Government (2018), while the new list to be censored was promised to become more comprehensive and include inappropriate but legal material. The focus remained on preventing child access to inappropriate online content (Conroy, 2007a).

The Minister in charge of the DBCDE(who was) in 2008 then revised the policy in the Senate. The revised policy reinforced the mandatory filtering for all illegal content as campaigned for in the 2007 election. The revised section of the policy added all refused classification and prohibited content, in addition to illegal content, into the proposal (Australian Government, 2008b). The determination of what content was to be restricted was to be made by ACMA and, as a result, could then include material that was legal for adults to own and possess. If the material has not already been classified by the ACMA then the same schedule allows for material to be classified as restricted and acted upon accordingly (Australian Government, 1992). The Department of Broadband, Communications and the Digital Economy web site had noted that material deemed "offensive" was also to be included in this list (DBCDE, 2009). The inclusion of offensive material was the most notable public instance that mandatory filtering of legal material was proposed to be enforced on all Australian Internet connections. While the Australian Government officially dropped support for this policy in 2012 (Conroy, 2012; Paula & Rhonda, 2014) in favour of the expansion of the ACMA blacklist of illegal material, the policy has been implemented in many other countries since then (Jakub et al., 2018).

When the Broadcasting Services Act was amended in 2000 to include Internet, many schools were already running their own content filtering systems (Williams & Dillon, 1998; Department of Education, 2019), as High Schools contain the exact demographic that the proposed policies were designed to protect. What resulted was a set of online environments where content filtering regimes attempted to protect a population with elements intent on circumventing it. Some of these high schools maintained systems similar to what was proposed by the Australian Government, which had been in place for some time and had been logging information about user behaviour. These systems were Internet proxies configured to restrict access by using URL filtering. Therefore, an opportunity exists to investigate the effectiveness of these systems to achieve the goals of preventing access to restricted content.

The argument for using Internet content filtering to protect minors has existed in Australia for over a decade (Conroy, 2006; NAIRN, 2007). Various systems, from Net Alert (Coonan, 2007), the mandatory content filter proposed by Conroy (2007a) to the current content filtering agreement for Internet Service Providers (ISPs) to block the content on the Australian Communications Media Authority's (ACMA) blacklist (Conroy, 2012) have all been predicated, in part or whole, on the principle of child protection. The "child protection" mindset has transcended to the inclusion of content filtering as part of the Students Online Policy for within Western Australian public schools (Department of Education, 2008, 2019).

1.2 Significance

There is extensive literature on the nature (Hunter, 2000b; Zittrain & Edelman, 2003; Palfrey, Roberts, & Zuckerman, 2009; Stem, 2017; Al Mugni, Herdiansah, Andhika, & Ridwan, 2019) and setup of URL-based content filtering (Chou et al., 2012; Mind Chasers Inc, 2019; Frost, 2020). This research evaluated the efficacy of a network-based URL content filter as a tool for preventing minors from accessing restricted material. ACMA (2008) reported that circumvention measures were not assessed.

For schools that possess a legal duty of care towards their students known as *locus parentii*, what any reasonable parent would do, the duty of care extends beyond the physical wellbeing and includes the material they access on the Internet (Williams & Dillon, 1998; Department of Education, 2019). Content filtering systems are put in place with the belief that the technical capability of these systems is sufficient to fulfil this duty of care. While there is data concerning the efficiency of these systems there is no published research regarding measurement of the efficacy of these systems. Even when research into content filtering addresses the issue of efficacy the research looks at subjective perceptions rather than actual efficacy (Vicks, 2013). If this belief is not borne out in the implementation of these systems, then the consequences can range from a misuse of investment into an ineffective system to the possibility of exposing users to what could be classified as harm.

This thesis demonstrates that the methods used to measure the expected performance of these content filters, that is efficiency, does not align to the performance in a live environment. A better understanding of the usage and limitation of these techniques can assist in the effective deployment and use of content filters. Additionally the areas of understanding how the complexity of configuration, the lack of training and the use of systems monitoring software impacts the efficacy of content filtering devices. The results of this thesis will assist content filter developers in understanding how to better build their systems, systems administrators in how to better implement and configure content filters integrated into their networks and inform government policy on what content filters are and are not capable of.

1.3 Purpose

The purpose of this study is to evaluate the efficacy of a URL-based content filtering at preventing students, aged 11 to 17, from accessing inappropriate content. This study provides empirical quantitative analysis of an existing environment that used these techniques so conclusions can be drawn as to the efficacy of this approach for protecting minors from accessing restricted content.

Believing that a tool functions to achieve a given purpose with no or inappropriate evidence can lead to a misunderstanding of the capabilities of that tool and what outcomes are likely to be achieved. If blacklist-based content filters are not preventing minors from accessing content they are not meant to access, then this can lead to other issues in child protection and risk management.

1.4 Thesis Roadmap

This thesis begins with an overview of the political landscape that brought Internet content filtering to prominence as a tool for protecting minors. This is followed by the purpose of this thesis, to investigate the efficacy of content filters as a means to protect high school students aged 11 to 17, and the research questions that provide the foundation for this research.

Chapter 2 is a review of the literature which covers how content filtering is seen as an important mechanism for protecting minors, the different types of content filtering and how the performance of content filtering is measured. Also presented is research that highlights the effectiveness-based methods that have been, and are currently, used to evaluate content filters and observes the lack of efficacy-based approaches to the evaluation of these same content filters.

Chapter 3 is an overview of research paradigms, why the case study methodology was chosen and the design of the research approach. This chapter begins with a discussion on methodology and research approaches, from the general to some of the more commonly used methods for the computer science discipline. The chosen method for this research being as a best fit for the problem of content circumvention investigation. This is followed by research design which describes the process used to examine the data. This chapter then finishes with the details of the ethics declaration associated with this research thesis.

Chapter 4 presents the results of the investigations, commencing with a discussion of how the data was found to be formatted and how it was structured. This is followed by an explanation of how the underlying premise of blacklist-based content filters creates the opportunity for exploitation. The next section looks to see if students attempt to access restricted content. Then there is an examination to determine if students access restricted content despite the filter being in place. The following section explores how students manage to circumvent the filter rules to access restricted content. The chapter ends by answering the questions of how effective the content filter is, efficacy vs efficiency and how pervasive circumvention of the content filter was.

Chapter 5 concludes the thesis by presenting the research questions and hypotheses, the methods used to test each research question and briefly discusses the outcome of each hypothesis.

Chapter 2 Review of the Literature

There are a number of reasons for why content filtering is examined and used in the ways they are. Information that can cause harm, including mental harm, exists <citation>. As a result, the dissemination of information has long sort to be controlled. In respect towards Internet content, there has long been a drive to protect minors from content that could be harmful by controlling access to that content. This drive, in time, turned to lobbying. Lobbying in turn resulted in policy, both local and governmental. In some cases, governmental policy has turned into law or some form of industry regulation.

The use of content filters has led to the development of different mechanisms to filter content. From the location these systems are places to the mechanisms used to identify restricted content. The self-control and self-governance local systems have vs the efficiency and ease of administration network-based systems. Identifying content originated with the real-world equivalents of blacklists and whitelists and in moving towards AI recognition systems attempting to automatically recognise and categorise content that should be restricted.

Once content on the Internet was restricted, the attempts to circumvent content filtering to access the restricted content. When content filtering algorithms were simple, so were the mechanisms to evade content filtering. When simple pattern matches were all that was used, simply using the IP address instead of the server name was enough. As algorithms evolved, the use of sites that fetch the content indirectly arose. Those sites could be banned so various forms of encryption could be used to obfuscate the request, either in the URL or the request in its entirety.

What content is filtered differs from environment to environment but in Australia the eSaftey commissioner oversees the blacklist that Australian ISPs block all access to. As an extension of that list, the Department of Education and Training of Western Australia maintains a content filter and filtering policy for schools.

As a result of this technical evolution and the difficulty in gaining data of an active user base, the performance metrics for content filters have always focused on experience or efficiency. User experience has focused on how users feel how effective various content filtering solutions are. Technical measures have always focused on how accurately a given content filter can match to items in their lists, black or white, how fast matches can be made or how few errors are made in the evaluation of requests. It is this focus that has overlooked the issue of efficacy.

2.1 History of content filtering

There exists a disconnect of opinions with respect to the initial development of internet filtering although Fourie, Bothma, and Bitso (2013) say it was in the early 1990’s. What can be shown is that the issue of filtering rose to prominence in the mid 1990's with the Communications Decency Act (CDA) and later the Child Online Protection Act (COPA), both of which were judged unconstitutional breach of the first amendment of the United States of America (Hunter, 2000a; Rosenberg, 2001). China began filtering the Internet with the Temporary Regulation for the Management of Computer Information Network International Connection. Drafted in 1993, announced in 1996 and verified in 1997. This regulation was the first of a number of Internet controlling regulations implemented by China (Qiu, 1999). Since then, Internet content filtering has expanded to many and is of enough concern that the level of content filtering is tracked by a number of organisations such as Freedom House (2018). What is agreed upon is that Internet content filtering has been in use for over 30 years.

2.2 The case for content filtering

The ties of moral protectionism to censorship have led in time to efforts to protect minors from information that could be deemed harmful (Aktay, 2018). Stark (2007) states that the content filtering systems of 2006 were originally fuelled by the perceived need to protect minors from the harmful effects of pornography. Stark (2007) also notes that legislation in the USA, such as the CDA of 1996 and COPA of 1998, continued to push content filtering primarily as a method for protecting minors.

The Australian Federal Government first amended the Broadcasting Services Act in 1999 to better enable the prohibition of restricted content accessed through the Internet. This act has since been updated since then with the latest amendments including the governance of online material through Internet Service Providers (Australian Federal Government, 2018). The responsibility for the management of this list has changed since then, originally being with the ACMA and now resides with the eSaftey Commisioner (2020).

The Cyber-safety policy, originally proposed as mandatory, was an opt-out clean feed where those adults wishing not to have their connection filtered would have to consciously request that the filtering be removed. The Cyber-safety policy was the first announcement of an Australian Government imposed network-based content filtering plan. The underlying principle for this policy was drawn from British Telecom's Cleanfeed (Beazley, 2006; Coonan, 2007) which, while still in use, has been shown to be overly broad in its implementation by Johnson (2008) and Schofield (2008).

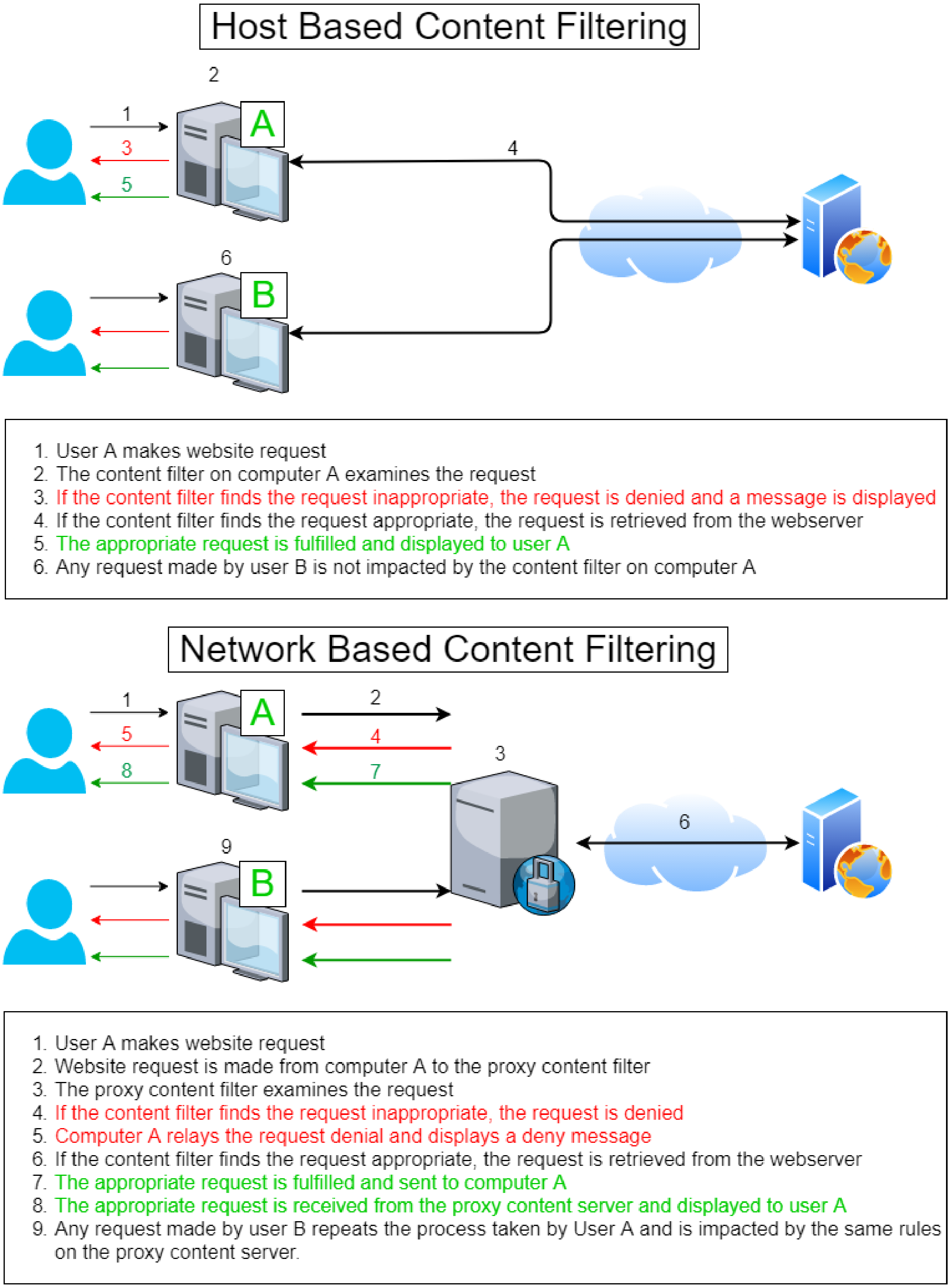
The Governing Australian coalition in 2007 made content filtering part of their "Protecting Australian Families Online" policy, and introduced a host based content filtering system (Coonan, 2007). The largest push for content filtering in Australia began not long before the 2007 election when the Australian Labor party released their Cyber-safety policy. This policy included an updated mandatory version of their previous “clean feed” policy (Conroy, 2007a). Senator Conroy made the first notable mention that the proposed filter would be mandatory in October of 2008 (Australian Government, 2008b). The announcement by Conroy (2007a) was the point in time many thought to be the first mention of the policy being mandatory (Graham, 2008). The debate on the effectiveness of such a filter has become a point of contention for supporters and opponents of the system (Lake, 2009).

When the context is switched to schools, content filtering is considered standard practice (Hills, 2018). The Department of Education (2019) policy filters the Internet at a base level based on category to reduce exposure to inappropriate content and allows for individual schools to implement additional filtering.

2.3 Web content filtering

2.3.1 How web content filters work

When content filters are used, where content is intercepted influences how many users are affected by the content filter and the impact the content filter has on performance. Content filters (figure 1) installed on a user’s computer affect only those people using that computer but can impact a computers performance and are susceptible to being disabled by a local user (Best, 2007). Network-based content filters (figure 1) allow for the application of content filtering of many users and specialised models on distributing the load of handling these requests (Lai, Ma, Yang, & Liu, 2010).



Host Based Content Filtering

1. User A makes website request
2. The content filter on computer A examines the request
3. If the content filter finds the request inappropriate, the request is denied and a message is displayed
4. If the content filter finds the request appropriate, the request is retrieved from the webserver
5. The appropriate request is fulfilled and displayed to user A
6. Any request made by user B is not impacted by the content filter on computer A

Network Based Content Filtering

1. User A makes website request
2. Website request is made from computer A to the proxy content filter
3. The proxy content filter examines the request
4. If the content filter finds the request inappropriate, the request is denied
5. Computer A relays the request denial and displays a deny message
6. If the content filter finds the request appropriate, the request is retrieved from the webserver
7. The appropriate request is fulfilled and sent to computer A
8. The appropriate request is received from the proxy content server and displayed to user A
9. Any request made by user B repeats the process taken by User A and is impacted by the same rules on the proxy content server.

Figure 1 Content filter operation

2.3.1.1 Whitelists and Blacklists

Content filtering systems can work on either a whitelist or blacklist method. A whitelist blocks all content by default unless the content is on the provided list. A blacklist allows all content by default unless the content is in the provided list. The act of filtering is achieved by denying, or blocking, content in accordance with a given list. Whitelists are restrictive and as such reduce the potential value of a resource like the Internet in an educational environment while keeping the whitelist relevant and useful requires a significant amount of work. Any World Wide Web (WWW) URL can be placed in a blacklist and the ideal outcome would be that any attempt to access a URL on the blacklist will be denied. A blacklist is more permissive that a whitelist and allows access to unaddressed material that could be useful. Given the vast number of URLs that could be filtered, a blacklist will block only a minute portion of the WWW and as such can prove ineffective against preventing access to categories of data instead of specific URLs.

The general types of content filtering systems range from rather simplistic key word approaches to more complex systems such as advanced URL filtering (Greenfield et al., 2001; Clayton, 2006). Key word systems look for the presence of a word or pattern of words that matches a predefined list (Greenfield et al., 2001; Palfrey et al., 2009). If one keyword or pattern is found the content is blocked, sometimes regardless of the context in which it was found or used. These keyword systems also are unable to check pictures which significantly reduces their usefulness (Greenfield et al., 2001; Palfrey et al., 2009; Ayre, 2012). This method, for example, will block all useful sex education material along with all other sex-related material. It will not, however, block any sex-related images. Due to the rigidity of this type of method, it has significant limitations (Narayanan, Moses, & Nirmala, 2018).

Internet Protocol (IP) packet filtering/dropping content filtering systems are a slightly more sophisticated filtering system than keyword filters. This type of system works by maintaining a list of IP addresses for hosts that contain material to be either permitted or denied. They are able to distinguish between specific Transmission Control Protocol (TCP) or User Datagram Protocol (UDP) ports but since they only deal with IP traffic, they are unable to distinguish between different web pages or even different web sites hosted on the same server (Clayton, 2006). This results in a situation where if a web server hosting multiple sites is blacklisted because of one website, all other websites on the same web server are also blocked.

Domain Name Service (DNS) blocking is where the Domain Name Server that resolves the URL into an IP Address is used to give the IP Address of a server other than the one the URL is supposed to represent. This method also suffers from blocking content that is not considered restricted but is somehow matched to a keyword or listed site. This result is known as overblocking. Since this method also does not deal with specific content, just preventing access to the host server, it is unable to deal with any contextual basis or even unrelated material hosted on the same server accessed with the same domain name. (Clayton, 2006; Reis, Godinho de Matos, & Ferreira, 2017) Many websites hosted on the same server now have different domain names reducing the impact of this limitation. Though easily circumvented by changing to a DNS that isn't poisoned this method has very little overhead making it cheap and quick.

2.3.1.2 URL Blocking

URL content filtering systems based on blacklists and whitelists allow more specific control than the IP-based or DNS-based systems. The list they use consist of universal resource locators and as such this type of system is also known as URL filtering. The system maintains a list of sites to either deny (blacklist) or a list of sites to be allowed (whitelist). They are capable of being far more granular than the previous methods as they can block access to a host, a web site, a web page or even a single component of a webpage such as a single image. For this type of approach to be fully effective it must be regularly updated. This maintenance can be manual on simpler systems (Greenfield et al., 2001; Collins et al., 2008) or on more modern systems this process can be automated by machine learning algorithms. (Lai et al., 2010)

There are systems (e.g., Web Sense, Netbox Blue and Netsweeper) that block access based on the category of the site. These categories are defined using a set of, usually externally, maintained URLs that conform to a specific category as deemed by the classifier. Category based filtering allows a client to select broad categories to be filtered using a more generalised process than simple URL filtering. This approach is not as granular as filtering individual URLs but it is simpler to use (Forte, Souza, & Prado, 2006).

All the aforementioned methods require manual intervention at some point unlike dynamic analysis, which is a different method. Dynamic filtering is dealt with as an extension of filtering technologies. In general, dynamic content filtering makes use of computing algorithms to determine whether something is likely to be of an undesired nature. (Greenfield et al., 2001; Collins et al., 2008).

Deep packet inspection (DPI) is an umbrella label given to several different techniques commonly used to refer to the inspection of the data payload of packets at the application layer. This allows for more refined evaluations of the data passing through the content filter. This method has several significant issues. As detection techniques progress up the OSI model (figure 2) it takes significantly more resources (CPU, RAM, time) to examine the data. Some of the more advanced techniques require the assembly of data before being able to dynamically categorise it. While there are advances being made to reduce the latency DPI introduces (Trabelsi, Zeidan, & Masud, 2016) these processes are always more resource intensive than in simpler mechanisms such as DNS poisoning. This translates into equipment that costs more, is more difficult to configure and maintain and introduces more latency than other methods.

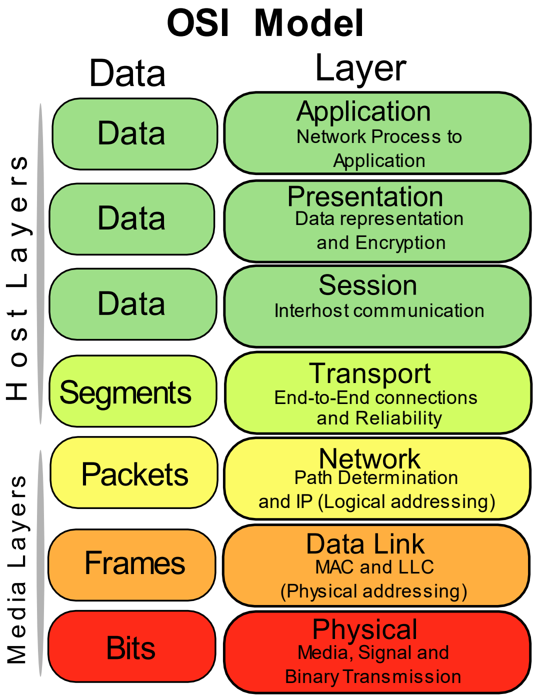

OSI Model
Data
Layer
Host Layers
Media Layers
Data
Application
Network Process to Application
Data
Presentation
Data representation and Encryption
Data
Session
Interhost communication
Segments
Transport
End-to-End connections and Reliability
Packets
Network
Path Determination and IP (Logical addressing)
Frames
Data Link
MAC and LLC (Physical addressing)
Bits
Physical
Media, Signal and Binary Transmission

Figure 2 OSI Network Model

2.3.1.3 AI-driven content filtering

Patel et al. (2019) proposed a Neural Network Classifier that, once trained, would be able to classify objectionable content as it is encountered. The test implementation used static datasets of both objectionable and unobjectionable material with part used for training and part for testing. Even with small training datasets (1000 objectionable and 1000 non-objectionable) the Objectionable Web Filtering System performed favourably, with correct classification in 96% to 99% instances, in comparison to selected commercial products such as URLflterDB, DansGuardian, and Net Nanny. However accurate these tests may be, they remain an examination in a closed environment with no users actively seeking to access the restricted content. This is a case of measuring efficiency rather than efficacy.

Faisal and El-Kassas (2018) state that most content filters now use machine learning or Artificial Intelligence but their supporting references, Hammami, Chahir, and Chen (2005) and Polpinij, Chotthanom, Sibunruang, Chamchong, and Puangpronpitag (2006) are both proposals for machine learning based content filtering systems. Hammami et al. (2005) is a proposal using decision trees with data populated from text analysis and skin tone analysis. Polpinij et al. (2006) presents two machine learning models for contextual text analysis in Thai and English to detect pornographic sites. Both of these papers fail to mention anything regarding the prevalence of machine learning in content filtering systems. This is a case of papers on new ways of using machine learning to filter Internet content being used as examples of how prevalent machine learning is in content filtering.

There exists little literature, beyond anecdotal, of exactly how prevalent machine learning is in content filtering either by product or in number of deployments. Where measures of the performance of machine learning based content filters exist, these are measures of efficiency rather than efficacy.

2.3.2 Content Filtering Circumvention

The methods of circumventing content filtering are widely publicised and easily available (Reshet, 2015; Mou et al., 2016). There are a number of different methods available and Reshet (2015) and The Citizen Lab (2007) describe some of these as well as a basic explanation of how to use them. For URL filters there are three main methods: substitution, content redirectors and encryption.

2.3.2.1 Substitution

Substitution is using the IP address to access a host instead of its URL. Some URL filters only match what is in their list and do not perform DNS lookups. A variation of this method is to use alternate or variations of the URL to the same site that may not be blocked. The approach here is to use the specific pattern matching nature of a URL blacklist to find a URL format that is not listed in the blacklist (Greenfield et al., 2001). An example of this would be using Facebook's mobile site https://m.facebook.com instead of the main site https://facebook.com. By doing the name resolution themselves a user can bypass some of the more simplistic URL content filters (The Citizen Lab, 2007).

2.3.2.2 Redirectors

Content Redirectors is a broad name for connecting to an intermediary site that will in turn fetch the desired content, this is a type of proxy. Since the user is not connecting to the undesired site directly, the blacklist does not have it on its list (Greenfield et al., 2001; The Citizen Lab, 2007). This method includes approaches such as public proxies, cache engines, connection anonymisers and translation sites (Greenfield et al., 2001; The Citizen Lab, 2007). This method can have varying success based on the method used to inspect the packets for traces of the final URL.

2.3.2.3 Encryption

The final method for circumventing URL content filters is encrypted or tunnelled connections (Greenfield et al., 2001; The Citizen Lab, 2007). Any site that uses secure sockets layer will be able to hide the contents of the packets used to communicate. Conceptually, encryption is hiding the information in a message so those for whom the message is not intended are unable to access it (IRMA, 2019). There are several methods of using encryption to defeat content filtering systems. How those are used can depend on the complexity of the content filtering system. At the simplest level, encoding the requested URL in Base64 is enough to defeat simple pattern matching content filters since the Base64 encoded form of the URL can be enough to no longer match the blacklist entry. For those content filters that do not support Secure Sockets Layer (SSL) filtering, a secure SSL connection can prevent a content filter from examining the contents of a connection. This is in fact the point of a SSL connection and to circumvent it would require a man in the middle attack (Durumeric et al., 2017). There are some network devices, such as the Cisco Firepower series, which can perform these interceptions. A virtual private network connection to an external, unrestricted server will allow complete unfettered access as is seen from the remote server (The Citizen Lab, 2007; Sovran, Libonati, & Li, 2008; Weiss, 2009). If a URL content filtering system permits encrypted connections through it, it is ultimately vulnerable to circumvention.

2.3.2.4 VPNs

An effective form of encryption use would be a Virtual Private Network (VPN) to an external network. This would allow a host to make requests that appear to come from the external network while having encrypted traffic pass through the content filter unexamined (Molina, Gambino, & Sundar, 2019). Virtual Private Networks allow for extending a network over a third party network through encapsulation (Tomsho, 2019). The network structure of a VPN is usually a private host or network connecting to another private network with the third-party network usually being, but does not need to be, the Internet. Additionally VPNs are promoted in the use of Internet Anonymity, allowing a host to appear to be accessing the Internet from a different location and/or by encrypting the traffic to the connected network to prevent eavesdropping (Molina et al., 2019; Tomsho, 2019). In the case of the circumvention of content control, VPNs allow for changing the point from which requests appear to originate. This means, depending on how these systems are implemented, traffic can be tunnelled through content control systems or bypass them completely. For those networks that limit what egress traffic is permitted, VPN over HTTP allows for utilising HTTP traffic to connect to external networks thus side stepping the point at which a control system normally examines HTTP traffic (Keijser, 2017). Deep Packet Inspection can allow some content control systems to examine the content of individual packets and counter this approach, unless the VPN uses some sort of encryption (Panchakarla, 2019).

2.4 Classification of material

Currently Internet Service Providers restrict access to prohibited content in a blacklist as defined by the eSaftey Commisioner (2020). The blacklist is compiled from the Refused Classification list (Leitch & Warren, 2015), which includes the Interpol child pornography list (Bambauer, 2013), content classified as X18+, R18+ that does not have restricted access system or is MA15+ and general is not subject to a restricted access system (Internet Industry Association, 2008). If the content is unclassified but likely to be classified as prohibited content, it will be blocked as prohibited content (Internet Industry Association, 2008).

The Department of Education and Training in Western Australia (DET) blocks content based on category and having been identified as unsuitable for the education market, where unsuitable is defined at the discretion of the DET. The DET also acknowledges that individual schools may apply additional local filtering for sites inappropriate for their environment (Department of Education, 2019).

2.5 Content filter performance and metrics

In the initial development of content filtering systems, discussions focused towards the effectiveness of the content filtering systems. Hunter (2000a) and later Stark (2007) focused on the technical efficiency of these systems to block a pre-defined list of undesirable sites, with underblocking and overblocking (table 1) being the primary points of focus. The primary difference between these studies is who created the sample list of sites for simulated Web use.

	Blocked	Permitted
Restricted Site	Expected Behaviour	Under-Blocking
Non-restricted Site	Over-Blocking	Expected Behaviour

Table 1 Under and Over Blocking

Underblocking and overblocking (table 1) are technology-specific terms for the false positive and false negative error types (table 2). In this case a false positive would be detecting a legitimate site as a site that should be restricted and a false negative as a restricted site that should be denied.

	Measured Positive	Measured Negative
Actual Positive	True Positive	False Negative
Actual Negative	False Positive	True Negative

Table 2 True and False Positives

Underblocking (figure 3) occurs when a system allows access to some material that should be denied. The focus here is on the fundamental failure of the function of the system (Rowe & King, 2015).

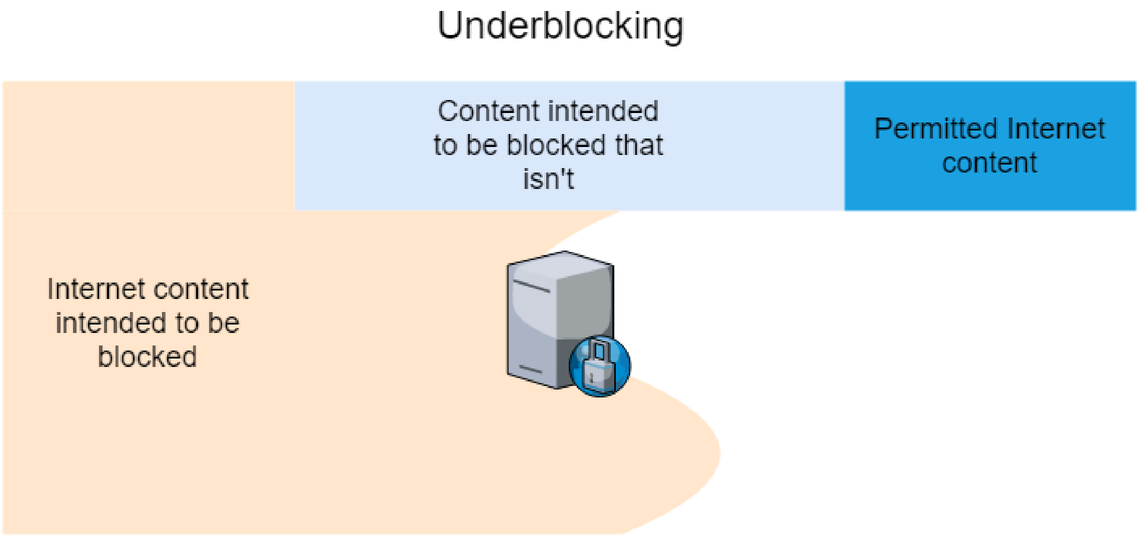

Underblocking
Content intended to be blocked that isn't
Permitted Internet content
Internet content intended to be blocked

Figure 3 Underblocking

Overblocking (figure 4) occurs when a system blocks material that should otherwise be permitted (Rowe & King, 2015). There are arguments for and against both underblocking and overblocking.

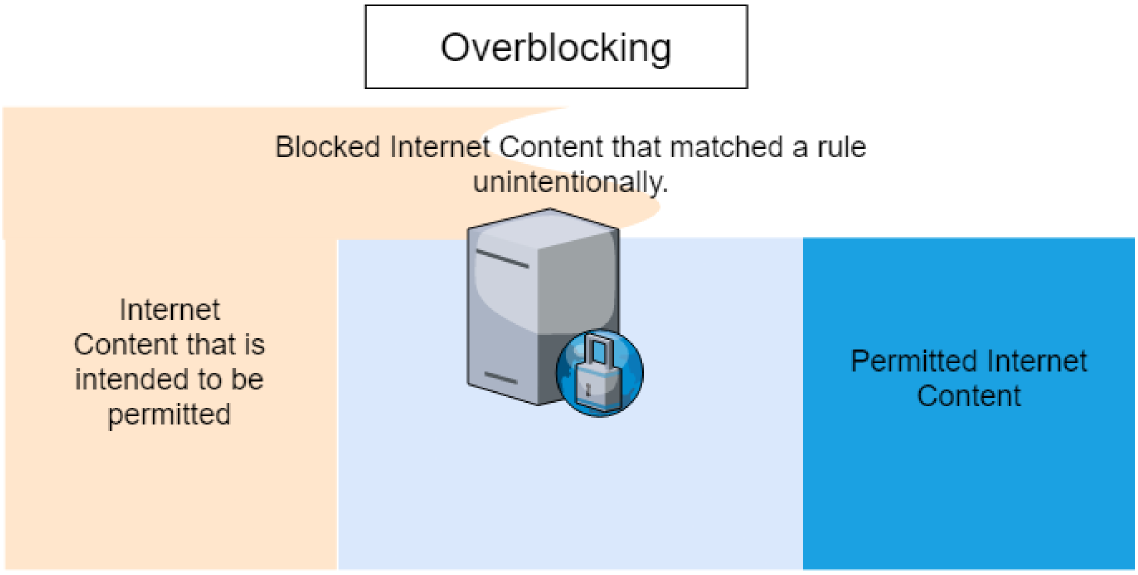

Overblocking
Blocked Internet Content that matched a rule unintentionally.
Internet Content that is intended to be permitted
Permitted Internet Content

Figure 4 Overblocking

Those in favour of overblocking have noted that content filtering systems are imperfect and have argued that if any harmful content is permitted through at all then that can cause harm. As such overblocking is an acceptable cost to pay for the better prevention of unwanted content (Stark, 2007). Although overblocking is considered a side effect it is still a form of denial of service. Those in favour of underblocking have also pointed out that content filtering systems are imperfect and that blocking legitimate content can undermine the purpose of the system from which they are filtering content, that a functional but imperfect system is better than no system at all (Stark, 2007). Additionally Rowe and King (2015) point out that overblocking can be viewed as censorship and impacts free speech. As such underblocking is an acceptable cost to pay to keep the original resource useful. Proponents of both sides have their extreme arguments. Some in favour of overblocking claim that if even one vulnerable individual is harmed by unwanted content then that is one too many. Those in favour of overblocking counter that perfect protection is easily achieved by simply removing the offending resource completely.

ACMA (2008) changed the approach slightly when they re-ran several previous tests they commissioned with some alterations. The same underblocking / overblocking tests were run but now the variable of latency was introduced. Unlike Stark (2007) and Rowe and King (2015), who focused solely on whether or not the material was blocked correctly or not, it was no longer enough for the ACMA (2008) that content filtering systems stop listed material but that they do so quickly. While these trials used a large pool (3930) of URLs, these trials still relied on a pre-defined simulated list. The closed environment test examined six systems in isolation. While the sample network was saturated before testing the systems, it was still a simulated use of the Web by a small number of simulated users in a controlled environment.

The statistics used by the Australian Government for citing the efficiency of content filtering are based around how well a filter blocks or does not block a given piece of content (Greenfield et al., 2001). While Greenfield et al. (2001) mention aspects such as ease of bypassing content filters and also tests for access to redirectors, its address of the efficiency at this task was based on a simple static list of defined redirectors. The closed environment testing report tabled by the ACMA (2008) focused on metrics such as

- Network performance when a content filter is present in both active and passive modes as well as the difference between them.
- How well the filters blocked material, they were supposed to and how often they blocked material they should not.

The tests that have been run on content filters, such as the ones run by Stark (2007), Greenfield et al. (2001) and ACMA (2008), focused on technical performance. These studies focused on technical questions such as How many URLs did the systems block that they were supposed to? How many URLs did the systems block that they should not? How much latency did the systems introduce to the network through their use? These studies focused on the efficiency of these systems rather than their efficacy.

Studies on how well content filters perform at preventing access to restricted content have been performed with simulated traffic or static sets. This is again a measure of efficiency, how well they function, rather than efficacy, how well they fulfil the role or function to which they are tasked. Typical examples of papers that sought to measure the performance of content filters include:

- Al-Hajery (2000) used the web logs gathered over 24 hours from an ISP in 1998. While the data used was gathered from actual users Internet use, it was still recorded data. There was no active user base that could react to access to sites being denied.
- Hunter (2000a) used three sets. 50 URLs generated by using Webcrawler's random links feature. 50 URLs generated by using 5 popular search terms and taking the first 10 URLs for each term. The last set consisted of 100 URLs purposefully selected as sites known to be troublesome in some aspect. There is no user base reacting to Internet sites being denied. This is a test of pattern matching ability rather than efficacy.
- The ACMA (2008) content filtering report used a more comprehensive list– 3930 in three categories (1000 URLs Prohibited content list, 933 selected URLs from the MA15+ to X18+ categories, 1997 selected URLs from the G to M range). Regardless of how extensive the list of URLs is, this is still a static list with no user base being impacted by the actions of the content filters being tested. This is a test of efficiency.
- Patel et al. (2019) used two static pools of web content, 1000 of objectionable and 1000 of unobjectionable content. Without a user based being impacted by the actions of these content filters this is, once more, a measure of efficiency rather than efficacy.

2.6 Summary

These studies all use static lists, even in the cases where the lists were user generated, that failed to impact a user base. Without a user base to respond to and then the possibility that the impacted user base would then attempt to circumvent the content filter, all the papers mentioned above remain measures of efficiency rather than efficacy. These systems, which if applied to an active user base could in turn, elicit a response or change in user behaviour could yield different results in preventing access to restricted content.

Obtaining a traffic set of what pages would be accessed by minors in various age groups is difficult (Stark, 2007). Stark (2007) noted the difficulty of obtaining such data even going so far as to state "To the best of my knowledge, such data do not exist and would be extremely expensive - - if not impossible - - to collect ". This is still discussing the collection of static data without active users actually subject to the rules of a content filters and able to react to the system.

As far as the question of efficacy is addressed, most research appears to avoid the issue of efficacy in favour of the assumption that content filtering works. Yeop et al. (2018) used a survey tool that in part asked about the implementation of content filtering as a mechanism to filter out inappropriate content. Yeop et al. (2018) is a case of using an inexact tool to measure the perceptions of efficacy rather than actual efficacy. Pons-Salvador, Zubieta-Méndez, and Frias-Navarro (2018) ask, in part, about the use of content filters to protect minors aged six to nine and parent's ability to install and configure content filters. The study uses a survey to report on the experience of using content filters and the lack of an empirical measurement of efficacy separates its intent from this thesis. Aktay (2018) queries the perceptions of teachers as to the need for and impact of content filtering in educational environments. These papers assume content filtering is an effective mechanism for protecting minors from material deemed inappropriate and all use imprecise tools to measure experiences and perception rather than efficacy. Quantitative measurement of an active user base is missing from all of these papers, primarily because what they sought to measure is perception and experience rather than efficacy.

When performance, rather than experience, is addressed the focus remains on efficiency rather than efficacy. Al-Hajery (2000) used pre-gathered web logs and ran them through various content filters. Hunter (2000a) used three sets of pre-gathered data from several search engines and ran these through various content filters. The ACMA (2008) drew their web sites from a list of content restricted by Australian law and set them against a selection of different content filters. Patel et al. (2019) used two static pools of web content to test a neural network-based content filter. All of these evaluated how efficiently content filters could pattern match and block access to selected sites from a pre-selected dataset. There is no allowance in any of these studies for a live papulation that would be impacted by the operation of these content filters nor is there any room for circumvention to arise as a result of that impact. The issue of efficacy is not examined in the literature.

This thesis examines the efficacy of an in-place content filter. This thesis does not evaluate the effectiveness of content filtering algorithms or implementations. This thesis likewise does not debate the usefulness or purpose of policy, either political or local. The focus of this thesis is exclusively on the efficacy of a content filter in a live environment. As such the issues addressed are attempts made to circumvent the content filter, do circumvention attempts succeed and what techniques are used to circumvent the examined content filter.

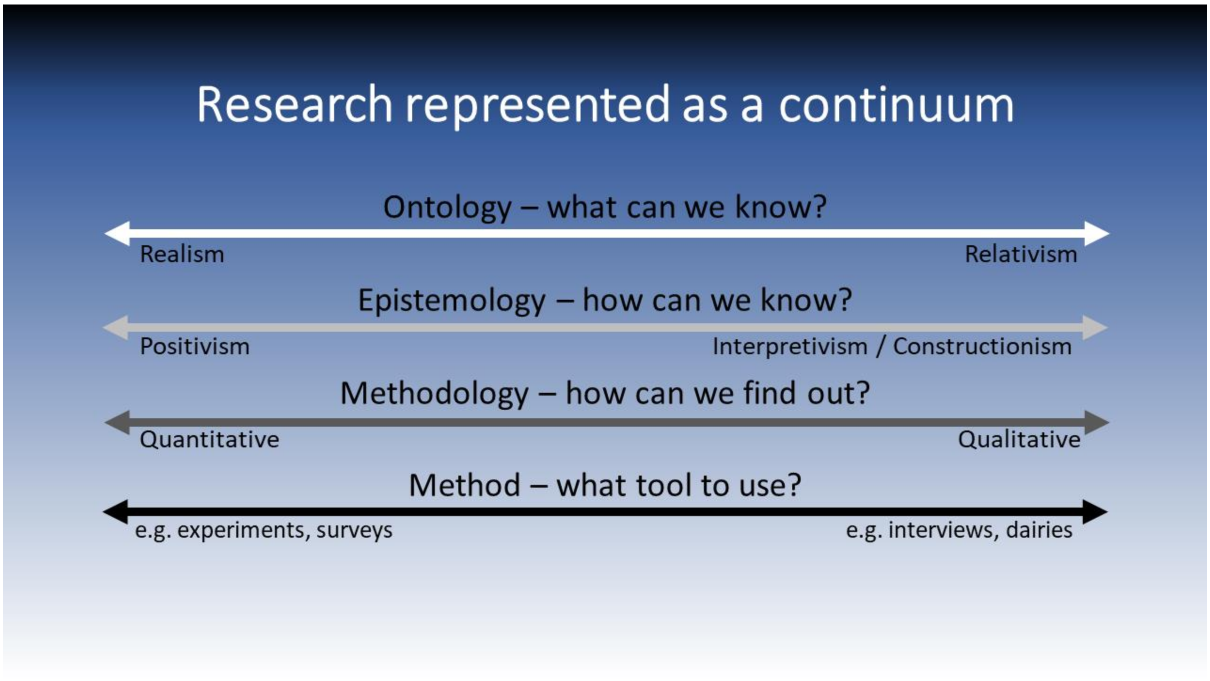
Chapter 3 Research Methods and Design

This chapter begins with a discussion on methodology and research approaches, from the general to some of the more commonly used methods in the information science discipline. The chapter then presents a discussion of research methods and justifies the chosen method for this research . The methods section is followed by the research design, describing the process used to examine the data. This chapter finishes with the limitations of the chosen research method and details of the ethics declaration associated with this research.

3.1 Methodology

Research can be viewed as a continuum between two diametrically-opposed world views on how knowledge is understood (Galliers, 1992; Williamson & Johanson, 2013). These world views inform the researcher's epistemology or how knowledge can be acquired. In turn, how researchers choose to approach the problem epistemology favours certain methods. The more quantitative methodologies tend to be used more by those with a positivist epistemology as they are quantifiable. As the perspective of an epistemology shifts more towards the interpretivist, the methodology tends to become more qualitative. In turn, certain methodologies tend to favour certain methods. This is not to say that particular methodologies must use certain methods but that the results that certain methods provided will be more or less quantitative and as such favoured by a particular world view (Williamson & Johanson, 2013).

The views of what we can know, ontology, are largely predicated on two opposing positions. Realism deals with the idea of an objective world that exists separate from the observer and that knowledge can be determined from the observation of that world. Relativism is commonly used to encompass a wide range of positions that have their basis, broadly speaking, in knowledge being relative from a given position. Research can be divided broadly into two main categories, Interpretivist and Positivist (figure 5). Other categories can be debated to exist in their own right, such as Critical Theory or Post-Positivist but there are other arguments that place these approaches as sub-categories of the first two (Critical Theory and Interpretivist) or are a combination of them (Williamson, 2002).


Research represented as a continuum

Ontology – what can we know?

Realism | Relativism

Epistemology – how can we know?

Positivism | Interpretivism / Constructionism

Methodology – how can we find out?

Quantitative | Qualitative

Method – what tool to use?

e.g. experiments, surveys | e.g. interviews, dairies

Figure 5 Research as a Continuum adapted from (Condie, 2012).

Positivist research is based on the classic scientific method. That is, the world can be described through objectively observable phenomena. Those phenomena can then be measured or quantified. The world view itself is strongly rooted in quantitative methods and deductive reasoning (Table 3). The classic perspective is that a hypothesis or idea is put forward to explain an observable event. The hypothesis is then tested, through further observations or experiments, with the aim of proving the hypothesis false. If proven false, the hypothesis is discarded. If not proven false, the theory is corroborated but not proven true. (Galliers, 1992; Williamson, 2002; B. C. Beins, 2012; Williamson & Johanson, 2013)

In contrast, Interpretivist research as an approach is based on the assumption that meaning in the social world is interpreted and that this interpretation happens through the perception of the researcher (table 3). The methods used in Interpretivist research tend towards qualitative methods and inductive reasoning. Interpretivist research is mainly associated with qualitative research methods and is sometimes simply referred to as qualitative research. This does not mean that quantitative methods cannot or are not used in Interpretivist research. Interpretivist research does not usually test hypotheses but rather is used to develop propositions. As the interpretivist is in the pursuit of meaning, such research is not always generalisable and the research is not always replicable. (Galliers, 1992; Williamson, 2002; B. C. Beins, 2012; Williamson & Johanson, 2013)

Research paradigm (world view)	Ontology	Epistemology	Research methodology	Modes of inquiry
Interpretive	Society is constructed and social reality is constantly interpreted	Knowledge is subjective and is generated through "exploration of the beliefs, feelings and interpretations of research participants" (Williamson & Johanson, 2013)	Qualitative	Interviews Case studies Field Experiment
Positivist	A theory, if not proven false, is corroborated but not proven true	Knowledge is objective and generated through observation.	Quantitative	Experimental Quasi-Experimental Surveys

Table 3 Research Methods adapted from (Williamson & Johanson, 2013)

The field experiment methodology is a pre-experimental or field study design (Williamson & Johanson, 2013). The researcher is interactive, placing this methodology close to the participant observation and results tend to be collected in a narrative form (Jackson, 2015). This methodology has the advantage of being conducted in a natural setting reflecting real life use which is applicable to these questions. However there still exists the need to introduce a variable into the environment to observe the changes (Williamson & Johanson, 2013) and interaction has the intention of altering the behaviour of those observed (Jackson, 2015). Since interaction with the environment involves the possibility of altering the behaviour of the observed or, more importantly, the educational outcomes of students this methodology has significant disadvantage.

The case study is often viewed as one of the oldest qualitative research methods (Jackson, 2015) and is a broad and flexible method that can be used in many ways, in many fields. This includes interpretive or positivist, deductive or inductive, investigates one or multiple cases and can use either qualitative or quantitative data (Williamson & Johanson, 2013; Yin, 2013). It is a useful methodology when the subject is vast and observing the events outside the context in which they naturally occur would not yield accurate findings (Williamson & Johanson, 2013). Since the collection of data may consist of analysis of recorded data (Williamson & Johanson, 2013; Yin, 2013) it is able to keep interaction with the environment as slight as possible and minimise the impact such interaction could have on altering the observations.

The disadvantage of a case study is that is that generalisations cannot be drawn from it as the subject could possibly be atypical (B. C. Beins, 2012; Yin, 2013; Jackson, 2015) as opposed to a representative sample selection (Williamson & Johanson, 2013). The lack of ability to impact the environment while providing the information required to answer the research questions makes the case study method applicable to this enquiry. This research examines data collected on World Wide Web (WWW) usage and access and analyses the attempts made to access restricted content. As such the research is investigating a real-world application.

An experimental approach would have involved influencing the real environment of minors to observe the results. The experimental approach requires independent controls and control of the environment to eliminate undesired variables (Galliers, 1992; B. C. Beins, 2012; Williamson & Johanson, 2013). Since the data comes from a real educational environment, introducing a change to the environment has the possibility of altering the behaviour of those observed (B. C. Beins, 2012; Jackson, 2015). It is because these constraints will be difficult to impossible to implement that this approach is rejected.

There are significant advantages over other methods to using a case study in that the answer to the research questions becomes self-evident in the behaviour of the subjects free of any influence other methods might impose on the subjects. The case study method also often makes suggestion of hypotheses for future study easily identifiable. The case study also has limitations. These include the possibility of the subject being atypical leading to erroneous generalisations and the issue of researcher bias leading to a subjective interpretation of data. These limitations, particularly the issue of researcher bias, were be kept in mind when using this method (B. C. Beins, 2012; Yin, 2013; Jackson, 2015).

3.2 Research Questions

The development of the research questions arose from informal discussions with school staff and anecdotal examples. Staff believed that circumvention was widespread but were uncertain as to the degree or the methods being used. It was first believed that URL redirectors were being used as this is one of the simplest methods to use and with many URL redirectors available, it is easy to move from redirector to redirector once one has been blocked. With these discussions in mind, the following research questions were developed.

RQ1 Do students successfully circumvent content filtering systems?

> If students do not manage to circumvent the content filter, then there can be no evidence to draw a distinction between the measurement of efficiency and efficacy.
>
> H_1 Students circumvent the content filter.
>
> H_2 The content filter blocks 95% or more of restricted content when directly accessed.

RQ1(a) Do students attempt to access restricted content?

> Students need to attempt to access restricted content to be able to measure any effectiveness or efficacy of the content filter.

H_3 Students attempt to access restricted content

H_4 Students attempt to circumvent the content filter.

RQ1(b) How prevalent is content filtering circumvention?

If students are attempting to circumvent the content filter and succeeding, then how often are they circumventing the filter?

H_5 80% or more students circumvent the content filter

RQ2 What are the techniques used by students to circumvent content filtering?

If the answer to RQ1 is in the affirmative, then what methods do students use to circumvent the content filter?

H_6 Students are using URL redirectors to circumvent the content filter.

RQ2(a) To what degree do students circumvent content filtering systems?

H_7 80% or more of students circumvent the content filter for 80% of their web requests.

RQ2(b) What does the pattern of access describe about circumvention behaviour?

H_8 Students use a common, or small selection, of methods to circumvent the content filter.

3.3 Research Design

The methodology chosen for this research will be a case study. Data analysis will begin with a pattern matching approach based on a large population. The data was gathered in the form of proxy logs over two periods. An analysis over such a length of time lends any findings more robustness by reducing the possibility, likelihood or effect of anomalous events such as statistical regression (Bernard C. Beins, 2004; B. C. Beins, 2012).

This research involves the analysis of World Wide Web (WWW) proxy logs and matching blacklist that have been collected from a proxy content filter in a high school in Western Australia. The data was copied to a secure environment for analysis and was unaltered. The data was processed according to the workflow depicted in figure 6. The data first needed to be cleaned (Williamson & Johanson, 2013; Yin, 2013) to remove data that is not relevant to the research questions. In this research this required the removal of data related to user accounts that are not students and as such subject to different access rules and a different blacklist. Once cleaned, a pattern matching method was used to examine the data for matches between restricted URLs and the recorded behaviour of users. The process looked for evidence of circumvention by looking at any matches identified in the previous step for any record that permitted access. For successful access records that are found the following questions are examined; what methods permitted the circumvention of the proxy content filter? how many users used this method? and how often do individual users use this method of circumvention? Finally, if evidence of circumvention is found, what restricted sites did the students manage to access?

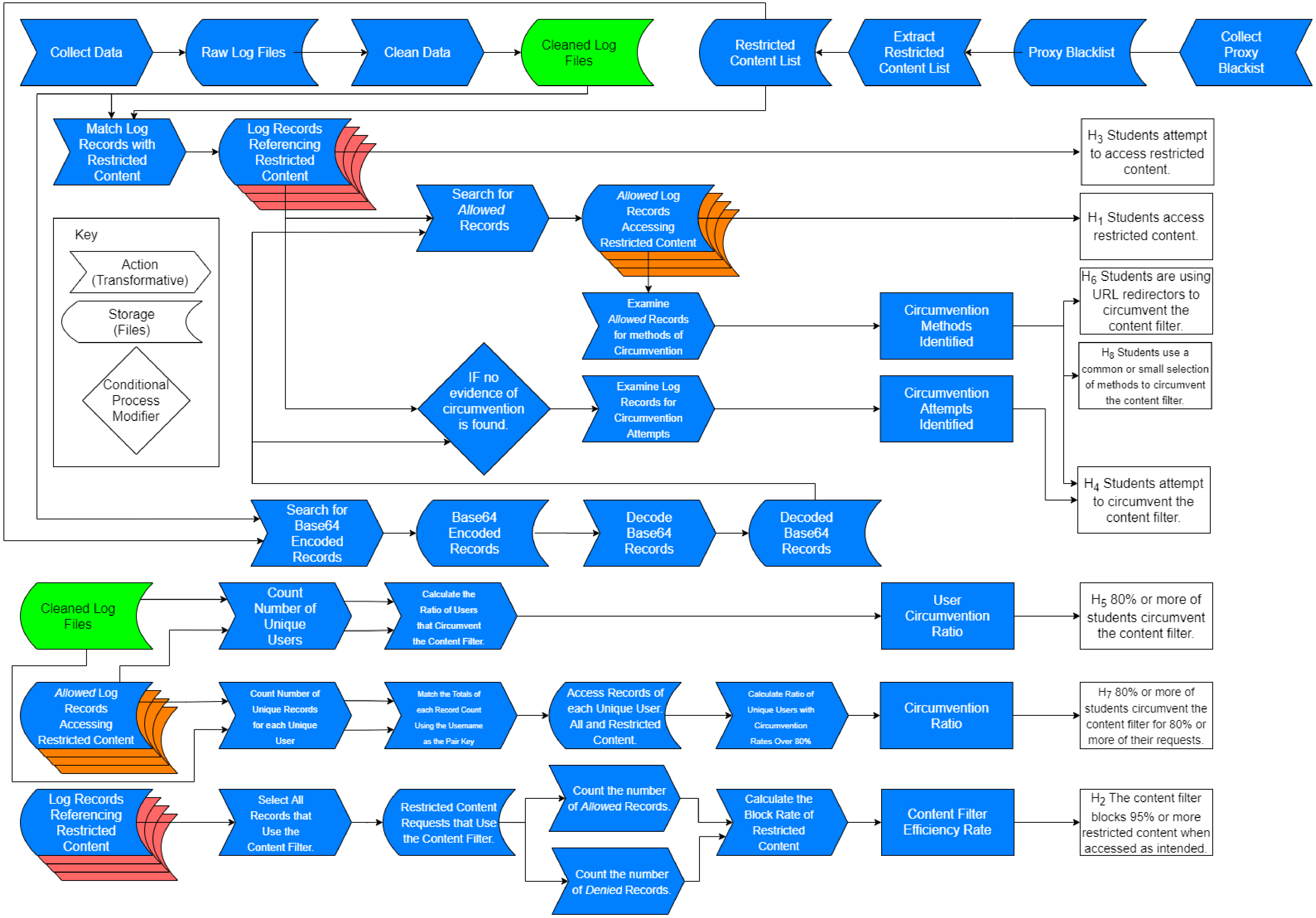

Collect Data
Raw Log Files
Clean Data
Cleaned Log Files
Restricted Content List
Extract Restricted Content List
Proxy Blacklist
Collect Proxy Blackist
Match Log Records with Restricted Content
Log Records Referencing Restricted Content
H_3 Students attempt to access restricted content.
Search for *Allowed* Records
Allowed Log Records Accessing Restricted Content
H_1 Students access restricted content.
Key
Action (Transformative)
Storage (Files)
Conditional Process Modifier
Examine *Allowed* Records for methods of Circumvention
Circumvention Methods Identified
H_6 Students are using URL redirectors to circumvent the content filter.
H_8 Students use a common or small selection of methods to circumvent the content filter.
IF no evidence of circumvention is found.
Examine Log Records for Circumvention Attempts
Circumvention Attempts Identified
H_4 Students attempt to circumvent the content filter.
Search for Base64 Encoded Records
Base64 Encoded Records
Decode Base64 Records
Decoded Base64 Records
Cleaned Log Files
Count Number of Unique Users
Calculate the Ratio of Users that Circumvent the Content Filter.
User Circumvention Ratio
H_5 80% or more of students circumvent the content filter.
Allowed Log Records Accessing Restricted Content
Count Number of Unique Records for each Unique User
Match the Totals of each Record Count Using the Username as the Pair Key
Access Records of each Unique User. All and Restricted Content.
Calculate Ratio of Unique Users with Circumvention Rates Over 80%
Circumvention Ratio
H_7 80% or more of students circumvent the content filter for 80% or more of their requests.
Log Records Referencing Restricted Content
Select All Records that Use the Content Filter.
Restricted Content Requests that Use the Content Filter.
Count the number of *Allowed* Records.
Count the number of *Denied* Records.
Calculate the Block Rate of Restricted Content
Content Filter Efficiency Rate
H_2 The content filter blocks 95% or more restricted content when accessed as intended.

Figure 6 Research Workflow

3.3.1 The Network Environment and Source Data Recording

The environment for this case study is a high school network as shown in Figure 7. The network had, at the time of data collection, over 1000 connected devices and all devices were required to access the world wide web (WWW) through the Microsoft Internet Security and Acceleration (ISA) proxy called BILL. It is important to note that there is no possible access to the WWW or any network, other than the server network, without going through the proxy BILL. The proxy serves several benefits such as caching web content for faster access to common requests and serves as a content filter to deny access to restricted sites.

As each request is made the proxy, records the details and outcome of each request in a file called a logfile. In this way there is a record of all activity sent out and coming into the network. Each logfile is rotated daily which means at the end of each day, the current logfile is closed and a new logfile is created for the new day and logging then commences in the new logfile. Each logfile is named in the format of ISALOG_YYYYMMDD_WEB_LLL.w3c where:

- YYYY is the year in full format, e.g.: 2010
- MM is the month number in two digits, e.g.: 09 being September
- DD is the day in two digits, e.g.: 08 being the 8th
- LLL is the log number in case the log file grew to such a size that a new log file is required. This was never encountered and as such was always 000.

As such an example logfile filename would be ISALOG_20100809_WEB_000.w3c

Looking at recorded proxy records has the advantage of identifying general patterns as they have occurred without assuming which variables may have caused the behaviour. This first part of the study is aimed at determining if users attempt to circumvent the content filter and if so, are they successful?

Once specific sites of interest have been identified, the research will shift to sites of statistical significance to attempt to identify behaviour patterns more clearly. This is where successful circumvention, if found, will be examined as to how this restricted content was accessed and what methods were favoured.

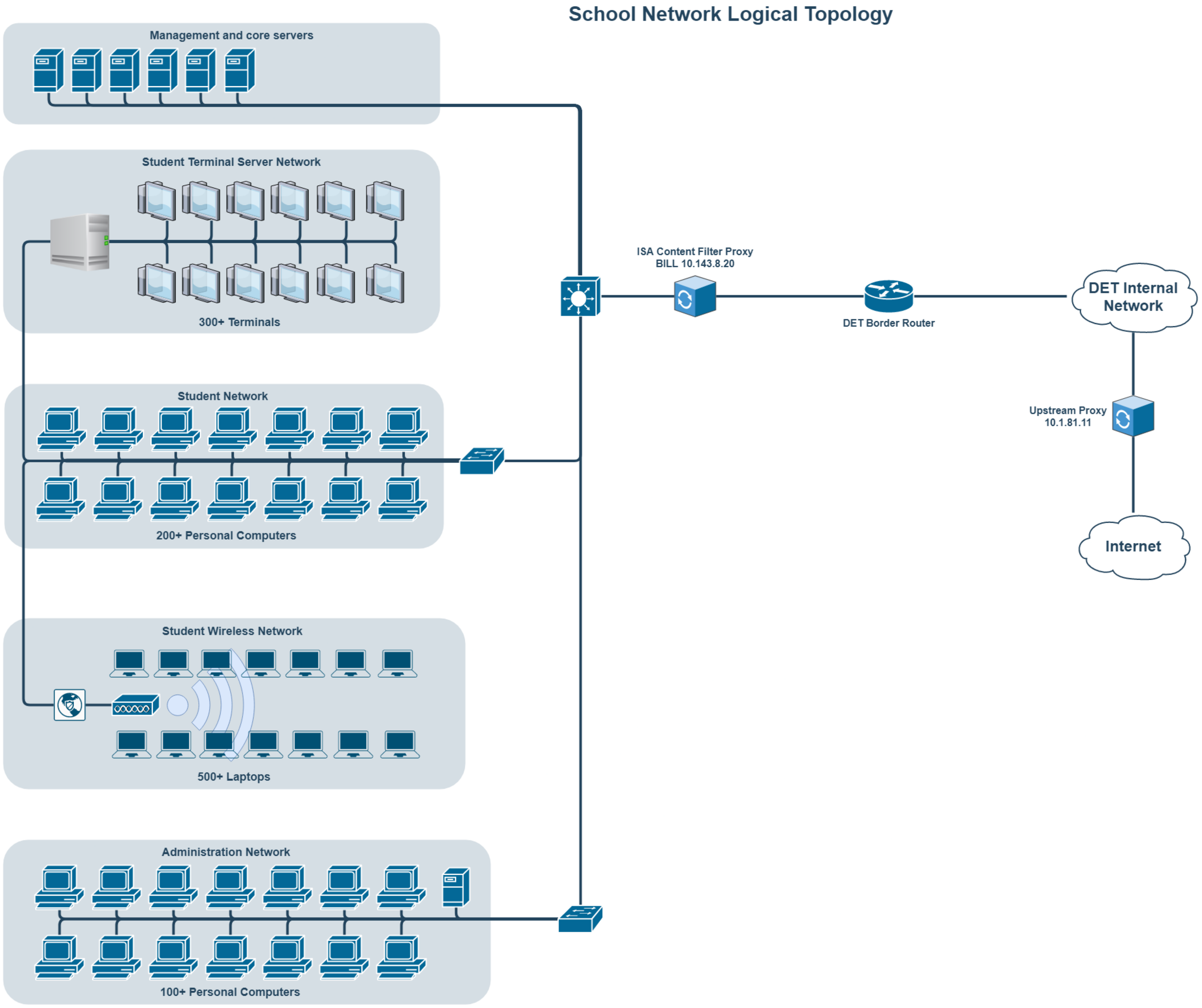


Figure 7 School Network Logical Topology

3.3.2 Step 1: Data collection and cleaning

The data sources consist of web proxy logs and the matching blacklist collected at the designated site at two intervals. The data was collected in two intervals. The data was cleaned to remove data records that were relevant to the study. Since the target demographic are minors 11-17 years old, all staff and administration data were removed. This is because the staff are both outside the age range being studied and because staff are permitted access to sites students are not. It is possible this would cause false positives for access to restricted content. In the case of the data available:

- staff have usernames with patterns unique from students beginning with an E followed by only numbers.
- users external to the network have easily identifiable usernames beginning in one of two easily identifiable patterns.
- machine accounts used for systems administration and updating all have usernames that end with a $

These records are easily identified and removed without inadvertently removing valid data. The script for data cleaning can be found in Appendix E.

The file containing the restricted site blacklist is stored in XML format (Appendix D). For the data in this file to be useful, the XML formatting needs to be stripped leaving just the list of restricted sites. This will allow for the restricted sites to be used in pattern matches in the data separation phase.

3.3.3 Step 2: Data Separation

Once the data was cleaned, the data was separated by restricted site name. For every restricted site listed in the blacklist, a search was conducted of the proxy logs for a record match. If there is a match for a restricted site, the record was copied into a file for the restricted site (Figure 8). This allows for the quantification of circumvention and attempted circumvention based on site. The custom script siteparse.sh (Appendix C) was used for this purpose.

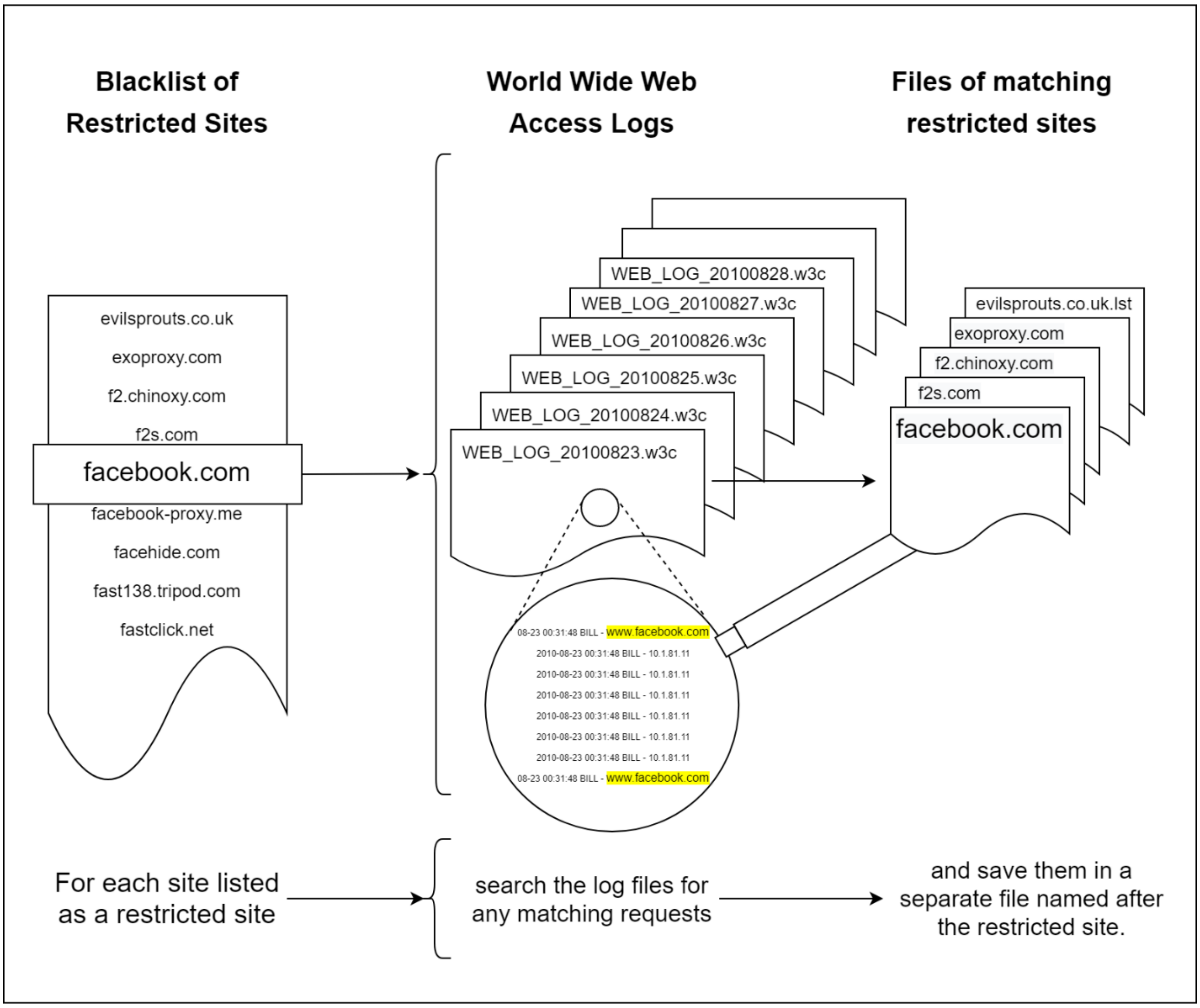

Blacklist of Restricted Sites
World Wide Web Access Logs
Files of matching restricted sites
evilsprouts.co.uk
exoproxy.com
f2.chinoxy.com
f2s.com
facebook.com
facebook-proxy.me
facehide.com
fast138.tripod.com
fastclick.net
WEB_LOG_20100828.w3c
WEB_LOG_20100827.w3c
WEB_LOG_20100826.w3c
WEB_LOG_20100825.w3c
WEB_LOG_20100824.w3c
WEB_LOG_20100823.w3c
08-23 00:31:48 BILL - www.facebook.com
2010-08-23 00:31:48 BILL - 10.1.81.11
2010-08-23 00:31:48 BILL - 10.1.81.11
2010-08-23 00:31:48 BILL - 10.1.81.11
2010-08-23 00:31:48 BILL - 10.1.81.11
2010-08-23 00:31:48 BILL - 10.1.81.11
2010-08-23 00:31:48 BILL - 10.1.81.11
08-23 00:31:48 BILL - www.facebook.com
evilsprouts.co.uk.lst
exoproxy.com
f2.chinoxy.com
f2s.com
facebook.com
For each site listed as a restricted site
search the log files for any matching requests
and save them in a separate file named after the restricted site.

Figure 8 The restricted site requests separation process

3.3.4 Step 3: Analysis

Once the data was separated, each file represented a restricted site in the proxy blacklist. For each of these sites, text pattern matching can be used to count how many users attempted to access that site and how many times each user was successful.

Following the extraction of the access request records for the restricted sites, a manual search of *Allowed* records was conducted. There are many possible circumvention techniques, and many are difficult to detect in an automated manner without knowing in advance what you are looking for. Custom scripts (Appendix C and E) and commands were written. The commands are presented in the analysis section as they are discussed with the information they were attempting to extract. The scripts are provided in the appendices.

The raw data has the format of (Figure 9):

```
Client IP Address       Username        User Agent String       Date    Time    Server
        Referring Server        Destination Host        Destination Host IP     Destination Port
        Processing Time         Bytes Received  Bytes Sent      Protocol        Operation
        URL     MIME Type       Result Code     Rule    Filter Information      Source Network
        Destination Network     Error Code      Action
```

Figure 9 ISA w3c log format

The fields in Figure 9 are explained in detail in Appendix B.

When an actual log record is examined it appears similar to (figure 10):

```
10.143.13.200   <username>      Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.6; en-US;
rv:1.9.2.2) Gecko/20100316 Firefox/3.6.2        2010-11-30      05:07:25        BILL    -
        static.ak.connect.facebook.com 10.143.8.20      80      1       445     181     http
        GET
        http://static.ak.connect.facebook.com/js/api_lib/v0.4/FeatureLoader.js.php/en_US
        -       12202   Deny Access to URL sets STUDENTS        Req ID: 0a3f696d        Internal
External        0x80    Denied
```

Figure 10 ISA W3C log example

Figure 10 is a request from a student to a Facebook server. This request matched the rule Deny Access to URL sets STUDENTS and the resulting action was the request was *Denied*. In a combined layout the fields would map as in table 4.

Field	**Example**
Client IP Address	10.143.13.200
Username	<username>
User Agent String	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.6; en-US; rv:1.9.2.2) Gecko/20100316 Firefox/3.6.2
Date	2010-11-30
Time	05:07:25
Server	BILL
Referring Server	-
Destination Host	static.ak.connect.facebook.com
Destination Host IP	10.143.8.20
Destination Port	80
Processing Time	1
Bytes Received	445
Bytes Sent	181
Protocol	http
Operation	GET
URL	http://static.ak.connect.facebook.com/js/api_lib/v0.4/FeatureLoader.js.php/en_US
MIME Type	-
Result Code	12202
Rule	Deny Access to URL sets STUDENTS

Filter Information	Req ID: 0a3f696d
Source Network	Internal
Destination Network	External
Error Code	0x80
Action	Denied

Table 4 Log file fields with examples

3.4 Methodology Limitations

The case study method is useful for studying events which, when stripped of the context of their environment, may yield different results. Given that the metrics surrounding content filters are measures of efficiency in isolated environments devoid of actual users, the case study method is appropriate for examining the efficacy of content filters in a live environment. The case study has serious limitations in that the results cannot be easily generalised to show larger patterns of behaviour.

It is important to note that this thesis is designed to look solely at the issue of circumvention in a live implementation. If circumvention is attempted, if circumvention succeeds and what methods are used to achieve circumvention. As a result, this thesis is limited in scope to these parameters. This approach cannot answer questions such as what motivates users to circumvent a content filter, how common content filtering is or how difficult users find circumventing a content filter.

3.5 Ethics

The research did not involve humans or animals.

Data was saved on a 1TB External Hard Drive and when not in use is stored in a locked cabinet.

An ethics declaration was approved by the ECU Ethics committee Ethics reference: 2019-00788-TURNER. The Department of Education and Training approved the original collection of data in written communication D10/0088840.

Chapter 4 Results and Analysis

This chapter begins with a description of the data used in this case study and then is followed by the results of the analysis. The chapter is presented in the order that the analysis was done as this reconstructs the journey undertaken and shows the results of one section informing the analysis of the next section. Each subsection, where analysis is presented, the results are first presented and then followed by the analysis. This structure is used because of the nature of exploring this case study. This chapter is written in the order the steps were taken and the results had an impacted on the steps taken in the next step. This environment made it logical to address the discussion in this manner. The results are presented in the manner of the hypothesis being tested, the method used to test the hypothesis, the result of the test and how this addresses the associated research question. The opening discusses how the data was found to be formatted and how it was structured. This is followed by an explanation of how the underlying premise of blacklist-based content filters creates the opportunity for exploitation. The next section looks to see if students attempt to access restricted content. Then there is an examination to determine if students access restricted content despite the filter being in place. The following section explores how students manage to circumvent the content filter rules to access restricted content. The chapter ends by answering the questions of how effective the content filter is, efficacy vs efficiency and how pervasive circumvention of the content filter was.

4.1 The data

The data as collected is clearly designated in to two, separate sets by a time gap. The first, dataset 1, ranges from 23rd of August until 31st of December 2010. The second, dataset 2, from the 3rd of August until the 31st of December 2011. The period from the 1st of January to the 2nd August 2011 is not recorded and provides a clear 7-month separation between the two samples.

The gap between the two samples was not deliberate but rather the result of unauthorised and untrained alteration of the content filter's log collection configuration. A new hire in the school decided to familiarise themselves with the content filter and altered the configuration. As a result of this tempering the content filter was rendered ineffective and the blacklist invalid. Further, the misconfiguration of the content filter caused the content filtering log collection to cease. This misconfiguration was not rectified until the researcher noticed the issue on a scheduled data collection visit and asked to correct the issue.

The original data are 135GB of World Wide Web Consortium extended log format (W3C) text files from a Microsoft ISA Proxy (table 5) with dataset 1 being 56GB consisting of 141,037,931 web request records and dataset 2 being 79GB consisting of 242,539,694 web request records. These data are then cleaned with the script sanatize.sh (Appendix E) to remove records that are not related to student activity. The removal of these records ensures that the results are not contaminated by requests that originate from users other than students. The raw data has the format as shown in figure 9. The resulting data are 19GB consisting of 41,548,548 web requests records and dataset 2 being 10GB

consisting of 30,509,943 records.

	Dataset 1	Dataset 2
Size	56GB	79GB
Number of records	141,037,931	242,539,694
Cleaned size	19GB	10GB
Cleaned number of records	41,548,548	30,509,943

Table 5 Data composition

4.2 Relevance

The collected data is from a blacklist proxy content filter system (figure 11). While technologies have evolved since this point, the same fundamental weaknesses in modern systems that use blacklists remain. The Internet consists of a large number of sites, over 1.6 billion, and is growing by around 1500 new sites per day says the Hosting Tribunal (2020). The surest, but not infallible, way to prevent access to restricted material would be to restrict access to only pre-approved material, i.e. a white list (Klang & Murray, 2016). The most flexible use the of Internet requires the user to be free to use unknown and new sources, behaviour which a white list either inhibits or can outright prevent. Blacklists, a list of known undesirable content, is the next step down that allows a user to access anything not specifically deemed restricted and this is where the inherent issue exists. So long as users of a content filter are able to access anything not predetermined to be undesirable, this is the area where those wishing to circumvent a content filter can operate (Wiley, 2016; Chen & Nguyen, 2018).

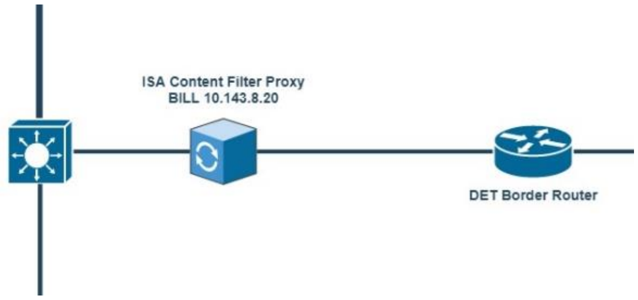


Figure 11 Content filter proxy topology

Content filtering technology has advanced since this data were collected. There are services that examine sites constantly and provide constant blacklist updates with categorisation (Klang & Murray, 2016). There are machine learning systems that are trained to look for the type of sites that are restricted, including sites that enable circumvention such as URL redirection sites (Chen & Nguyen, 2018). The same is also true for circumvention techniques. The open source proxy, Squid, is freely available for any individual to set up a previously unknown proxy. VPNs are now able to be used over HTTP protocols to secure traffic through a web content filter (Dixon, Ristenpart, & Shrimpton, 2016). This is a competition in development between content filters and circumvention methods that has evolved not only since this data was collected but since the effort to restrict information began. The data examined in this thesis in just another step in the long journey of information control.

4.3 Extraction of restricted site requests

Hypothesis H_3 stated that students attempt to access restricted content. To determine this, the cleaned log files were matched against the list of restricted sites (Appendix A) using the custom script siteparse.sh (Appendix C). Any match of a request for a restricted site was copied into a file having the name of the matched restricted site.

Most of the sites in the blacklist of restricted sites resulted in no attempted access but a smaller list of restricted sites resulted in requests for access. Of the 1393 restricted sites contained in the blacklist, Dataset 1 returned 974 (70%) sites and dataset 2 returned 1025 (74%) sites with no requests. There are many possible reasons for the existence of so many sites on the restricted blacklist, most of them mundane, that no one attempts to access. This is unsurprising since the blacklist was a living list that, rather than being maintained and curated, was simply added to over time. However, some sites showed significant traffic. The more interesting results were that dataset 1 yielded requests to 419 restricted sites with a total of 4,745,629 requests (11.42% of all requests in dataset 1). Dataset 2 resulted in requests to 368 restricted sites with a total of 2,960,906 requests (9.27% of all requests in dataset 2) as shown in figure 12.

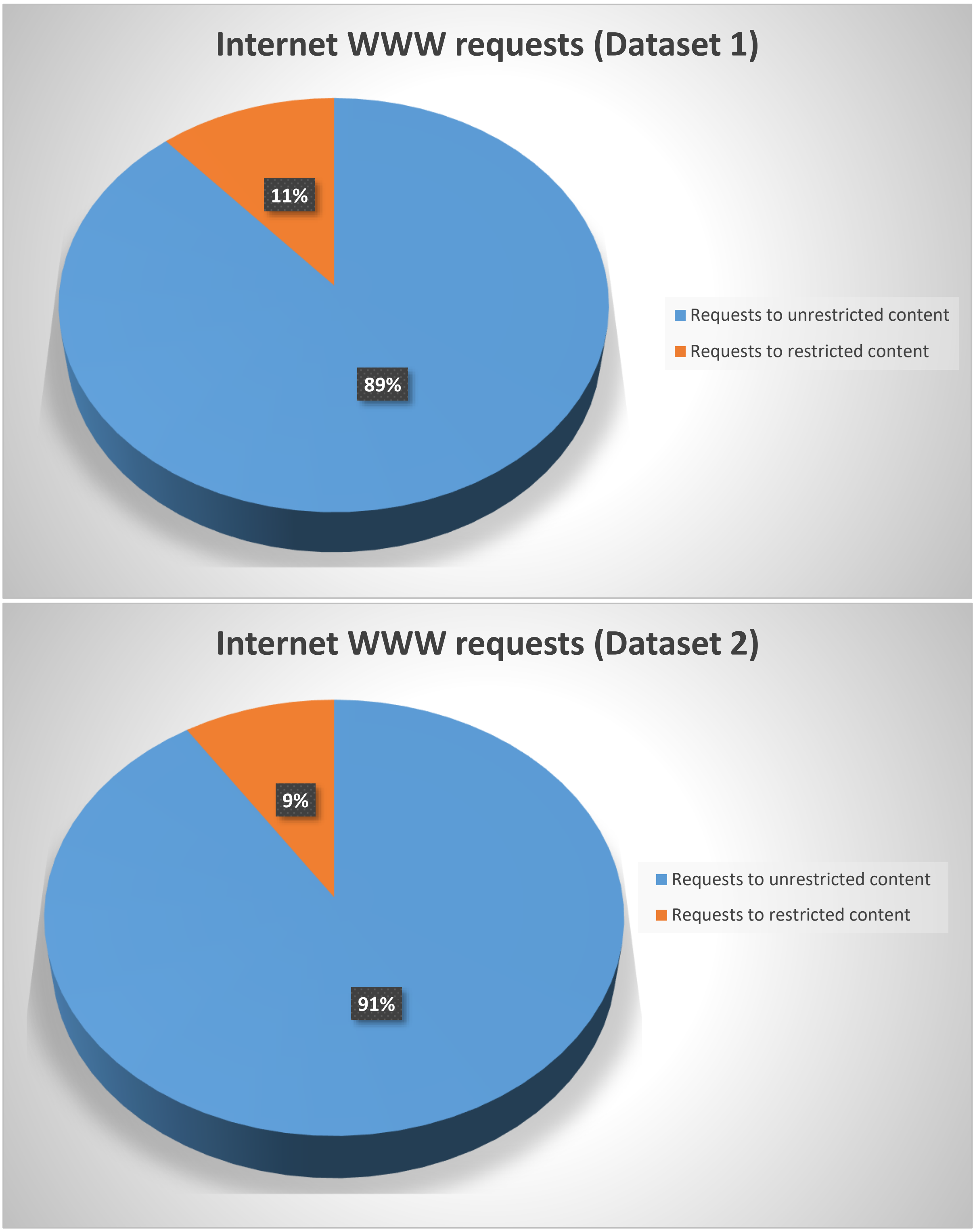


Figure 12 Pie chart - Internet WWW requests

The most attempted access for restricted sites in the first data set (2010), along with the size of the log data is listed in table 6.

Site name	**# of requests**
doubleclick.net	588720
ninemsn.com.au	624174

www.facebook.com	674677
google-analytics.com	522122
surfagain.com	314290
googlesyndication.com	317850
All .info domains	277533
youtube.com	140868
yieldmanager.com	160774
tornposter.com	120657

Table 6 Top restricted sites accessed from Dataset 1 (2010)

Dataset 2 resulted in a slightly different list (table 7).

Site name	**# of requests**
facebook.com	844561
doubleclick.net	721837
google-analytics.com	318515
gravatar.com	264393
yieldmanager.com	144399
All .info domains	157068
youtube.com	133349
googlesyndication.com	118402
www.jango.com	55095
www.unblockyoutube.com	49982

Table 7 Top restricted sites accessed from Dataset 2 (2011)

Once the traffic to restricted sites was removed to specific files, each file was inspected. Each file contained records where the access was requested to restricted sites as expected. These top accessed sites can be categorised into 5 general categories. These are:

- Social media (facebook, gravatar): These sites were restricted to reduce students wasting time in class. Gravatar, while not a social media site itself, is linked to other site with a social presence, such as a comment section, since this when calls to the avatars Gavatar makes available occur.
- Media streaming (youtube, jango): These sites are, again, restricted to reduce students wasting time in class.
- Advertisement (doubleclick, google-analytics, googlesyndication, yieldmanager): This section is probably indicative of a much larger problem. These are advertising sites that are not accessed directly but rather are requested as imbedded components in other pages. While circumventing a content filter in this case is resulting in advertisements being displayed when they would otherwise not be, this is a clear case of content that would be otherwise unwanted being displayed. That is once the content filter is circumvented it is circumvented for all sites in the content filters blacklist, for malicious sites as well as beneficial.

- Redirector proxies (surfagain, tornposter, unblockyoutube): These are external proxy redirectors designed to fetch requested content on the user's behalf and then displaying the content in its own webpage. Since the content is technically being displayed by the proxy redirector site and not by a restricted site, it is a tool that can be used to bypass a content filter so long as the redirector proxy is not on the content filter's blacklist. It is important to note that a redirector proxy is a tool and not just a means of evading content filters. Among their uses, they can also be used for cyber safety in preventing sites from gathering identifying information about user's through connection information or to anonymise access.
- News (ninemsn): Ninemsn is both a news website and the default webpage for Internet Explorer in 2010 and 2011. There is evidence that some users accessed ninemsn for news, but most were to the home page indicating that this resulted from simply opening Internet Explorer for use.

Not listed is the .info domains. This listing in the blacklist was a wildcard listing of *.info. That is, every .info domain was banned regardless of content. There is no way to know the reasoning as to why all .info domains were restricted.

Among the restricted sites that requests were made for were sites designed to circumvent proxy blacklists. The sites surfagain.com and tornposter.com, specifically, state that circumventing school restrictions is in part why they existed. This result confirms the hypothesis H_3 and answers the research question RQ1(a)" Do students attempt to access restricted content?" in the affirmative.

4.4 Search for allowed access to restricted sites

Hypothesis H_1 states that students circumvent the content filter or answers the question, do students access restricted content? To determine this, the log files of the restricted sites created in the previous step were examined. If a student attempted to access a restricted site and failed, the log record would look similar to figure 13:

```
10.143.12.29    <username>    Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.6; en-US;
rv:1.9.2.2) Gecko/20100316 Firefox/3.6.2        2010-08-23    00:43:32      BILL    -
        static.ak.connect.facebook.com 10.143.8.20    80      1       434     181     http
        GET
        http://static.ak.connect.facebook.com/js/api_lib/v0.4/FeatureLoader.js.php/en_US
        -       12202   Deny Access to URL sets STUDENTS      Req ID: 0c571fbd        Internal
        External        0x80    Denied
```

Figure 13 Denied record example

This demonstrates that a request to a restricted site directed to the content filter proxy the request is correctly denied. A text search for the action of "*Allowed*" resulted in records for restricted sites, in the case of figure 14 the exact same site, that were permitted noting that a different rule has been matched. Closer inspection of the records in the files of restricted sites found that all those with a destination field that listed the proxy were *Denied*. That is, when the intended destination of the request was the proxy, the rules matched as expected and access to restricted sites was *Denied*.

```
10.143.13.201   <username>    Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0;
STUDENT; InfoPath.2; .NET CLR 3.0.04506.30; .NET CLR 2.0.50727; .NET CLR 3.0.4506.2152; .NET
CLR 3.5.30729; .NET CLR 1.1.4322; .NET4.0C; .NET4.0E; STUDENT)        2010-08-23
        01:26:48        BILL    -       10.1.81.11      10.1.81.11      8080    328     858
        671     http    GET     http://facebook.com/    Upstream        301     Allow Access To
Servers Req ID: 0c59489e        Internal External       0x580   Allowed
```

Figure 14 Allowed record example

When the *Allowed* records of the restricted sites were examined there was an anomaly evident. All records showed 10.1.81.11 in the destination field. Different from the URL field where the URL the user is requesting is found, the destination address is the server that is requested to fulfil the request. What would commonly be expected is that this destination field be the content filter server itself. As an example, the destination field in figure 13 shows that the request is to be fulfilled by 10.143.8.20, the content filter. There was another prominent destination listed in the records, that being 10.1.81.11. When searched for (figure 15), any request to a restricted site that was Allowed was directed to 10.1.81.11 and no request to a restricted site (figure 17) that was Allowed was directed to any other proxy, including the content filter proxy. Likewise, any request for a restricted site that was not directed to 10.1.81.11 was correctly *Denied*. This not only shows circumvention, it demonstrates that the mechanism used for circumvention was the upstream proxy, 10.1.81.11, accessed through the content filter proxy. It is interesting to note that within the top 10 sites with the largest number of records are sites such as doubleclick.net, google-analytics.com and googlesyndication.com. These are sites that are generally not accessed or requested directly but rather embedded within and requested as a part of other pages. These entries would likely be included to prevent the loading of advertisements and mitigate data privacy issues. Circumvention of the content filter to access restricted sites has also had the effect of circumventing any protections this content filter had in place (RQ2). When the intended destination of the request was directed through the proxy to 10.1.18.11, the requests were *Allowed* far more often than they *Denied*.

This result confirms the hypothesis H_1 and answers research question RQ1 “Do students successfully circumvent content filtering systems?” in the affirmative.

4.5 Identifying the method of circumvention.

This section is designed to identify any methods of circumvention. Manual examination of the records of restricted sites revealed a common field of all *Allowed* records. The commonality being the destination host field being 10.1.81.11. To confirm this, several commands were run over the restricted site files. This first being (figure 15):

```
cat *.txt | grep Allowed | grep 10.1.81.11 > ./10.1.81.11.txt
```

Figure 15 Shell command - Allowed requests to restricted sites using the upstream proxy

This command put all requests to restricted sites that were *Allowed* and directed to the host 10.1.81.11 into the file 10.1.81.11.txt. This resulted in a file of significant size, 2,643,046 requests (55.69%) for dataset 1 and 1,143,390 requests (38.61%) for dataset 2 of requests to restricted sites that were *Allowed* and used the 10.1.81.11 proxy (figure 16).

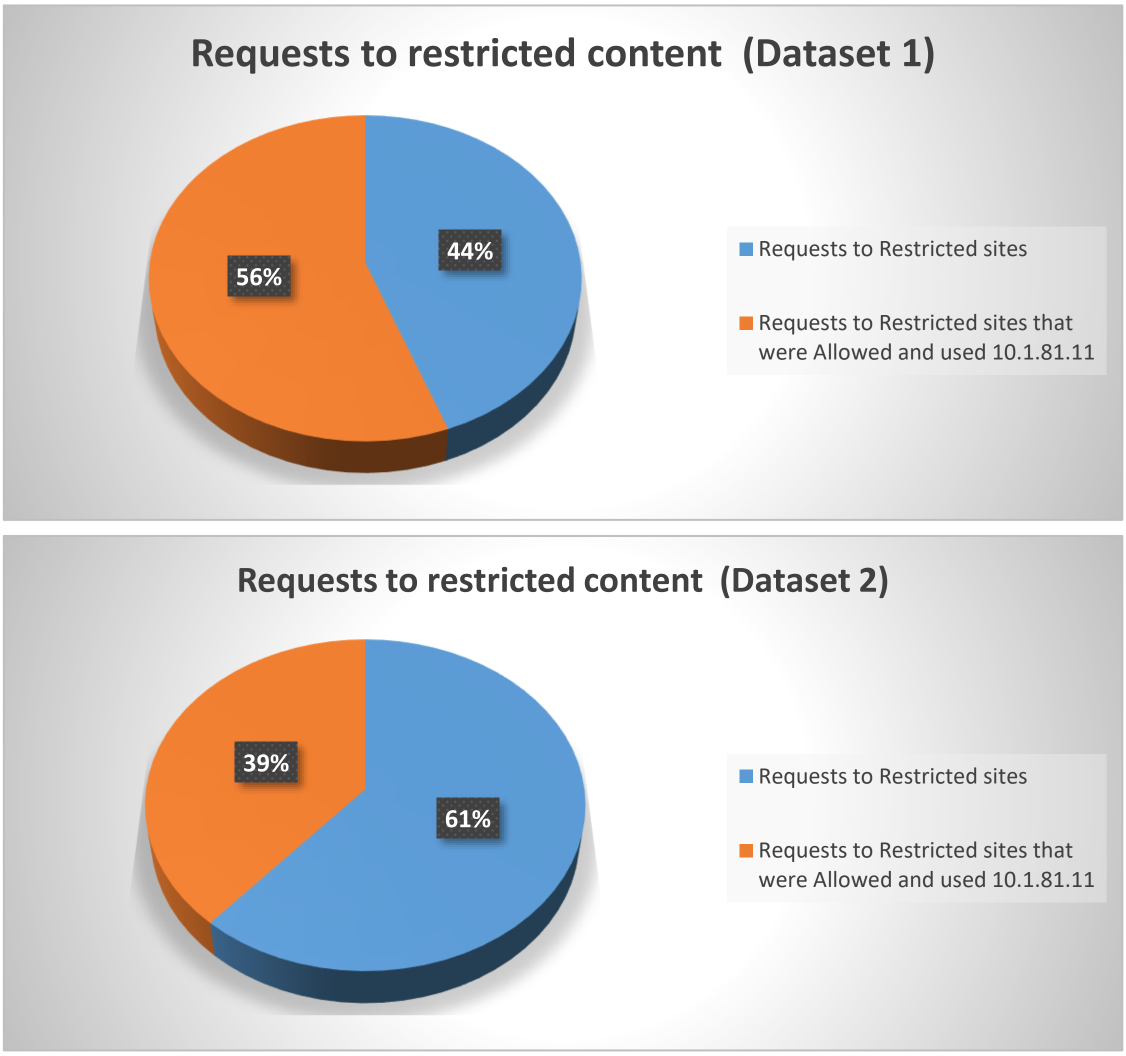


Figure 16 Pie Chart - Requests to restricted content

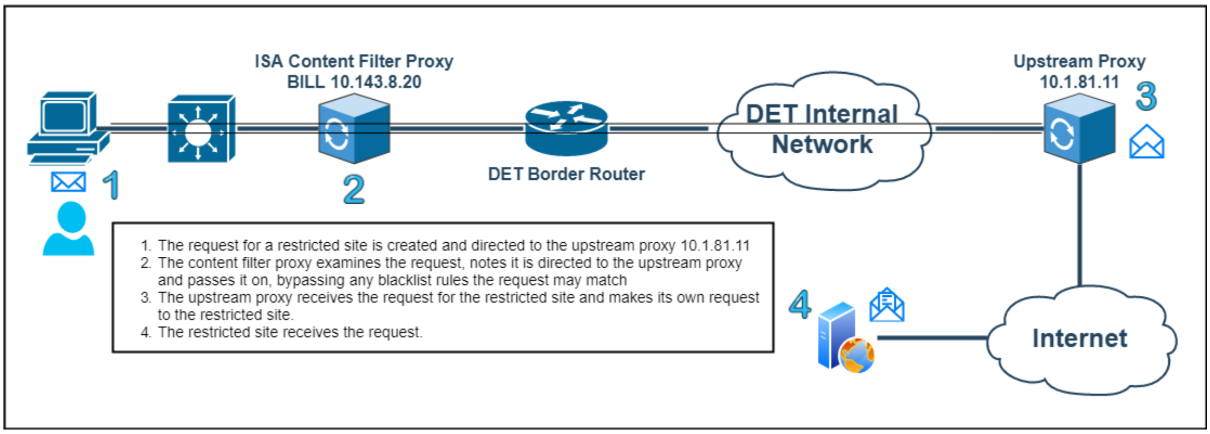
The second command being (figure 17):

```
cat *.txt | grep Allowed | grep -v 10.1.81.11
```

Figure 17 Shell command - Allowed requests to restricted sites not using the upstream proxy

This command selected all requests to restricted sites that were Allowed but without any occurrence of 10.1.81.11. This was an empty list.

Obtaining a large number of requests to restricted sites (55.69% and 38.61%) as a result of the first command, where access was *Allowed* to a restricted site and that request was directed towards the upstream proxy, has identified a circumvention mechanism (figure 18). The second command demonstrated that in no instance was access *Allowed* to a restricted site when that request was directed towards to any host other than the upstream proxy. Additionally, directing a request to a blacklisted site that otherwise enables circumvention would not actually allow circumvention. If requests to the requested site are not forwarded through, then no communication will occur let alone any circumvention of the blacklist. This identified that not only was the primary method of circumvention the upstream proxy 10.1.81.11 but that it was the sole method of circumvention of the content filter being examined.



1. The request for a restricted site is created and directed to the upstream proxy 10.1.81.11
2. The content filter proxy examines the request, notes it is directed to the upstream proxy and passes it on, bypassing any blacklist rules the request may match
3. The upstream proxy receives the request for the restricted site and makes its own request to the restricted site.
4. The restricted site receives the request.

Figure 18 A request directed to the upstream proxy

The next step of the process, according to the workflow (figure 10), was to identify any Base64 encoded records, decode them and examine those for restricted sites. Base64 is a simple form of encoding that can obfuscate a URL request from text pattern matching. The purpose of this step was to check for a simple form of encryption as a possible means of circumvention and if no evidence of circumvention was otherwise discovered, this may be an avenue of circumvention being used. This proved to be a difficult step since Base64 MIME encoding uses standard legitimate characters to encode three bytes as four bytes of text characters. While Base64 decryption is a simple task given a single record, identifying a record as a Base64 encoded from amongst 30-40 million records using pattern matching techniques proved challenging. How filenames are used can vary from site to site.

What version of Base64 a website uses can change if certain characters are valid. Each site can change how that data is formatted for reassembly and decoding at the other end. Finally, how these records can be identified as different from temporary filenames generated from random characters or just normal requests is difficult since Base64 encoded data, by design, uses standard valid characters. The closest to a simple identification pattern is the "=" character that is used to pad any data out to four bytes. Data that ends with "==" is likely to be a Base64 encoded string but it is not a certain identification, and this also misses any Base64 encoded data that uses less or has no padded bytes. Considering the previous results and the likely small impact on these results this effort would entail, this was reduced to the manual decoding of several records. While there were some requests to restricted sites, no decoded record that was *Allowed* to a restricted site was directed to any host other than 10.1.81.11.

That no requests to restricted sites were *Allowed* unless the upstream proxy was used affirms H_8 that Students use a common method to circumvent the content filter.

To confirm, H_6 Students are using URL redirectors to circumvent the content filter, it must be confirmed that the host 10.1.81.11 is the requested URL in the URL field. To do that the restricted site files were searched for 10.1.81.11 in the URL field using the command in figure 19. This yielded no results. This means that while the requests to restricted sites were directed to 10.1.81.11, they were not requested of 10.1.81.11. There was no webpage on 10.1.81.11 that was used by the students to then request another page. This ?? is proxy behaviour and not the behaviour of a redirector. H_6 was proven to be false.

```
cat *.txt | cut -f16 | grep 10.1.81.11 > redirect.txt
```

Figure 19 Shell command - searching for the upstream proxy in the URL field

4.6 Quantifying the degree of circumvention

H_5 states that 80% or more students circumvent the content filter, restated as how prevalent is circumvention of the content filter? Now that the method of circumvention has been identified, the degree can be quantified. To do this the restricted sites log files were searched for 10.1.81.11. The commands in figure 20 counts the number of requests for each unique user; first in the cleaned log files and the second the requests for restricted sites. This resulted in a file with the usernames of each user that used the 10.1.81.11 proxy, one name per line.

For the number of users in the cleaned logs:

```
cat *.txt | cut -f2 | sort | uniq | wc -l
```

For the number of users in the restricted site files, since a request can match more than one site label (using surfagain.com to access facebook.com for example):

```
cat *txt | sort | uniq | cut -f2 | sort | uniq | wc -l
```

Figure 20 Shell command - counting the number of requests for each user. All and restricted sites.

In dataset 1, 4,745,629 requests were made to restricted sites by 1610 students with 2,643,046 requests being *Allowed* from 1589 students. 1589 students attempted to access content through the upstream proxy with 1589 succeeding. With 1672 students in the logs who used the Internet, 96.29% (1610) of students attempted to circumvent the content filter with 95.03% (1589) succeeding (figure 20). Traffic that circumvented the filter with the upstream proxy 10.1.81.11 comprised 96.21% of all traffic but circumvented traffic that was directed towards restricted sites was only 6.36% of all traffic. Circumventing the content filter and comprised 96.21% of all student Internet requests (figure 21).

In dataset 2, 2,690,906 requests were made to restricted sites by 1754 students with 1,143,390 requests being *Allowed* from 1752 students. 1711 students attempted to access content through the upstream proxy with 1523 succeeding. With 1771 students in the logs who used the Internet, 96.61% (1711) students attempted to circumvent the proxy with 85.99% (1523) succeeding (figure 21).

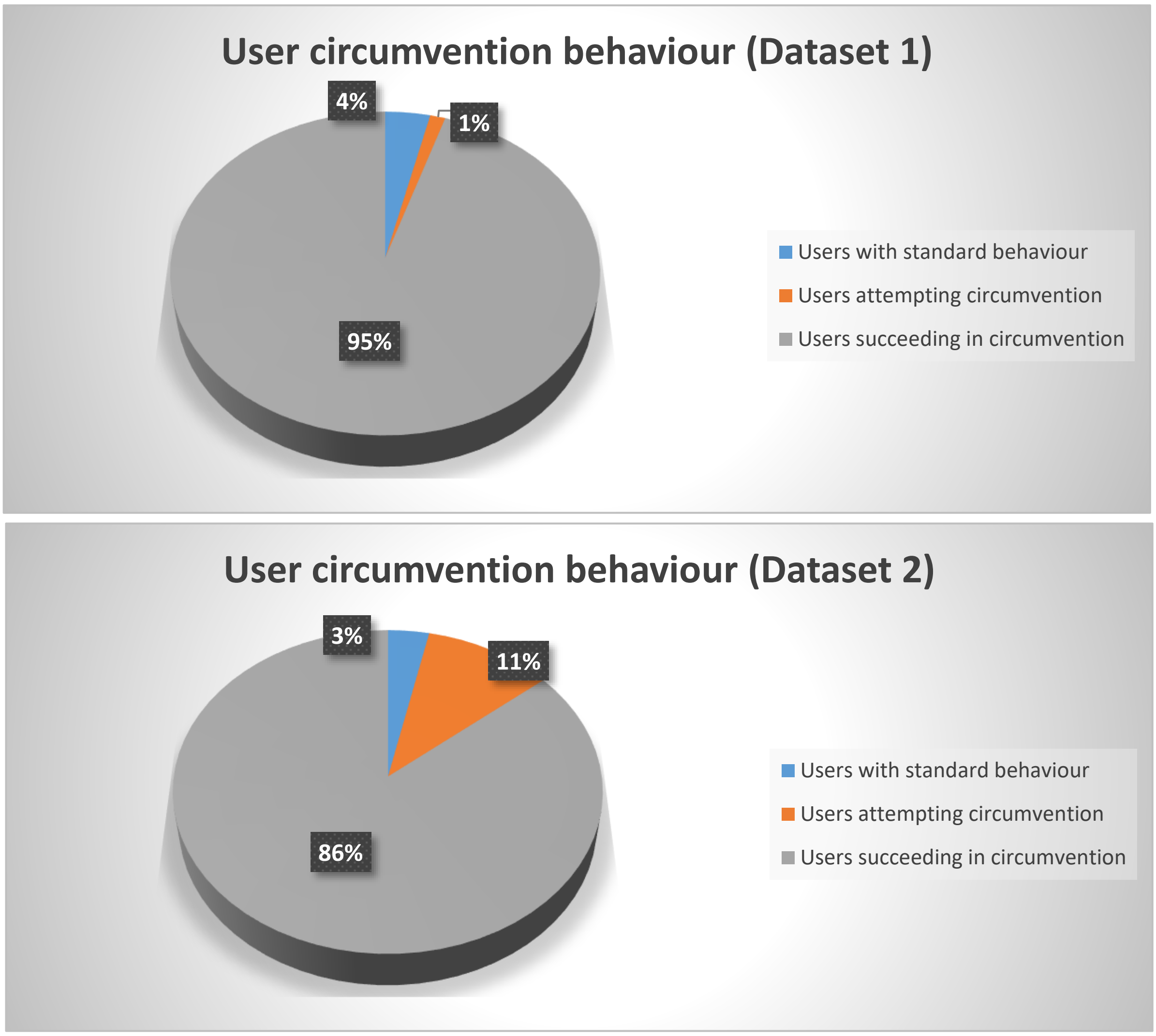


Figure 21 Pie chart - User circumvention behaviour

Traffic that circumvented the filter with the upstream proxy 10.1.81.11 comprised 36.45% of all traffic but circumvented traffic that was directed towards restricted sites was only 3.58% of all traffic (figure 22). Circumventing the content filter and comprised 36.45% of all student Internet requests. Traffic in Term 4 of dataset 2 dropped dramatically which coincided with a change in routing rules that *Allowed* direct access to 10.1.81.11 without having to go through the local proxy. As a result, H_5 holds true for both dataset 1 and dataset 2.

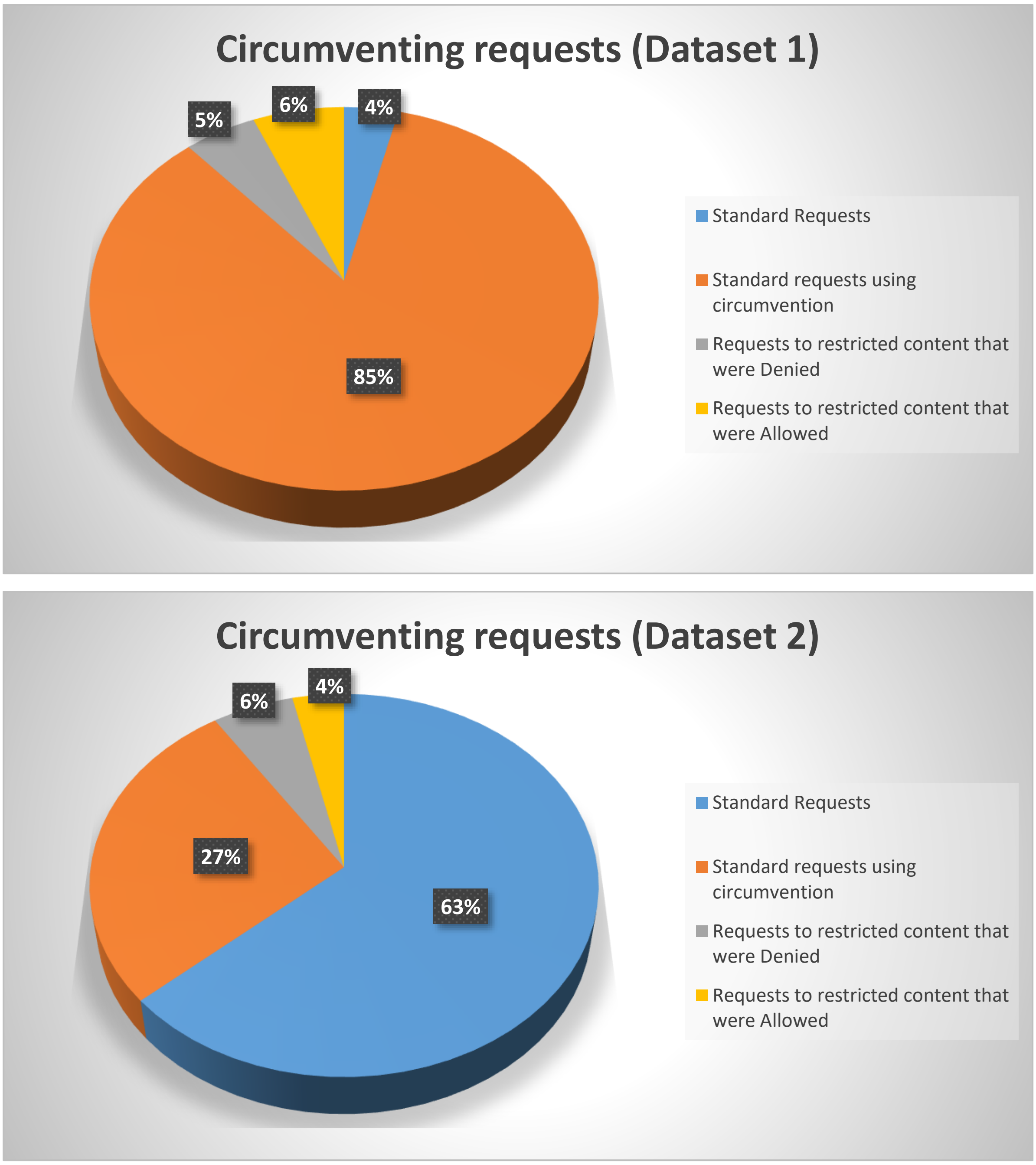


Figure 22 Pie chart - Circumventing requests

The drastic difference in Internet requests directed towards the upstream proxy, 10.1.81.11, between the two time periods indicates a significant difference in the first dataset. The data available does not allow the discovery or analysis of why this difference exists, just that it does. There are many factors that could influence this from an extensive student social network, support of circumvention by teachers, a support technician providing circumvention software or one of many other possibilities.

The large amount of general traffic requests that use the circumvention method of the upstream proxy in relation to requests towards restricted content while using the method suggests an in place tool of some description that sent all URL requests through the upstream proxy, regardless of destination. The discrepancy between all requests and upstream proxy requests could be accounted for by the number of computers this tool was installed on. The large number of users circumventing the filter might be explained as mostly unwitting users using a computer automatically configured to bypass the content filter. The drop in circumvention, in terms of overall requests, in dataset 2 could possibly been explained by most of the computers with the tool having the tool removed. The high number of students using the circumvention method could also be explained by a core number of computers all students would be required to access, say a computer lab, remained active.

All these possibilities are pure speculation as there is nothing in the data that can point to who or how the method of circumvention was disseminated, if the method used involved any level of technical skill or involved elevated privileges to utilise. This thesis does not, and cannot, draw inferences as to how the circumvention was implemented, just that it occurred and what method was used to facilitate it.

H_2 The content filter blocks 95% or more of restricted content when directly accessed or, how effective is the content filter? To do this a list of all requests to restricted content sites that were not directed towards 10.1.81.11 and were also *Denied* is created. Once selected, the URL site is extracted and the number of each *Denied* site was requested. This is done with the command in figure 23:

```
cat *.txt | grep -v 10.1.81.11 | grep Denied | cut -f16 | cut -d ‘/’
-f3 | sort | uniq -c | sort -rn > Denied.txt
```

Figure 23 Shell command - Denied requests that do not use the upstream proxy

Then to get a comparison and create a percentage of blocked sites to unblocked (*Allowed*) sites the command in figure 24 was used. This command, when run over the restricted content requests, excludes all records that were directed towards 10.1.81.11 and then selects those that were *Allowed*.

```
cat *.txt | grep -v 10.1.81.11 | grep Allow
```

Figure 24 Shell command - Allowed requests that do not use the upstream proxy

The process of extracting the requests that were *Allowed* and those requests that were *Denied* separately was going to go further to create a percentage hit : miss ratio. However, dataset 1 and 2 both came back with an empty list. This created a 100% block rate for matches when presented directly to the proxy. This confirms $H_{2.}$

All requests to restricted sites that were directed to the local proxy were properly *Denied*. All requests to restricted sites that were *Allowed* were directed to 10.1.81.11 through the local proxy although not all requests were permitted. The likely explanation for this is the upstream proxy 10.1.81.11 had its own, different, blacklist of sites and while requests to this proxy bypassed the local blacklist, requests were still subject to the restrictions of the fulfilling proxy. This ?? is further supported through the *Denied* requests from the upstream proxy and the subsequent use of redirector sites that would have been *Denied* in the local proxy's blacklist.

H_7 80% or more of students circumvent the content filter for 80% of their web requests or how pervasive is circumvention of the content filter? To test this the process was to, for each user, get the total number of requests and the total number of requests directed to 10.1.81.11 and then calculate the percentage of requests that used the upstream proxy. To do this it was needed to create a separate list of all requests made through the identified proxy 10.1.81.11. This was done with the command in figure 25:

```
cat *WEB_000.wc3.txt | grep 10.1.81.11 > 10.1.81.11.search
```

Figure 25 Shell command - finding all records that use the upstream proxy.

Then count the number of requests each user made in first the cleaned datasets and then the upstream proxy file 10.1.81.11.search. This was done for all requests with the command in figure 26:

```
cat *WEB_000.wc3.txt | cut -f2 | sort | uniq -c | sort -rn >
allrequests.lst
```

Figure 26 Shell command - count the number of requests each user made.

and for the upstream proxy requests with the command in figure 27:

```
cat  10.1.81.11.search  |  cut  -f2  |  sort  |  uniq  -c  |  sort  -rn  >
10.1.81.11.requests.lst
```

Figure 27 Shell command - number of requests each user made using the upstream proxy.

Searching through both the logs and the file of requests that used the upstream proxy allows us to count the number of requests for every user along with the number of requests utilising the upstream proxy. To match the numbers in each file up with the corresponding usernames imported these files into a spreadsheet. Using a vlookup() to match the username in the all request list with the appropriate number of requests for each user in the 10.1.81.11.search list. Once each dataset has both request numbers for each user, the percentage of requests made through the upstream proxy can be calculated as a percentage.

Once this data is populated the number of users is counted along with the number of users with a percentage of 80% or higher. This is then represented as a percentage for each dataset.

In dataset 1, 99.88% of students circumvented the content filter over 80% of their requests (figure 28).

In dataset 2, 62.79% of students circumvented the content filter over 80% of their requests (figure 28).

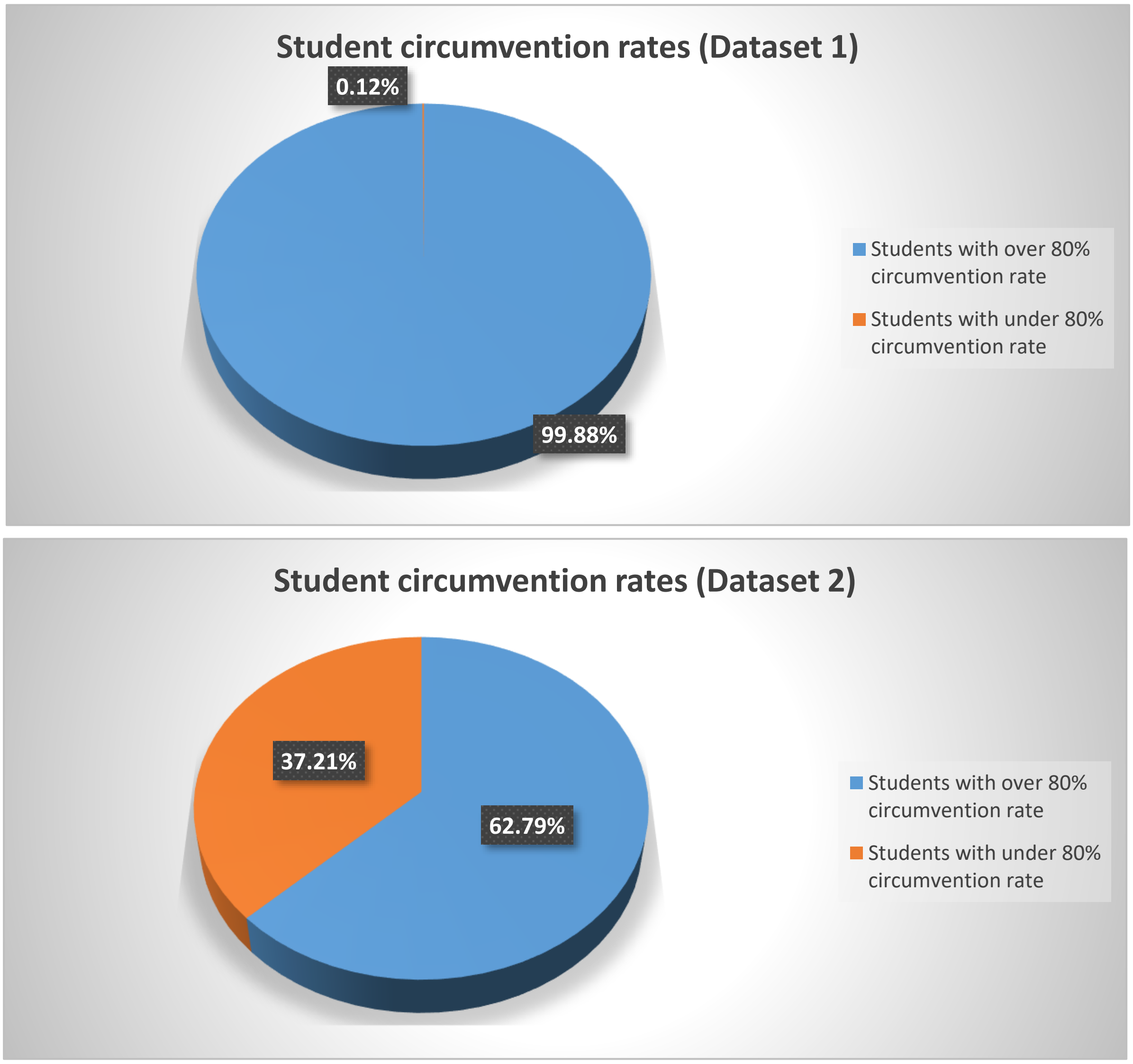


Figure 28 Pie chart - Student circumvention rates

H_7 holds only for dataset 1 but fails for dataset 2. With a circumvention rate of nearly 63% in dataset 2, it is safe to conclude that circumvention is prevalent, and students circumvent the content filter often. So RQ2(a) To what degree do students circumvent content filtering systems? is answered as in 99% and 67% of requests the content filter was circumvented.

4.7 Circumvention resolutions

Several issues arose during this research. Some issues were encountered, for example the unauthorised and untrained alteration of the content filter configuration or the upstream proxy, 10.1.81.11, used as the common circumvention method. These issues and the result that circumvention was achieved by students gives rise to three main challenges and their possible resolutions. These challenges are training, implementation and configuration and monitoring.

4.7.1 Training

The issue of the unauthorised and untrained alteration of the content filter configuration highlights the issue of skills and training. This is evidenced by:

- Someone without appropriate training or knowledge was permitted to alter the configuration of the content filter.
- The content filter was inactive for an extended period unnoticed.
- Once noticed, no-one knew how to correct the configuration of the content filter.
- No-one noticed that the content filter was being circumvented or could show that the content filter was being circumvented.

The flaw in the configuration that allowed the identified method of circumvention, the upstream proxy 10.1.81.11, could have been easily and simply rectified either by disabling the proxy chaining referral or blacklisting the upstream proxy from direct student requests. For this correction to happen someone would have had to have been examining the logs, notice the repeating pattern of the upstream proxy in those logs and then know how to modify the configuration to prevent this method of circumvention.

The most obvious recommendation is training for the content filter administrators. In this environment, the content filter was misconfigured. Content filter applications step their users (administrators in this instance) through the configuration process but if the user does not know the terminology or concepts the application uses, then even an experienced administrator can misconfigure a content filter. Training would make administrators of these devices more likely to configure a content filter correctly and be more likely to know when a content filter was not functioning as it should.

Training was an issue when the data in this research were originally collected in 2010/2011, it was later identified by the Australian Government as a core issue in 2016 <citation>, and continues to be an ongoing issue as identified in 2020. In 2016, five years after the collection of the last of the data examined, the Australian Government released its document Australia’s Cyber Security Strategy (Australian Government, 2016), in which the need for training would feature heavily. The superseding document, Australia's Cyber Security Strategy 2020 (Australian Government, 2020), continues to heavily feature the need for training in its recommendations. Therefore, training has been identified as a key factor in resolving configuration issues.

4.7.2 Implementation and configuration

Often the configuration aids (eg. Microsoft wizard) within content filtering applications, that assist and guide administrators in configuring the content filter, contain with vendor-specific terminology and are not necessarily intuitive. A configuration checking application that could take a configuration file, validate the configuration as valid or invalid and break down the behaviour of the content filter into simple terms would allow administrators to potentially identify misconfigurations. Such a tool could also look for and generate alerts on syntactical misconfigurations and, more importantly, common logical misconfigurations. While logical misconfigurations can be syntactically valid and allow a content filter to run, they can also lead to unintended behaviour of the content filter and undesired outcomes. There are IT systems with configurations that are sufficiently complex to have external configuration validation tools. For an example, Batfish is an open source configuration validation tool for checking network and firewall configurations (Beckett, Gupta, Mahajan, & Walker, 2017). Batfish takes configuration files from multiple vendor specific configurations and converts them into a vendor neutral terms along with the ability to highlight what may be common problems. The existence of an equivalent tool for validating content filter configurations and highlighting potential common configuration problems would improve the efficacy of content filtering systems.

4.7.3 Systems Monitoring

Systems monitoring is a proven area in systems administration but there is currently no commercial offering that monitors traffic for restricted content. This would be a log or packet stream monitoring tool that could examine the traffic of a content filtering system and alert to behaviour that could be deemed undesirable. Log or packet stream monitoring tools exist in many areas of information technology and can trigger alerts on any number of configurable events. Intrusion Detection Systems (IDS) are a likely parallel. IDS function by examining traffic and either comparing that traffic to pre-defined rules, looking for anomalous behaviour or both. Having such a tool in this environment, that could examine the logs or traffic stream for any occurrence of restricted sites or anomalous traffic that could be an indication of circumvention behaviour, would have been able to alert the administrators of the content filter that restricted content was being accessed and could have highlighted the undesirable behaviour. There is significant scope in a content filtering monitor for:

- Development in expert and AI system driven examination.
- Impact on and detection of content filter misconfiguration.
- Implementation oversights for undesired behaviour.

- Implementor confidence in the content filtering system.

A content filter monitor that looks for evidence of circumvention has the possibility to feed back into AI driven content filters themselves. In addition, a valid command given to a tool that produces a result different from the intention of the operator is not an uncommon occurrence. Hence, a content filtering monitor that examines the activity of a content filter rather than the instructions given to a content filter can help draw attention to the behaviour that results from the misconfiguration that lies between instruction and intent. A system monitor of this type could be either a separate program from the content filter or an integrated feature of a more advanced content filter.

Chapter 5 Conclusion

This thesis began with the goals of examining the efficacy of an active content control system in an environment of high school students, if the content control system is being circumvented and if so, what circumvention methods are being used. This chapter summarises the research outcomes, explains how the hypotheses answer the research questions and finishes with a discussion of limitations and future work.

5.1 Research outcomes

Two primary and four sub-questions were created to identify the attempted and actual circumvention of a proxy used as a content filter. Using the research method design, seven hypotheses were developed to examine these questions. Table 8 identifies each research question and its associated hypotheses.

Table 8 Research Outcomes

Research question	**Related Hypotheses**
RQ1 Do students successfully circumvent content filtering systems?	H_1 Students circumvent the content filter. H_2 The content filter blocks 95% or more of restricted content when directly accessed.
RQ1(a) Do students attempt to access restricted content?	H_3 Students attempt to access restricted content H_4 Students attempt to circumvent the content filter.
RQ1(b) How prevalent is content filtering circumvention?	H_5 80% or more students circumvent the content filter

RQ2 What are the techniques used by students to circumvent content filtering?	H_6 Students are using URL redirectors to circumvent the content filter.
RQ2(a) To what degree do students circumvent content filtering systems?	H_7 80% or more of students circumvent the content filter for 80% of their web requests.
RQ2(b) What does the pattern of access describe about circumvention behaviour?	H_8 Students use a common, or small selection, of methods to circumvent the content filter.

Several methods were used to test the stated hypotheses as shown in table 9.

Table 9 Methods and Hypotheses

Method	**Related Hypotheses**
Logged URLs matched against list of restricted sites.	H_3, H_5
List of requested restricted sites searched for records that were *Allowed*.	H_1
Search of requested restricted sites for *Allowed* records that use the identified upstream proxy.	H_8
Search of requested restricted sites for *Allowed* records that do not use the identified upstream proxy.	H_8
Search of upstream proxy in the URL field.	H_6
Search of requested restricted sites for unique users using the identified upstream proxy and compared to the number of total unique users in the unfiltered cleaned logs.	H_5
Search of requested restricted sites for records that do not use the identified upstream proxy. Of these records, the number of *Allowed* records was compared to the number of *Denied* records.	H_2
The unfiltered cleaned logs were searched for the number of requests made by each user and the number of requests made using the identified upstream proxy for each user. These two figures were then used to create a percentage of requests that used the identified upstream proxy for each user. The number of users over 80% were then counted.	H_7

H_3: Do students attempt to access restricted content? was shown to hold true by matching the cleaned logs against the list of restricted sites. Not only were there matches, there were a significant number of matches, 4,745,629 and 2,960,906 requests in total for each dataset, as stated in tables 4 and 5. While not the most important question it is the start of the investigation process because for any of the other hypotheses to be true, students would have had to begin by attempting to access restricted content.

H_4: Do students attempt to circumvent the content filter? held to be true by a search for restricted sites. There were matches for sites known to be proxy sites and matches for *Allowed* access to restricted sites. Matches to various restricted proxy sites demonstrate an attempt to circumvent the proxy content filter. The *Allowed* access to restricted sites through use of the upstream proxy 10.1.81.11 demonstrates not only an attempt to circumvent the proxy content filter but also success in doing so.

H_1: Do students circumvent the content filter? held true by a search for *Allowed* records in the sites found when searching for restricted sites. Once *Allowed* records were found in the logs of restricted sites, this is evidence of the filtering failing. Examination of the *Allowed* records then showed the mechanism of circumvention, that being the upstream proxy 10.1.81.11. This upstream proxy was not restricted by the restricted sites blacklist that was applied to the local content filter and was being used to access content otherwise not permitted by the restricted sites blacklist. Circumvention was shown and the mechanism by which it was circumvented identified.

H_8: Do students use a common, or small selection, of methods to circumvent the content filter? was shown to hold true. With the identification of a mechanism of circumvention, the *Allowed* records to restricted sites was searched for both records that used the upstream proxy 10.1.81.11 and records that excluded the upstream proxy 10.1.81.11. The search for *Allowed* records using 10.1.81.11 resulted in many URL requests (2,643,046 for dataset 1 and 1,143,390 for dataset 2) while the search for *Allowed* records not using 10.1.81.11 returned no records. This shows that the upstream proxy 10.1.81.11 was not only a common circumvention mechanism; it was the only circumvention mechanism with no record that was *Allowed* using anything other than the upstream proxy being found.

H_2: Does the content filter block 95% or more of restricted content when directly accessed? Held true. Once the mechanism of circumvention was found the logs of the restricted sites was searched for all records that did not have the destination address of the upstream proxy 10.1.81.11 in the destination field. The results of this search were in turn searched for *Allowed* records which returned no records. The results in turn were searched for *Denied* records which returned a large number of records. This means that when the upstream proxy was not used to circumvent the proxy content filter, 100% of restricted sites were successfully *Denied*.

H_6: Do students use URL redirectors to circumvent the content filter? was proven false. Once it was found that the sole method to circumvent this proxy content filter was the upstream proxy 10.1.81.11 the requests directed towards the upstream proxy were examined. The URL field was stripped from all records featuring the upstream proxy. The resulting URL field was then examined for 10.1.81.11 to produce any URLs directed towards 10.1.81.11. This search produced no results. This shows that 10.1.81.11 was acting as a proxy and not as a URL redirector. Users were not browsing to 10.1.81.11 to request a restricted site through a webpage interface, as they would with a URL redirector, but rather sending the restricted site request directly to the upstream proxy, as a proxy, through 10.1.81.11.

H_5: Do 80% or more students circumvent the content filter? held true. The number of unique usernames in the cleaned logs was counted to get the total number of users accessing the Internet through this proxy content filter. Then the number of unique usernames found in the restricted sites logs that were also *Allowed* were counted to find the number of students that circumvented the content filter. These results were compared to get the percentage of users that circumvented the content filter. The results of 95.03% and 85.99% of students circumventing the content filter were over 80% in both datasets.

H_7: Do 80% or more of students circumvent the content filter for 80% of their web requests? Was found to be false. This was demonstrated by counting the number of requests each unique user made in the cleaned log files and comparing this against the number of unique requests that were *Allowed* by each unique user in the logs of restricted site request logs. Knowing how many requests each unique user made and how many requests each unique user also made to a restricted site that was successful, this can be converted to a percentage. With each user having a percentage of requests that circumvented the content filter, the number of usernames with a circumvention percentage at or over 80% was counted and compared to the total number of usernames. With 99% of users circumventing the content filter for 80% or more requests in dataset 1 and 63% of users circumventing the content filter for 80% or more requests in dataset 2, this held true only for dataset 1. This only held true for one dataset 1 but since this hypothesis was a general statement, the failure of dataset 2 to reach 80% disproved H_7.

RQ1(a) Do students attempt to access restricted content? was answered by demonstrating hypothesis H_3 and H_4 which affirms that students did attempt to access restricted content. RQ1(b) How prevalent is content filtering circumvention? was answered in the positive by H_5 being shown true with 95% and 86% of students circumventing the content filter. Following from RQ1(a), H_1 Students circumvent the content filter, held true showing that students did manage to circumvent the filter despite H_2, the content filter blocks 95% or more of restricted content when directly accessed, also holding true. Answering RQ1(a) and RQ1(b) though the testing of H_3, H_4 and H_5 and the examination of H_1 and H_2 holds RQ1, do students successfully circumvent content filtering systems? to be true.

RQ2 What are the techniques used by students to circumvent content filtering? was answered by the negative result of H_7. The method of circumvention was not the expected method but rather a second, unrestricted proxy. RQ2(b) What does the pattern of access describe about circumvention behaviour? was answered by H_8, that the method used to circumvent the studied proxy content filter was solely the upstream proxy and the use of the upstream proxy was nigh universal. H_6 Students are using URL redirectors to circumvent the content filter was proven false with H_8 demonstrating that the method of circumvention was not the expected URL redirector. Although H_6 and H_7 were not proven the results yielded from investigating these hypotheses did demonstrate that the circumvention was likely not actively engaged in by all students and likely was permanently configured on most devices that students used to access the Internet. RQ2 What are the techniques used by students to circumvent content filtering? was answered with the proxy rules were bypassed by directing requests through the content filter with a destination of the upstream proxy.

The final lesson learned from this case study is that, in this case, the efficiency and the efficacy of this system varied greatly. The efficiency of this proxy content filter was 100% when operated as intended. This would be the likely result of any test mentioned earlier where sample or pre-recorded traffic would be run through this content filter. The efficiency of this system is drastically different where access when desired, while not consistently ubiquitous, was highly prevalent to the point it could be said to be universal.

5.2 Critical Review

During the course of this research there were several lessons learnt. These have been grouped into the three main areas of implementation, design and methodological.

5.2.1 Implementation

During implementation it was considered appropriate that the data collection system was secure. Physical access and remote access to the content filter proxy was restricted to the Systems Administrators who had agreed to only access the proxy filter for essential operational reasons. This assessment did not account for the new hire of a Systems Manager who, using Administrator privileges, would subsequently alter configurations without sufficient understanding of either the data collection purpose of the content filter or how to configure the content filter correctly.

This mis-configuration led to a time gap in the collected data samples and resulted in separated datasets, 1 and 2 instead of one contiguous dataset that would have covered 17 months. There is a distinct difference in the circumvention rates between dataset 1 (85%) and dataset 2 (27%) with no indication of how this change took place. The impact this had on the research outcomes was that the rate of change in circumvention between dataset 1 and dataset 2 was not observable. A gradual change could have indicated the need for future research into cultural pressure on circumvention. If this was a sudden and rapid change, this could have indicated a technical system change that stopped the enabling of circumvention. Without the intervening data these questions are speculative. In future research, the integrity of the data collection units will need to be more secure. In the context of the content filtering in this research, the Domain Administrator group could have been removed from the Local Administrator group and replaced with the researchers account through a group policy. This would not have prevented a determined and knowledgeable user, but it would slow access, and raise questions for the user as to why such precautions were taken. Consequently, an explanation from the existing staff, to explain the proxy filter's purpose, may have arisen.

5.2.2 Design

The content filter was integrated with the site's Active Directory (AD) which, in practical terms, meant that the content filter could have been configured to add AD information, concerning each student, to the logs. As an example, every log record could have included the student's year group or home room which would have enabled a deeper analysis into what years, or even what home groups, were circumventing the content filter. This may have provided further explanation into how circumvention was used by age or social group. Combined with the possible contiguous 17-month data set that was originally intended to be collected, the potential data by year and time may have revealed other significant insights. Future research should consider what additional information can be gathered and how that could relate a deeper understanding of the research questions.

5.2.3 Methodological

The method chosen for this research was the case study method. This was the appropriate method to answer the research questions. This is because the research questions were questions of state. That is, what is the state of a given condition. A case study provides evidence of the given state but has issues with generalisability. There are no insights into motivation or influencing factors that can be gained through this method, such as:

- Did students choose to circumvent the content filter or was there some existing system condition that enabled this state for them?
- If some students chose to circumvent the content filter, why did they choose to do so?
- If some technical ability was involved in circumventing the content filter, how did the student acquire this knowledge?
- If the technical ability was acquired through other students how did social groups, such as home room, affect the acquisition?
- Did the presence of the proxy content filter and the knowledge it recorded all access to the Internet have any effect on the choice to access restricted content?

Future research should consider what discovery could be provided by the inclusion of additional instruments, such as survey tool for students or a review of the IT environment. This could provide information into the behaviour and decisions that lead to the circumvention of the content filter.

While this research did answer the questions proposed at the outset, the possibility existed in this opportunity to create a deeper and more foundational analysis of the circumvention issue beyond confirming the existence of circumvention itself.

5.3 Limitations and Future Work

This study was a cursory examination of proxy logs to find evidence of circumvention. Tools have moved on and while proxy content filters are still used in some environments, there are more sophisticated tools and techniques. Some environments are likely to accept the network performance penalties that newer tools may impose for the greater difficulty involved in circumventing them. This study was one school with one tool with a particular user base. The results, while informative in identifying the method of circumvention, did not have the data required to identify how students used that method.

There is also a danger if the researcher has an inherent bias that this can lead to poor selection of the data, leading to a preconceived conclusion (Williamson & Johanson, 2013; Jackson, 2015). In this case the data was what was available rather than a selective choice of options. While the data is not chosen to support a particular outcome in this case, care would need to be taken in data collection.

While the approach taken has a solid base to show patterns of behaviour and avoidance that is all it can do. The deeper questions as to motivation or predisposition are not addressed. Also not addressed are any solutions to any issues that might be uncovered. The study is a look into existing behaviour trends only and does not purport to be exhaustive of the methods that could be used, what variables could be used to control outcomes or the appropriateness of network-based URL filtering as a content filtering tool. Additionally, as a case study, what this research reveals is not necessarily representative of all high schools.

Replication of this study in other high schools would be required to allow conclusions to be generalised beyond a single case. This may or may not introduce more advanced methods of content restriction and circumvention methods. Studies may also need to be broadened into the behaviours that lead to the desire for circumvention. The more samples that can be examined, the better the problem can be understood. As a single case study, any conclusions drawn can only apply to this one single environment. Additional further work in this area could be undertaken from several given viewpoints.

From the perspective that content filters are a solution then this case study has shown that there are weaknesses that, not only can be but, are exploited to circumvent content filters. In this case a quick configuration change of blocking the upstream proxy from direct access from student requests in the blacklist would have prevented this exploit. This does not negate that other exploits exist or that the students in this case study would not have found another exploit had this exploit been addressed.

Further work could revolve around configuration tools or assistants that could look for possible misconfigurations and help in correcting such mistakes.

From the perspective that content filters do not work, research could include examining why students seek to circumvent a content filter or if education coupled with psychological resilience would be a better tool for dealing with the harm some believe can be inflicted by the consumption of inappropriate content accessed on the Internet. There is evidence provided in this thesis that content filters are circumvented and there is an argument to be made that it is prudent to assume that if a content filter is employed, it will be circumvented at some point.

Finally, from the perspective that content filters may not be a perfect defence, findings in this thesis suggest that the mere known presence of a content filter is a phycological deterrent. That the majority of student requests bypassed the content filter in both datasets but the requests for restricted content were only 4% - 6% of all requests suggests that while students could get to restricted content, students refrained from accessing restricted content. The role of a content filter as a deterrent rather than as a barrier in influencing student behaviour is a question worth pursuing.

Chapter 6 Appendices

Appendix A – Partial listing of the blacklist

The full blacklist is available upon request.

djhybridstorm.com
dongtaiwang.com
dontfilter.us
dontshowmyip.info
doubleclick.com
doubleclick.net
download.im1music.net
download-free-games.com
downloadgames2.com
dragongamez.com
drprox.com
dsouth.net/anon
dtunnel.com
duproxy.com
dutysite.info
dzzt.com
easymusicdownload.com
easypro.cz.cc
easyproxy.org
easysurf.com
easyvisit.info
eatmoreblueberries.com
eatmybrowser.com
eazysurfer.com
ebay.com
ebay.com.au
ebuddy.com
ebutechnologies.com
ebypass.org
egogo.ru
elanceconnect.com
eliminatespam.info
elitegate.info
embedproxies.com
emil-zittau.de
emyspaceunblocker.info
enstealth.com
e-ronin.com
etype.hostingcity.net
eurostretch.ru
evilsprouts.co.uk
exoproxy.com
f2.chinoxy.com
f2s.com
facebook.com
facebook-proxy.me
facehide.com
fast138.tripod.com
fastclick.net
fastgames.com
fastproxynetwork.com
fax-gateway.info
fearofmidgets.com/cgiproxy
fearofmidgets.com/cgiproxy/nph-proxy.pl/000010A/http/www.web.freerk.com/c/
feathermud.com
ferienwohnung-in-masuren.de
fightclubvideos.com
filebahn.com
filehippo.com
filter2005.com
filtersnoop.info
finderly.net/proxy
findnot.com
findproxy.org
finxe.com
firefox.con
fireprox.com
firewalldown.net
flasharcade.com
flash-game.net
flashgamecodes.com
flashgames.com
flashgames247.com
flaxads.com
floon.com
flylikeaturtle.com
footballscorelive.com
forumwhore.com
fosho.us
foxyproxy.net
fpflashfarm.com
fr.search.yahoo.com/images
free2.surffreedom.com/nph-free.cgi
free2surf.org
freebieproxy.com
freedom2surf.net
freedomdown.net
freefronthost.com
freegamesonline.dk
freeonlinegames.com
freeonlineproxy.com
freepr0xy.com
freeproxy.ru
freeproxy.us
freeproxyserver.net
freeproxysite.com
freesteam.org
freeunblocker.net
freeusaproxy.com
freevisit.info
freewebarcade.com
freewebproxy.org
frozenraindrop.com
fsurf.com
ft888.net
ftpplanet.com
ftunnel.com
fxprofile.com
gamefudge.com
gamegarage.co.uk
gamegecko.com
gamegeko.com
gamersbanner.com
gamershell.com
gamesbannernet.com
gamesproxy.com
gasterixx.de
gatekeeper.rdi-electronics.com
gator.com
getpast.com
getproxies.be
getunblocked.info
ghzm.com
gmail.com:443
gnet30.gamesnet.de
go.icq.com
go2-vn1.appsport.com
go-fish.info
goldfishandchips.co.uk
goldproxies.info
google.com/+

google.com/+:443
google-analytics.com
googlehammer.com
googlesyndication.com
goproxing.com
gosneak.com
gouc.fr
gravatar.com
greatproxy.net
greenrabbit.com
greenrabbit.org
gridironglory.com
groothuijsen.nl
grooveshark.com
gsfiles.com
guardster.com
gunshin.net/cgi-bin/nph-proxy.cgi
hackit.us
hacksurfing.com
halomovies.org
handsoffmycomputer.com
hannes-wacker.de
happyface.brainsoft.biz
heartsmoke.com
heartsmoke.com/
heavygames.com
hentaiclips.us
heshan18.com

Appendix B – ISA field descriptions.

Client IP Address: The IP Address to the Client requesting the resource

Username: The Username of the account making the request. This can vary on the type of authentication configured or even if authentication is configured. If authentication is not used then ANONYMOUS is used instead.

User Agent String: This is the name and version of the browser or application sent in the HTTP agent header.

Date: The date the event was logged in the format of YEAR-MONTH-DATE

Time: The time the event was logged in the format of HOUR:MINUTE:SECOND

Server: The name of the ISA server logging the event

Referring Server: Reserved – Not used

Destination Host: The domain name, or if unresolved the IP Address, of the remote server that is to provide the service. A – in this field indicates that this was pulled from the local cache.

Destination Host IP: The IP Address, of the remote server that is to provide the service

Destination Port: The port number to be used to connect to the requested service on the remote server.

Processing Time: The time in milliseconds taken to process the connection.

Bytes Received: The number of bytes sent from the remote server and received by the client.

Bytes Sent: The number of bytes sent from the client and sent to the remote server.

Protocol: The protocol used to fulfil the request. Most likely values are HTTP, HTTPS, or FTP.

Operation: The action requested through the protocol. Values are most likely HTTP actions such as GET, POST, PUT, etc.

URL: The URL requested

MIME Type: The MIME type used if any MIME encoding is used

Result Code: A cumulative numeric code used to represent a number of error or status conditions for protocol transmissions.

Rule: The name of the rule matched in the ISA proxy configuration for this request

Filter Information: A cumulative numerical code used to feedback technical information on the request status.

Source Network: The network label from which the request came from.

Destination Network: The network label to which the request is, or in the case of a denial would be, sent.

Error Code: A cumulative numerical error code for ISA processes.

Action: The action taken as a result of the rule matched. This is most likely *Allowed* or *Denied*.

Appendix C – siteparse.sh

```
#!/bin/bash
   #set -x
   START=$(date +%d/%m:%H:%M:%S)
   printf "Processing started at $START\n"
   printf "Filename: $1\n"
   exec<$1
   printf "****\n"
   while read LINE
   do
         LINE=`echo $LINE | sed 's/\\r//g'`
            printf "Entry: $LINE\n"
         FILENAME="${ LINE }.txt"
         echo -e "$FILENAME --\n "
         printf "Searching for occurrences of %s in logfiles.
Output: %s\n" "$LINE" "$FILENAME"
         fgrep "$LINE" *.w3c.txt > ./$2/$FILENAME
         WORD_FREQUENCY=$LINE"_users.txt"
         printf "Searching for user frequency in data subset file
./$2/$FILENAME. Saving in ./$2/$WORD_FREQUENCY\n"
         cat ./$2/$FILENAME | cut -f2 | sort | uniq -c | sort -nr >
./$2/userparse/$WORD_FREQUENCY
         printf "****\n"
   done
   FINISH=$(date +%d/%m:%H:%M:%S)
   RUN="$(($FINISH-$START))"
   printf "Processing finished at $FINISHED with a run time of
$RUN\n"
```

Appendix D – Partial listing of the original blacklist: FirewallPolicy.xml

```
<fpc4:URLSet StorageName="{EDAF1420-2B36-4AA3-A83D-BB1C5570D6F7}" StorageType="2">
        <fpc4:Name dt:dt="string">Inappropriate</fpc4:Name>
        <fpc4:Predefined dt:dt="boolean">0</fpc4:Predefined>
        <fpc4:URLStrings>
                …
                <fpc4:Str dt:dt="string">*imhaha.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*paypal.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*torrentbay.de</fpc4:Str>
                <fpc4:Str dt:dt="string">*hotgamestown.com</fpc4:Str>
                <fpc4:Str dt:dt="string">mygirlyspace.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*4players.de</fpc4:Str>
                <fpc4:Str dt:dt="string">*arcadejoint.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*jayisgames.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*loadfreegames.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*fastgames.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*2keygames.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*thatvideosite.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*smashingames.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*transformersgame.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*onlinefreegaming.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*partypoker.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*music.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*arcadebomb.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*ftpplanet.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*thegamehomepage.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*notdoppler.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*bebo.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*videocopilot.net</fpc4:Str>
                <fpc4:Str dt:dt="string">*facebook.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*stoptazmo.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*promtgames.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*onemanga.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*stopazmo.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*theereadingroom.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*proxybomb.net</fpc4:Str>
                <fpc4:Str dt:dt="string">*reallyfunarcade.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*promptgames.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*thespectrum.net</fpc4:Str>
                <fpc4:Str dt:dt="string">*three.com.au</fpc4:Str>
                <fpc4:Str dt:dt="string">*agame.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*actionflash.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*youtube.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*vista-server.com</fpc4:Str>
                <fpc4:Str dt:dt="string">*bungie.org</fpc4:Str>
                <fpc4:Str dt:dt="string">*uploaded.to</fpc4:Str>
                <fpc4:Str dt:dt="string">*youtube.com</fpc4:Str>
```

Appendix E – Sanitize.sh

```
#!/bin/bash

THISDIR=$1

for filename in ./*.w3c; do

    FILE=$filename".txt"

    echo "Processing $filename";

    cat $filename | awk -F $'\t' '!/\$/ {print $0}' | awk -F $'\t' '!/^#/
{print $0}' | grep -v [Aa]nonymous | grep -v <Admin Domain> | grep -v
'<Domain>\\<School number>' | grep -v '<Domain>\\[Ee][0-9]' | grep -v
<Administrator 1> | grep -v < Administrator 2> | grep -v < Administrator
3> | grep -v '<Domain>\\<Administrator 4>' | grep -v '<Domain>\\vmadmin' |
grep -v '<Domain>\\teststaff' | grep -v '<Domain>\\administrator' | grep -
v '<Domain>\\< Administrator 5>' | grep -v '<Domain>\\ICL' | grep -v
'<Domain>\\<Service Account>' | grep -v cafe | >
../$THISDIR.cleaned/$FILE;

done
```

References

ACMA. (2008). *Closed Environment Testing of ISP-Level Internet Content Filtering*. Retrieved from https://www.academia.edu/7988830/Closed_Environment_Testing_of_ISP_Level_Internet_Content_Filters_Report_to_the_Minister_for_Broadband_Communications_and_the_Digital_Economy

Aktay, S. (2018). Teacher Perspective on Internet Censorship in Turkey. *Universal Journal of Educational Research, 6*(2), 296-306.

Al-Hajery, E. S. (2000). *Evaluating Web filters: A practical approach.* INET 2000.

Al Mugni, A., Herdiansah, M. F., et al. (2019). *DNSBL for Internet Content Filtering Utilizing pfSense as The Next Generation of Opensource Firewall.* 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI).

Australian Federal Government. (2018). *BROADCASTING SERVICES ACT 1992 - SCHEDULE 5 Online services* Australian Federal Government Retrieved from https://www.legislation.gov.au/Series/C2004A04401.

Australian Government. (1992). Broadcasting Services Act 1992. Retrieved from http://www.austlii.edu.au/au/legis/cth/consol_act/bsa1992214/

Australian Government. (2008a). Parliamentry Debates, Senate, 3 December 2008. Retrieved from http://www.aph.gov.au/hansard/senate/dailys/ds031208.pdf

Australian Government. (2008b). Standing Committee on Environment, Communications and the Arts, Supplementary Budget Estimates, 20 October 2008. Retrieved from http://www.aph.gov.au/hansard/senate/commttee/S11346.pdf

Australian Government. (2016). *Australia's Cyber Security Strategy*. Internet: Australian Government Retrieved from https://www.homeaffairs.gov.au/cyber-security-subsite/files/PMC-Cyber-Strategy.pdf.

Australian Government. (2020). *Australia's Cyber Security Strategy 2020*. Internet: Australian Government Retrieved from https://www.homeaffairs.gov.au/cyber-security-subsite/files/cyber-security-strategy-2020.pdf.

Ayre, L. (2012). Filtering Worst Practices: Keyword Filtering and Blocking by File Type Retrieved from https://galecia.com/blogs/lori-ayre/filtering-worst-practices-keyword-filtering-and-blocking-file-type

Bambauer, D. E. (2013). Censorship v3. 1. *IEEE Internet computing, 17*(3), 26-33.

Beazley, K. (2006, 2006). Labor's Plan To Protect Kids From Internet Pornography. Retrieved from http://web.archive.org/web/20060422120043/http://www.alp.org.au/media/0306/msloo210.php

Beckett, R., Gupta, A., et al. (2017). *A general approach to network configuration verification.* Proceedings of the Conference of the ACM Special Interest Group on Data Communication.

Beins, B. C. (2004). *Research Methods: A tool for life*: Pearson Education.

Beins, B. C. (2012). *Research Methods: A tool for life*. Boston, MA: Pearson Higher Ed.

Best, J. (2007). Teen cracks AU$84 million porn filter in 30 minutes. Retrieved from https://www.zdnet.com/article/teen-cracks-au84-million-porn-filter-in-30-minutes/

Chen, J., & Nguyen, U. T. (2018). *A Robust Protocol for Circumventing Censoring Firewalls.* 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData).

Chou, L.-D., He, Z., et al. (2012). *Design and implementation of content-based filter system on embedded linux home gateway.* 2012 14th International Conference on Advanced Communication Technology (ICACT).

Clayton, R. (2006). Failures in a Hybrid Content Blocking System *Privacy Enhancing Technologies* (Vol. 3856/2006, pp. 78-92): Springer Berlin / Heidelberg.

Collins, L., Love, P., et al. (2008). *FEASIBILITY STUDY ISP LEVEL CONTENT FILTERING*. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.168.5185&rep=rep1&type=pdf

Condie, J. (2012). Researching people using questionnaires and interviews. Retrieved from https://www.slideshare.net/jennacondie/researching-people-using-questionnaires-and-interviews

Conroy, S. (2006). Coonan Out Of Touch On Porn Filtering. Retrieved from http://web.archive.org/web/20080104112145/http://www.alp.org.au/media/0406/mscomit120.php

Conroy, S. (2007a). Labor's Plan for Cyber-safety. Retrieved from http://www.alp.org.au/download/now/labors_plan_for_cyber_safety.pdf

https://s3-ap-southeast-2.amazonaws.com/ad-aspi/2018-07/2007-Labor-Cybersecurity-Policy-Statement-Labors-Plan-for-Cyber-safety.pdf?4.KU86Jun1uWOhBfUk.PYtx_mU5_AXBm

Conroy, S. (2007b). Labor’s Plan for Cyber Safety. *Election.* Retrieved from http://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;query=Id%3A%22library%2Fpartypol%2FKOZO6%22

Conroy, S. (2012). Child abuse material blocked online, removing need for legislation [Press release]. Retrieved from https://web.archive.org/web/20121114042606/http://www.minister.dbcde.gov.au/media/media_releases/2012/180

Coonan, H. (2007, 10 August). NetAlert: Protecting Australian Families Online. Retrieved from http://www.minister.dcita.gov.au/coonan/media/media_releases/netalert_-_protecting_australian_families_online

DBCDE. (2009). Online Content Regulation. Retrieved from http://www.dbcde.gov.au/internet/online_content_regulation

Department of Education. (2008). Students Online. Retrieved from https://web.archive.org/web/20100329033635/http://www.det.wa.edu.au/policies/detcms/policy-planning-and-accountability/policies-framework/policies/students-online.en?oid=au.edu.wa.det.cms.contenttypes.Policy-id-3784406

Department of Education. (2019). *Students Online in Public Schools Policy v3.1*. Retrieved from http://det.wa.edu.au/policies/detcms/policy-planning-and-accountability/policies-framework/policies/students-online-in-public-schools-policy.en?cat-id=3457100.

Dixon, L., Ristenpart, T., et al. (2016). Network traffic obfuscation and automated internet censorship. *IEEE Security & Privacy, 14*(6), 43-53.

Durumeric, Z., Ma, Z., et al. (2017). *The Security Impact of HTTPS Interception.* NDSS.

eSaftey Commisioner. (2020). Illegal and harmful content. Retrieved from https://www.esafety.gov.au/key-issues/Illegal-harmful-content

Faisal, I., & El-Kassas, S. (2018). *Limited Proxying for Content Filtering Based on X. 509 Proxy Certificate Profile.* International Conference on Security for Information Technology and Communications.

Forte, M., Souza, W. L. d., et al. (2006). *A content classification and filtering server for the Internet*. Paper presented at the Symposium on Applied Computing, Dijon, France. http://delivery.acm.org/10.1145/1150000/1141553/p1166-forte.pdf?key1=1141553&key2=7703588421&coll=GUIDE&dl=GUIDE&CFID=45666207&CFTOKEN=90958713

Fourie, I., Bothma, T. J., et al. (2013). Trends in transition from classical censorship to Internet censorship: selected country overviews. *Innovation: journal of appropriate librarianship and information work in Southern Africa, 2013*(46), 166-191.

Freedom House. (2018). The Rise of Digital Authoritarianism. Freedom on the Net 2018: New York.

Frost, C. (2020). Web Content Filtering and Security. Retrieved from https://support.opendns.com/hc/en-us/articles/227988047-Web-Content-Filtering-and-Security

Galliers, R. (1992). *Information systems research: Issues, methods and practical guidelines*. Oxford: Blackwell Scientific.

Graham, I. (2008). AU Gov't Mandatory ISP Filtering / Censorship Plan. Retrieved from http://libertus.net/censor/ispfiltering-au-govplan.html

Greenfield, P., Rickwood, P., et al. (2001). *Effectivness of Internet Filtering Software Products*. Retrieved from http://www.security-science.com/pdf/effectiveness-of-internet-filtering-software-products.pdf

Hamilton, C. (2009, Feburary 16). Web doesn't belong to net libertarians. Retrieved from https://www.theaustralian.com.au/business/technology/web-doesnt-belong-to-net-libertarians/news-story/1a43bbdc8df477bbc10cc771b2e082ed

Hammami, M., Chahir, Y., et al. (2005). Webguard: A web filtering engine combining textual, structural, and visual content-based analysis. *IEEE Transactions on Knowledge and Data Engineering, 18*(2), 272-284.

Hills, E. (2018). 2. A Survey on the Cybersecurity of K-12 Schools.

Hosting Tribunal. (2020). How Many Websites Are There? How Many Are Active in 2020? Retrieved from https://hostingtribunal.com/blog/how-many-websites/

Hunter, C. D. (2000a). *Internet filter effectiveness (student paper panel): testing over and underinclusive blocking decisions of four popular filters.* Proceedings of the tenth conference on Computers, freedom and privacy: challenging the assumptions.

Hunter, C. D. (2000b). Social Impacts: Internet Filter EffectivenessŠTesting Over-and Underinclusive Blocking Decisions of Four Popular Web Filters. *Social Science Computer Review, 18*(2), 214-222.

Internet Industry Association. (2008). *Internet Industry Code of Practice Contenet Services Code*. Retrieved from https://www.commsalliance.com.au/Activities/ispi.

IRMA. (2019). *Internet and Technology Addiction: Breakthroughs in Research and Practice: Breakthroughs in Research and Practice*. Hershey, PA: IGI Global.

Jackson, S. (2015). *Research methods and statistics: A critical thinking approach*. Boston, MA: Cengage Learning.

Jakub, D., Lex, G., et al. (2018). Planet Netsweeper. Retrieved from https://citizenlab.ca/2018/04/planet-netsweeper/

Johnson, B. (2008, 8 December). Wikipedia falls foul of British censors. *The Guardian*. Retrieved from http://www.guardian.co.uk/technology/2008/dec/08/wikipedia-censorship

Keijser, J. J. (2017). *OpenVPN Cookbook*. Birmingham, UK: Packt Publishing Ltd.

Klang, M., & Murray, A. (2016). *Human rights in the digital age*. London, UK: Routledge.

Lai, Y., Ma, Q., et al. (2010). *Framework of Web content filtering for IPv6.* High Performance Computing and Simulation (HPCS), 2010 International Conference on.

Lake, C. (2009, 24 Feburary 2009). Web filter debate descends into slanging match at Kickstart Forum 2009. *news.com.au,* p. 1. Retrieved from http://www.news.com.au/technology/story/0,28348,25100456-5014239,00.html

Leitch, S., & Warren, M. (2015). Applying classification controls to Internet content in Australia. *Journal of Information, Communication and Ethics in Society, 13*(2), 82-97. doi:doi:10.1108/JICES-08-2014-0037

Malone, M. (2009, 10 Feburary). Education, not filtering, the answer: iiNet. *Australian IT*. Retrieved from http://www.australianit.news.com.au/story/0,24897,25035546-5013046,00.html

McMenamin, B. (2009, 5 Feburary). Why Australia needs to trial net filters. Retrieved from https://womensphere.wordpress.com/2009/02/14/why-australia-needs-to-trial-net-filters/

Mind Chasers Inc. (2019). Create your own Web Content Filter with Squid and Linux Retrieved from https://mindchasers.com/dev/app-squid-redirect

Molina, M. D., Gambino, A., et al. (2019). *Online Privacy in Public Places: How do Location, Terms and Conditions and VPN Influence Disclosure?* Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems.

Mou, Y., Wu, K., et al. (2016). Understanding the use of circumvention tools to bypass online censorship. *new media & society, 18*(5), 837-856.

NAIRN, G. (2007). NetAlert—Protecting Australian Families Online. Canberra: Department of Communications. *Information Technology and the Arts*.

Narayanan, B. K., Moses, S., et al. (2018). Adult content filtering: Restricting minor audience from accessing inappropriate internet content. *Education and Information Technologies, 23*(6), 2719-2735.

Newton, M. (2009, 12 Feburary). ISP filtering not scalable: SAGE. *Australian IT*. Retrieved from http://www.australianit.news.com.au/story/0,24897,25045088-5013038,00.html

Palfrey, J., Roberts, H., et al. (2009). Circumvention landscape report: Methods, uses, and tools.

Panchakarla, B. P. (2019). *Design and Implementation of Firewall to Inspect Traffic in Encrypted VPN Tunnels.* University of Missouri--Kansas City.

Patel, O. P., Bharill, N., et al. (2019). Advanced Quantum Based Neural Network Classifier and Its Application for Objectionable Web Content Filtering. *IEEE Access, 7*, 98069-98082.

Paula, P., & Rhonda, J. (2014). *Australian Governments and dilemmas in filtering the Internet: juggling freedoms against potential for harm*. Retrieved from https://parlinfo.aph.gov.au/parlInfo/download/library/prspub/3324230/upload_binary/3324230.pdf;fileType=application/pdf

Polpinij, J., Chotthanom, A., et al. (2006). *Content-based text classifiers for pornographic web filtering.* 2006 IEEE International Conference on Systems, Man and Cybernetics.

Pons-Salvador, G., Zubieta-Méndez, X., et al. (2018). Internet Use by Children Aged six to nine: Parents' Beliefs and Knowledge about Risk Prevention. *Child indicators research, 11*(6), 1983-2000.

Qiu, J. L. (1999). Virtual censorship in China: Keeping the gate between the cyberspaces. *International Journal of Communications Law and Policy, 4*(Winter), 125.

Reis, F., Godinho de Matos, M., et al. (2017). The Impact of DNS Blocks on Digital Piracy Activity.

Reshet, N. (2015). The Effectiveness of Filtering. Retrieved from http://www.netivei-reshet.org/en/node/103

Rosenberg, R. S. (2001). Controlling access to the Internet: the role of filtering. *Ethics and Information Technology, 3*(1), 35-54.

Rowe, M., & King, R. (2015). An investigation into the performance of UK internet providers' web filters.

Schofield, J. (2008, 8 December). Wikipedia page censored in the UK for 'child pornography'. *The Guardian*. Retrieved from http://www.guardian.co.uk/technology/blog/2008/dec/08/internet

Sovran, Y., Libonati, A., et al. (2008). *Pass it on: Social Networks Stymie Censors*. Paper presented at the International Workshop on Peer-to-Peer Systems, Tampa Bay, Florida. http://www.cs.toronto.edu/iptps2008/final/73.pdf

Stark, P. B. (2007). The Effectiveness of Internet Content Filters (pp. 19). University of California, Berkeley: University of California, Berkeley.

Stem, L. E. (2017). Censorship: Filtering Content on the Web. *The Southeastern Librarian, 64*(4), 3.

The Citizen Lab. (2007). *Everyone's Guide to By-passing Internet Censorship for Citizens Worldwide*. Retrieved from

The Detail. (2020, June 10). NZ's new internet laws: Censorship or necessary for our safety? *NZ Herald*. Retrieved from https://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=12338595

Tomsho, G. (2019). *Guide to networking essentials* (8th ed.). Boston, MA: Cengage Learning.

Trabelsi, Z., Zeidan, S., et al. (2016). *Network packet filtering and deep packet inspection hybrid mechanism for IDs early packet matching.* Advanced Information Networking and Applications (AINA), 2016 IEEE 30th International Conference on.

Vicks, M. E. (2013). An examination of internet filtering and safety policy trends and issues in south carolina's k-12 public schools.

Weiss, A. (2009). A Digital Trail is Forever. *netWorker, 13*(2), 14-19.

Wiley, B. K. (2016). *Circumventing network filtering with polymorphic protocol shapeshifting*. Cambridge, MA.

Williams, P., & Dillon, K. (1998). The internet and the law: Emerging issues for Australian schools. *Austl. & NZJL & Educ., 3*, 3.

Williamson, K. (2002). *Research methods for students, academics and professionals: Information management and systems*. Melbourne, Vic: Elsevier.

Williamson, K., & Johanson, G. (2013). *Research methods: Information, systems and contexts*. Cambridge, MA: Tilde University Press.

Yeop, Y. H., Othman, Z. A., et al. (2018). *Key Factors to Implement BYOD in Schools.* 2018 Cyber Resilience Conference (CRC).

Yin, R. K. (2013). *Case study research: Design and methods*: Sage publications.

Zittrain, J., & Edelman, B. (2003). Internet filtering in China. *IEEE Internet computing, 7*(2), 70-77.