University of Crete
Computer Science Department

CensMon :
A Web Censorship Monitor

Andreas Sfakianakis

Master's Thesis

February 2012
Heraklion, Greece

University of Crete
Computer Science Department

CensMon
A Web Censorship Monitor

Thesis submitted by

Andreas Sfakianakis

in partial fulfilment of the requirements for the
Master of Science degree in Computer Science

THESIS APPROVAL

Author: ______________________________
Andreas Sfakianakis

Committee approvals: ______________________________
Evangelos P. Markatos
Professor, Thesis Supervisor

____
Sotiris Ioannidis
Principal Researcher

____
Maria Papadopouli
Assistant Professor

Departmental approval: ______________________________
Angelos Bilas
Professor, Chairman of Graduate Studies

Heraklion, February 2012

Abstract

The Internet has traditionally been the most free medium for publishing and accessing information. It is also quickly becoming the dominant medium for quick and easy access to news. It is therefore not surprising that there are significant efforts to censor certain news articles or even entire web sites. Furthermore, last year's incidents of Arab Spring as well as SOPA and PIPA acts brought web censorship in the forefront of attention.

Most existing sources of Internet censorship do not provide detailed technical information. They are mainly based on accessibility tests and a lot of current censorship measurement efforts are mostly conducted by journalists without technical background. As a result, there is a lack in measuring who censors what, how, when on an ongoing basis since what is blocked changes over time.

In this thesis we present an approach to detect censorship in a technically sound way. First, we studied current filtering technologies as well as existing approaches to map web filtering. Then, we designed and implemented a web censorship monitor, called CensMon. CensMon is distributed in nature, operates automatically and does not rely on Internet users to report censored web sites, can differentiate access network failures from possible censorship, and uses multiple input streams to determine what kind of censored data to look for. Our evaluation shows that CensMon can successfully detect censored content and spot the filtering technique used by the censor.

Supervisor: Professor Evangelos Markatos

Περίληψη

Το Internet είναι παραδοσιακά το πιο ελεύθερο μέσο για πρόσβαση και δημοσιοποίηση πληροφορίας. Επίσης, το Internet εξελίσσεται ταχύτατα ως το κυρίαρχο μέσο για γρήγορη και εύκολη πρόσβαση για ειδήσεις. Έτσι, δεν προκαλεί έκπληξη το γεγονός ότι γίνονται προσπάθειες για να λογοκριθούν συγκεκριμένα ενημερωτικά άρθρα στο Διαδίκτυο ή ακόμα και ολόκληροι ειδησιογραφικοί ιστότοποι. Επίσης, τα πρόσφατα γεγονότα της Αραβικής Άνοιξης καθώς και τον νόμων SOPA και PIPA έφεραν τα θέματα της λογοκρισίας στο Διαδίκτυου στο επίκεντρο του ενδιαφέροντος.

Οι περισσότερες πηγές για λογοκρισία στο Διαδίκτυο δεν παρέχουν τεχνικές λεπτομέριες. Επίσης, είναι περισσότερο βασισμένες σε ελέγχους προσβασιμότητας ενώ πολλές μετρήσεις γίνονται από δημοσιογράφους, οι οποίοι υπολείπονται σε τεχνικό υπόβαθρο. Σαν αποτέλεσμα, υπάρχει ανάγκη να ξέρουμε τι φιλτράρεται στο Διαδίκτυο, από ποιον, πως και πότε μέσω μιας συνεχής παρακολούθησης, μιας και το τι λογοκρίνεται αλλάζει με τον χρόνο .

Σε αυτήν την εργασία παρουσιάζουμε μια προσπάθεια να εντοπίσουμε διαδικτυακή λογοκρισία εστιάζοντας κυρίως σε τεχνικές παραμέτρους. Αρχικά μελετάμε τις σύγχρονες τεχνολογίες λογοκρισίας καθώς και τις κύριες υπάρχουσες προσπάθειες για την μελέτη της λογοκρισίας στο Διαδίκτυο. Στην συνέχεια σχεδιάζουμε και υλοποιούμε ενα σύστημα παρακολούθησης της λογοκρισίας στο Διαδίκτυου που ονομάζεται CensMon. Το CensMon είναι από την φύση του αποκεντροποιημένο, λειτουργεί αυτόματα καθώς δεν βασίζεται στους χρήστες του Διαδικτύου για να αναφέρουν λογοκριμένες σελίδες στο Internet, μπορεί να ξεχωρίσει δικτυακά σφάλματα από πιθανές απόπειρες λογοκρισίας και τέλος, χρησιμοποιεί πολλαπλές ροές εισόδου για να ψάξει για λογοκριμένο υλικό. Η αξιολόγηση του συστήματος μας έδειξε ότι το CensMon μπορεί να εντοπίσει επιτυχώς λογοκριμένο περιεχόμενο στο Διαδίκτυο καθώς και την τεχνική φιλτραρίσματος που χρησιμοποιεί ο λογοκριτής.

Επόπτης: Καθηγητής Ευάγγελος Μαρκατος

Acknowledgments

I would like to thank Professor Evangelos Markatos as well as Dr. Sotiris Ioannidis for providing me the opportunity to conduct my master's research under their guidance and whose patience and academic experience have been invaluable to me.

It has been a pleasure to be a member of Distributed Computing Systems laboratory and I owe thanks to all the people of the lab for their support, feedback and happy times spent together during my master studies.

Many thanks to my friends in Heraklion for helping me get through the difficult times and for all the emotional support, entertainment, and caring they provided. Moreover, I am indebted to all the guys in Kallithea for providing me with an environment to write this thesis while serving military duty in Athens.

Finally, I would like to thank my parents, Nikos and Elissavet, who have been a constant source of support - emotional, moral and financial - during these years, and this thesis would certainly not have existed without them.

Contents

List of Figures

List of Tables

1

Introduction

Nowadays, the Internet plays a significant role in the economic, political, and social fabrics of global society. Having more than 1.7 billion users [22], the Internet fosters an estimated $1.5 trillion in annual global economic benefits [54], and it is widely agreed to offer a lot of societal benefits as well as a great promise for improving the communication capabilities of many users. As its growth continues (interdomain traffic has an annual growth rate of 44.5% [81]), the Internet has proven susceptible to emerging patterns of overt and more subtle disruption.

Our even increasing dependence on networked communications makes it easier for organizations to control, monitor, or block user communications. ISPs and governments routinely restrict access to Internet content and services, either by censoring access to the information or by degrading the performance of various services (e.g. violating network neutrality). Indeed, although we think of the Internet as enabling the "democratization" of communications, free and open access is at risk according to OpenNet Initiative (ONI) [29].

1.1 Internet's Filtering Map

Being the major source of knowledge concerning Internet filtering, the OpenNet Initiative [29] is a joint project whose goal is to monitor and report on Internet filtering and surveillance practices by nations. The project employs a number of technical means, as well as an international network of investigators, to determine the extent and nature of government-run Internet filtering programs.

Despite the wide range of topics filtered around the world, according to ONI's studies there are essentially three motives or rationales for Internet filtering: politics and power, social norms and morals, and security concerns. Accordingly, most of the topics subject to filtering (see Table 1.1) fall under one of three thematic headings: political, social, and security. A fourth theme -Internet tools- encompasses the networking tools and applications that allow the sharing of information relating to the first three themes. Included here are translation tools, anonymizers, blogging services, and other Web-based applications categorized in Table 1.1 (found in [70]).

Political Content

Figure 1.1 (found in [32]) depicts the Internet filtering map (according to ONI) based on content that expresses views in opposition to those of the current government, or is related to human rights, freedom of expression, minority rights, and religious movements.

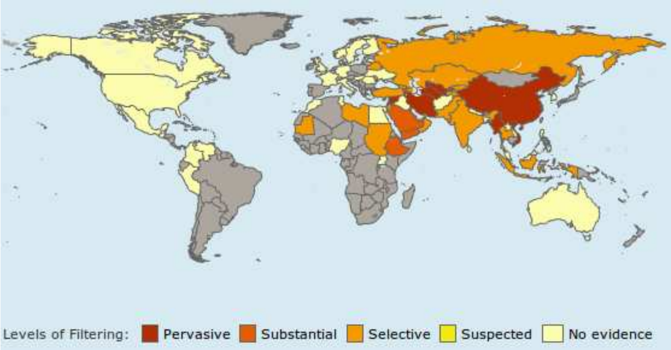


FIGURE 1.1: Internet filtering map based on political content (found in [32]).

Social Content

In Figure 1.2 (found in [33]) we can see the Internet filtering map based on content related to sexuality, gambling, and illegal drugs and alcohol, as well as other topics that may be socially sensitive or perceived as offensive.

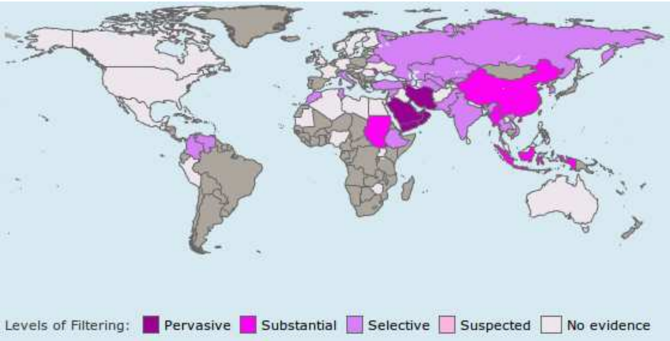


FIGURE 1.2: Internet filtering map based on social content (found in [33]).

Conflict & Security

Figure 1.3 (found in [30]) depicts the Internet filtering map based on content related to armed conflicts, border disputes, separatist movements, and militant groups.

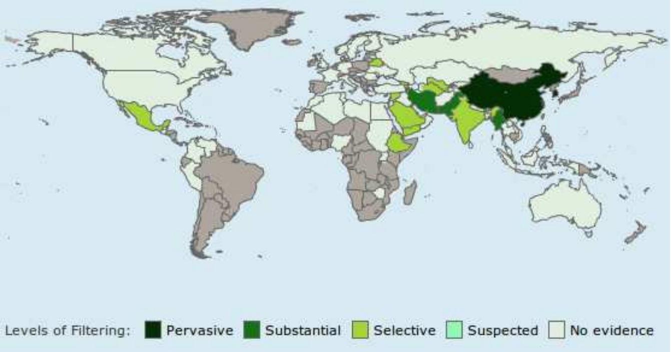


FIGURE 1.3: Internet filtering map based on conflict & security content (found in [30]).

Internet Tools

Figure 1.4 (found in [31]) depicts the Internet filtering map based on content concerning with web sites that provide e-mail, Internet hosting, search, translation, Voice over Internet Protocol (VoIP) telephone service, and circumvention methods.

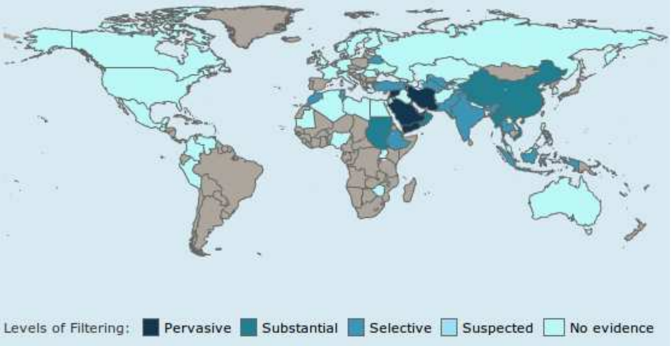


FIGURE 1.4: Internet filtering map based on Internet tools (found in [31]).

1.2 Recent Internet Censorship Incidents

In 2011, two major incidents brought Internet censorship in the forefront of attention: Arab Spring as well as SOPA (Stop Online Piracy Act) and PIPA (Protect IP Act) acts.

Arab Spring

During the Arab Spring, the Internet and mobile technologies, particularly social networks such as Facebook and Twitter, played a significant role in organizing and spreading the protests and making them visible to the rest of the world. This use of digital media lead to web censorship resulting the complete loss of Internet access for periods of time in Egypt and Libya in 2011.

As described in [64], on the evening of January 27, 2011 Egypt vanished from the Internet. The Egyptian government ordered a complete Internet shutdown [10] while popular anti-government protests were calling for the resignation of Egyptian President Hosni Mubarak. The heavy-handed attempt to block communications in the country did not quell the protests, and may have even increased the number of people in the streets; protests intensified and continued even after Internet connectivity was restored five days later. Under political pressure from inside and outside Egypt, President Hosni Mubarak resigned, turning command over to the military on February 11.

Four days later, similar protests erupted in Libya, calling for an end to the Gaddafi regime. On February 17 major protests took place across the country [23]. On the night of February 18 (Friday) the government imposed an "Internet curfew", blocking all Internet access until morning (08:01 local time), and repeating it the next day (Saturday) [36]. In the following days, Libyan traffic to popular sites increased steadily [16] until Internet access was disabled again, this time for nearly four days.

Protect IP Act and Stop Online Piracy Act

The Protect IP Act is a proposed law with the stated goal of giving the US government and copyright holders additional tools to curb access to "rogue websites dedicated to infringing or counterfeit goods", especially those registered outside the U.S. The Stop Online Piracy Act (SOPA) is a bill, introduced in the United States House of Representatives on 26 October 2011, that expands the ability of U.S. law enforcement and copyright holders to fight online trafficking in copyrighted intellectual property and counterfeit goods.

According to both these Acts, non-authoritative domain name servers would be ordered to take technically feasible and reasonable steps to prevent the domain name from resolving to the IP address of a website that had been found by the court to be "dedicated to infringing activities". The website could still be reached by its IP address, but links or users that used the website's domain name would not reach it. Moreover, search engines such as the already protesting Google would be ordered to "remove or disable access to the Internet site associated with the domain name set forth in the [court] order; or not serve a hypertext link to such Internet site". In addition to domain-name filtering, SOPA would impose an open-ended obligation on Internet Service Providers (ISPs) to prevent access to infringing sites.

Preventing access to specific sites would require ISPs to inspect all the Internet traffic of its entire user base the kind of privacy-invasive monitoring that has come under fire in the context of deep packet inspection for advertising purposes. Finally, SOPA would also hold the web publishers and hosting services responsible for curbing their users from posting copyright-infringed content.

Steve Crocker et al. [53] suggest that the DNS filtering provisions in the bills "raise serious technical and security concerns" and would "break the Internet". They claim that "From an operational standpoint, a resolution failure from a nameserver subject to a court order and from a hacked nameserver would be indistinguishable. Users running secure applications have a need to distinguish between policy-based failures and failures caused, for example, by the presence of an attack or a hostile network, or else downgrade attacks would likely be prolific". Furthermore, a group of engineers, networking specialists, security experts and other specialists deeply involved with the Internet's development and growth have sent a letter [2] to lawmakers criticizing the highly controversial SOPA and PIPA bills and imploring them not to pass the legislation, which they say would stifle innovation and "threaten engineers who build Internet systems or offer services that are not readily and automatically compliant with censorship actions by the U.S. government."

A browser plugin called MAFIAAFire Redirector [24] already exists that redirects visitors to an alternative domain when a site's primary domain has been seized. The Mozilla Foundation says that United States Department of Homeland Security (DHS) requested by phone that Mozilla remove the plugin [21], a request with which they have not yet complied [25]. Instead, Mozilla's legal counsel has asked for further information from the DHS, including legal justification for the request.

Finally, on January 18 2012, widespread online protests against SOPA and PIPA were held [46]. Internet users defended their right for a free and open Internet since they believe that these bills would stifle expression on the world wide web.

1.3 The Need for Web Censorship Monitoring

As we have seen from the previous section, censorship on the world wide web appears to be taking place more than ever before. The OpenNet Initiative reports that there are almost 60 countries that either filter or are suspected of filtering web content showing that it is not only typically-considered "oppressive regimes" that censor the web access of their Internet users. Furthermore, in 2011, the Freedom House released a report that examined Internet-freedom in 37 countries around the world, and found that there is a diverse and growing threat to Internet-freedom [11]. Also, from 2000 onwards, many web censorship-related stories can be found by performing a simple web search in Slashdot's Your Rights Online section (YRO) [44]. This growing trend towards blocking, tampering with, or otherwise restricting communications on the Internet calls for better techniques for monitoring the state of restrictions on Internet content and communications (i.e., improving “transparency”).

Web users can learn about filtered content in countries from various articles on popular news websites. However, the information provided is usually very sparse and in most cases, limited to a few filtered URLs per country. Thus, there is a need for a service that provides more detailed information about up-to-date web censorship, such as ONI’s Herdict Web [35].

Censorship is a phenomenon that changes over time. The main criteria used for censorship-checking by all these web services is whether a specific web site is accessible or not. ONI’s Herdict Web [35], which is the major web censorship monitoring service, depends on user-generated feeds to determine the accessibility or inaccessibility of a web site. This fact alone may result in false positives, since often users cannot differentiate network failures from actual censorship. Therefore, there is need for a service that does *not* depend on user input, and runs transparently to track all the changes in the accessibility state of web pages.

In this thesis, we present the design and implementation of CensMon, a system that offers users real-time information about filtered web content, without *actually* depending on web users’ experience. Specifically our systems has three design characteristics: (i) it uses PlanetLab’s [40] nodes, to create a worldwide web censorship monitor, (ii) it uses plug-in feed modules, that stream newly published, possibly sensitive, content to our system for censorship-checking, and (iii) we maintain historical data by continuously monitoring sites that have been censored. Our results, by using many different web sources, show that our system can exploit these sources to detect censored content and identify the filtering technique used.

1.4 Contributions

The contributions of this thesis are the following :

- We present the design and implementation of CensMon, a system that offers users real-time information about filtered web content, without *actually* depending on web users’ experience.

- CensMon uses information streams automatically extracted from a plethora of sources for censorship checking.

- CensMon monitors sites that have been found censored in order to check the censorship state.

- Our system can detect censored content and identify the filtering technique used.

- Finally, we present some early interesting observations about the behaviour of “Great Firewall of China”.

1.5 Thesis Outline

The rest of this thesis is organized as follows. Chapter 2 provides technical background concerning censorship on the world wide web and current filtering technologies. In Chapter 3, we describe the architecture and the overall design of CensMon. Chapter 4 presents the experimental evaluation of CensMon. Chapter 5 presents previous work related to Internet censorship, net neutrality as well as censorship circumvention. In Chapter 6, we propose future directions of our work while Chapter 7 is the discussion section of the thesis. Finally, in Chapter 8 we draw our conclusions.

TABLE 1.1: Categories subject to Internet Filtering (found in [70])

Free expression and media freedom	Political transformation and opposition parties
Foreign relations and military	Political reform, legal reform, and governance
Militants, extremists, and separatists	Human rights
Minority rights and ethnic content	Women’s rights
Environmental issues	Hate speech
Economic development	Sensitive or controversial history, arts, and literature
Sex education and family planning	Public health
Pornography	Provocative attire
Gay/lesbian content	Dating
Gambling	Gaming
Minority faiths	Religious conversion, commentary, and criticism
Alcohol and drugs	Anonymizers and circumvention
Hacking	Blogging domains and blogging services
Voice over Internet Protocol (VOIP)	Search engines
Free e-mail	Web hosting sites and portals
Multimedia sharing	P2P
Groups and social networking	Commercial sites

2
Censorship Technologies

According to Wright et al. [102] “The Internet has expanded through the accretion of protocols, services and applications that have been extended and improved far beyond their original purpose since its development was neither carefully planned nor accurately predicted. As a result, many of the protocols provide opportunities both for filtering technologies and for attempts to bypass or study those technologies.”

There are a lot of methods applied to filter Internet connections at a national level. The major filtering techniques have been categorised by Murdoch and Anderson [88] as follows in the next sections.

2.1 TCP/IP Header Filtering

The Network Layer

The network layer (layer 3) of the OSI model is primarily responsible for logical addressing and routing of data. IP, the Internet Protocol, is the fundamental protocol by which traffic passes across the Internet, encoded in IP packets. An IP packet consists of a header followed by the data (payload) the packet carries. Normally, routers must inspect the packet header which details the numerical address of the packet’s destination. As a result, filtering may occur via inspection of the header of an IP packet which may therefore be filtered according to lists of banned destination IP addresses.

The following sample rules are typical Access Control Lists for Cisco devices that will deny all TCP and UDP traffic to or from the IP address (139.91.151.170) associated with the website of the ICS-FORTH (www.ics.forth.gr):

deny ip host 139.91.151.170 any
deny ip any host 139.91.151.170

If these rules are added to a central networking device, there would be no way of accessing the website of the ICS-FORTH unless the filtering is somehow circumvented. This holds also true for any other service listening on that particular host. The advantage of layer 3 filtering is that, in theory, processing such rules requires only minimal resources on any networking device and can be done very efficiently.

However, in practice, given the vitality of IP addresses and websites these rulesets often tend to become very large in size and cause a huge performance loss. Additionally managing, distributing and synchronizing them among all network devices involved is another difficult challenge for the operator of the network infrastructure. Lastly, due to the lack of granularity in the filtering mechanism itself, layer 3 filtering does not provide a way of limiting the blocking to a specific service or port. Consequently, taking also into account the potential for services to change or to have multiple IP addresses, the filtering might be too broad and may unintentionally block access to a particular host or service.

The Transport Layer

Layer 4 (transport layer) is “primarily responsible for the formatting and handling of the transport of data in a transparent manner”. It provides “reliable and accurate delivery of the data to the next layer” [89] and uses protocols such as TCP, UDP as well as ICMP. The UDP and TCP protocols both include information (i.e. a port number) about the type of service (e.g. port 80 for HTTP) a packet was most likely generated by or is destined for. Together with the source and destination addresses of a packet, this application-specific information provides a finer distinction and division of network traffic when compared with OSI layer 3.

Since each host may provide multiple services such as hosting both web sites and e-mail servers, blocking based only on IP addresses will make all services on each blacklisted host inaccessible. By additionally blacklisting the port number, which is also in the TCP/IP header, slightly more precise blocking can be achieved. Common applications on the Internet have characteristic port numbers, allowing routers to make a crude guess as to the service being accessed. As a result, in order to block just the web traffic to a site, a censor might block only packets destined for port 80 (the default port for web servers).

An example of layer 4 filtering would be:

deny tcp any host 139.91.151.170 eq 80

In the example above, web traffic from any host with any source port to destination port 80 (HTTP) on 139.91.151.170 is denied. If such a rule is deployed, any host affected by this filtering would be unable to communicate with host 139.91.151.170 on port 80. Although layer 4 filtering offers greater flexibility and precision in terms of the scope of the filtering, it may also block access to resources it should not block (overblocking). For expamle, Dornseif [68] mentions the HTTP protocol in which one server with a single IP address may host several (up to hundreds or thousands) other websites (so-called "name virtual hosting"). Hence, if access to the web server is blocked, then access to all other websites that are hosted on the same server is also blocked.

2.2 TCP/IP Content Filtering

TCP/IP header filtering can only block communication based on where packets are going to or coming from, not on what they contain. This can be a problem if it is impossible to establish the full list of IP addresses containing prohibited content, or if some IP address contains enough non-infringing content to make it unjustifiable to totally deny all communication with it.

Rather than inspecting the header, a filter may search the content of traffic for banned terms using "*deep packet inspection*". This approach is far more flexible, allowing packets to be blocked only if the include banned keywords or the traffic patterns of particular applications. Deep packet inspection refers to "the capabilities of a firewall or an Intrusion Detection system (IDS) to look within the application payload of a packet or traffic stream and make decisions on the significance of that data based on the content of that data" [69]. Initially used as a technology to detect and defend against known and unknown network-based attacks, deep packet inspection is also a suitable technique for performing content filtering if an appropriate set of signatures and keywords is employed. However, since routers do not normally examine packet content but just packet headers, extra equipment may be needed. Typical hardware may be unable to react fast enough to block the infringing packets, so other means to block the information must be used instead.

Common techniques for TCP/IP content filtering are the following:

- *Uniform Resource Locator (URL) filtering* : URL strings are scanned for target keywords regardless of the domain name specified in the URL. Typical circumvention methods are to use escaped characters in the URL, or to use encrypted protocols such as VPN and TLS/SSL.

- *HTML response keyword filtering* : In HTML response filtering, when a keyword is detected within an HTTP response transfer, the censor attempts to interrupt the connection and stop the transfer. Typical circumvention methods are to use encrypted connections - such as VPN and TLS/SSL - to escape the HTML content, or by reducing the TCP/IP stack's MTU/MSS to reduce the amount of text contained in a given packet.

- *Connection reset* : If a previous TCP connection is blocked by the filter, future connection attempts from both sides can also be blocked for some variable amount of time. Depending on the location of the block, other users or websites may also be blocked, if the communication is routed through the blocking location. A circumvention method is to ignore the reset packet sent by the firewall.

As packets have a maximum size, the full content of the communication will likely be split over multiple packets. Thus, while the offending packet will get through, the communication can be disrupted by blocking subsequent packets. This may be achieved by blocking the packets directly or by sending a message to both of the communicating parties requesting they terminate the conversation. Another effect of the maximum packet size is that keywords may be split over packet boundaries. Devices that inspect each packet individually may then fail to identify infringing keywords. For packet inspection to be fully effective, the stream must be reassembled, which adds additional complexity.

Deep packet inspection approach can be partially defeated by using encrypted connections. Nevertheless, filters may choose simply to block all encrypted connections in response, or to block traffic according to identifying traffic signatures that can occur even in encrypted protocols. The most significant limitation of this approach is that inspection of traffic content comes at a significant computational cost.

2.3 DNS Tampering

The Domain Name System (DNS) is a globally deployed hierarchical database to resolve hostnames (e.g. www.ics.forth.gr) into the corresponding IP addresses (e.g. 139.91.151.170). Since most Internet communication uses domain names rather than IP addresses, DNS is thus critical for most user-focused services such as the Web.

Thus, if the domain name resolution stage can be filtered, access to infringing sites can be effectively blocked. With this strategy, the DNS server accessed by users is given a list of banned domain names. When a computer requests the corresponding IP address for one of these domain names, an erroneous (or no) answer is given. Without the IP address, the requesting computer cannot continue and will display an error message.

Although it was never intended to be used as a filtering mechanism, it nowadays “seems to be the preferred way of blocking” [68] due to the simplicity and yet effectiveness in which manipulations can be done. Dornseif was the first to study this order in 2003 and identified the major techniques for performing DNS tampering [68]:

- *Refused*: The easiest way to stop users from connecting to a certain host is to simply refuse to resolve that given domain. Therefore the DNS standard defines the reply “REFUSED” which means that “the name server refuses to perform the specified operation for policy reasons” [87]. Consequently this is likely to cause a “host not found” or “connection refused” error message.

- *Nxdomain*: A manipulation in which the existence of a particular domain is denied (“NXDOMAIN, non-existing domain”) by the recursive DNS server of the provider. To invalidate a domain, the provider has to pretend to be authoritative for that domain and hence breach the DNS standard. For the user this forgery will also cause a “host not found” error message and will prevent the user from connecting to the target host.

- *Name redirection*: Refers to a deliberate modification in which the user’s request to resolve a certain domain is answered with bogus data. This will typically result in the user being unintentionally redirected (“hijacked”) to another site.

- *Name invalidation*: A technique similar to “name hijacking” in which resolving a domain results in invalid rather than bogus replies. This will cause a “could not connect” error message. Dornseif refers to this method as “name astrayment”.

- *Silence*: Another way of refusing to resolve a particular domain is silently not to respond to such a request at all. This will result in a delay or even a timeout and will eventually cause a “host not found” error.

- *Provoked server failures*: This type of tampering will cause a server-generated error message to be send to any client trying to resolve a certain domain. Hence the user will experience some sort error message (e.g. “could not connect”) and will be unable to resolve or connect to the destination domain.

A typical circumvention method is to find a Alternative DNS root that resolves domain names correctly, but domain name servers are subject to blockage as well, especially IP address blocking. Another workaround is to bypass DNS if the IP address is obtainable from other sources and is not itself blocked. Examples are modifying the Hosts file or typing the IP address instead of the domain name as part of a URL given to a Web browser.

2.4 HTTP Proxy Filtering

A more sophisticated approach is to pass all Internet traffic through an intermediary “proxy” service that fetches and, typically, caches information for users. This is a common Internet service that can be used to speed up Internet connections and reduce traffic. However, as well as improving performance, an HTTP proxy can also block Web sites. A suitably enabled proxy can employ sophisticated filtering on certain destinations, whilst leaving other connections alone. This approach can, by ignoring the majority of traffic, be efficient on a national scale while still allowing for detailed filtering similar to TCP/IP content filtering.

Taking into account ISPs, a transparent HTTP proxy may intercept outgoing web requests and send them to a proxy server. While being quite complex to set up, this option avoids any configuration changes on the user's computer (contrary to non-transparent proxies). This gives it the opportunity of seeing both the Web site domain name and which page is requested, allowing more precise blocking when compared to TCP/IP header or DNS filtering.

2.5 Other Approaches

We could consider social pressure and legislation as means of filtering however in this section we will concentrate mostly on other technology based approaches.

Denial of Service

Where the organization deploying the filtering does not have the authority (or access to the network infrastructure) to add conventional blocking mechanisms, Web sites can be made inaccessible by overloading the server or network connection. This technique, known as a Denial-of-Service (DoS) attack, could be mounted by one computer with a very fast network connection; more commonly, a large number of computers are taken over and used to mount a distributed Dos (DDoS).

Domain Deregistration

The domain name system is organized hierarchically, with country domains such as ".uk" and ".de" at the top, along with the nongeographic top-level domains such as ".org" and ".com." The servers responsible for these domains delegate responsibility for subdomains, such as example.com, to other DNS servers, directing requests for these domains there. Thus, if the DNS server for a top-level domain deregisters a domain name, recursive resolvers will be unable to discover the IP address and so make the site inaccessible. Country-specific top-level domains are usually operated by the government of the country in question, or by an organization appointed by it. Thus, if a site is registered under the domain of a country that prohibits the hosted content, it runs the risk of being deregistered.

Server Takedown

Servers hosting content must be physically located somewhere, as must the administrators who operate them. If these locations are under the legal or extra-legal control of someone who objects to the content hosted, the server can be disconnected or the operators can be required to disable it.

Surveillance

The above mechanisms inhibit the access to banned material, but are both crude and possible to circumvent. Another approach, which may be applied in parallel to filtering, is to monitor which Web sites are being visited. If prohibited content is accessed (or attempted to be accessed) then legal (or extra-legal) measures could be deployed as punishment. If this fact is widely publicized, it will discourage others from attempting to access banned content, even if the technical measures for preventing it are inadequate. This type of publicity has been seen in China with Jingjing and Chacha, two cartoon police officers who inform Internet users that they are being monitored and encourage them to report suspected rulebreakers.

Search result removal

Search engines, may exclude web sites that they would ordinarily include. This renders a site invisible to people who do not know where to find it. When a major portal does this, it has a similar effect as censorship. Sometimes this exclusion is done to satisfy a legal or other requirement, other times it is purely at the discretion of the portal. For example Google.de and Google.fr remove Neo-Nazi and other listings in compliance with German and French law [14].

Internet kill switch

A technically simpler method of Internet censorship is to completely cut off all routers, either by software or by hardware (turning off machines, pulling out cables). This appears to have been the case on 27/28 January 2011 during the 2011 Egyptian protests, in what has been widely described as an “unprecedented” internet block [55]. About 3500 Border Gateway Protocol (BGP) routes to Egyptian networks were shut down from about 22:10 to 22:35 UTC 27 January [9]. This full block was implemented without cutting off major intercontinental fibre-optic links, with Renesys stating on 27 January, ”Critical European-Asian fiber-optic routes through Egypt appear to be unaffected for now.” [9] Full blocks also occurred in Myanmar/Burma in 2007 [43] and Libya in 2011 [55].

3
CensMon Architecture

CensMon is a system that conducts extensive accessibility tests, trying not only to detect the presence of filtering but also to spot the root cause of it, if possible. It consists of two basic building blocks: the central server and the network of sensing nodes. We refer to the sensing nodes as CensMon's agents. In this chapter, we describe the properties of CensMon's central server and demonstrate the design of our system by providing a test case.

3.1 General Overview

The architecture of CensMon is illustrated in Figure 3.1. Figure 3.1 shows a user accessing CensMon's web front end, which is one of the possible systems inputs (in the upper right part of there figure are the inputs used during our evaluation period). The user can choose if they want to forward the URL in question to a specific alive agent or to all of CensMon's agents. Then the central server will forward the query and will try to detect possible filtering of the user request. Afterwards, server stores the results reported by the agents in the database. Finally, CensMon informs the user via web front-end for the results of their query.

3.2 Central Server

The central server is the headquarters of the entire system. The web front-end runs there as well as all the scripts handling CensMon's input. Furthermore, the analysis of the information collected by the agents takes place here. Finally, it runs the database that stores all probing information and the filtering history. Being the core of our system, central server's functionalities are significant for our system.

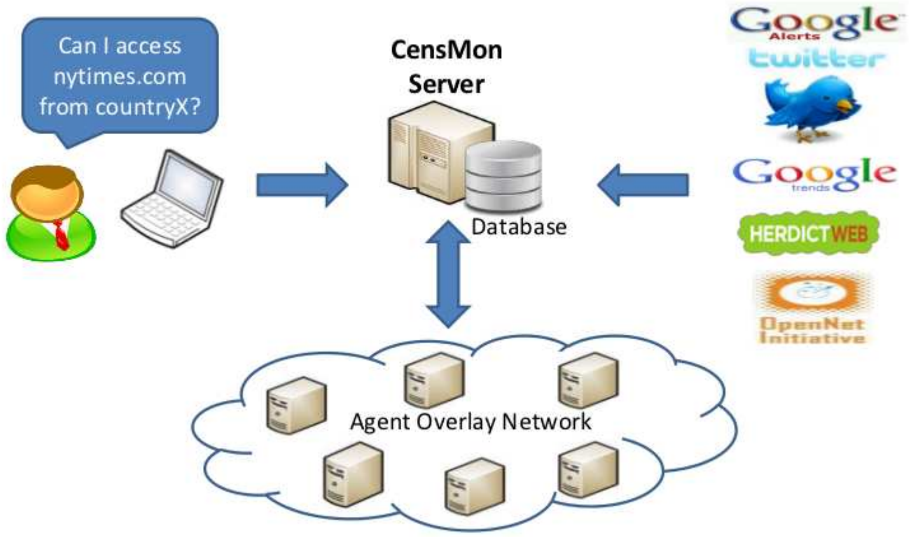


FIGURE 3.1: CensMon's architecture.

3.3 Agent Network

Our system's agent network comprises of nodes which collect accessibility information in various levels. The network of agents is responsible for CensMon's distributed nature. This way, information is collected from the agents and thereby sent to the central server for further analysis and storage.

3.4 Filtering Detection Procedure

We will now describe the methodology followed by our system. There are eight steps that take place in our system as illustrated in Figure 3.2:

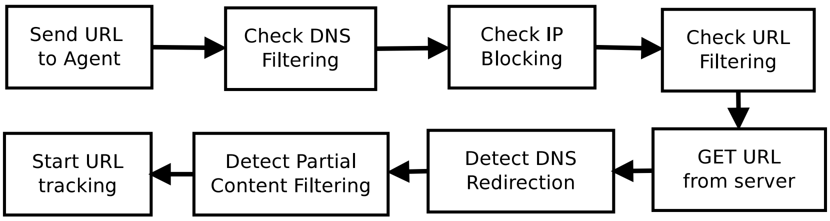


FIGURE 3.2: Filtering Detection Procedure.

1. At the start, the central server receives as input a URL to test. It then forwards this URL to all alive agents in CensMon's network to be tested. To

avoid a flood of messages to all agents, we have inserted a small time-out between messages that are sent to the agents.

2. Once an agent receives a URL from the server, its initial task is to make a DNS request for the domain of the URL so as to get the corresponding IP address or addresses. If no IP address is returned from the DNS server then the agent reports to the central server the probable cause of the DNS failure (e.g. connection refused, connection timed out, non-existing domain etc) to take further action.

3. If the agent successfully resolves the IP address of the domain in question, it tries to connect to that IP address at port 80, in an attempt to detect whether IP address blacklisting takes place. Upon successful connecting to the remote port, the agent continues to the next step of our protocol, otherwise it reports the connectivity problem to the central server.

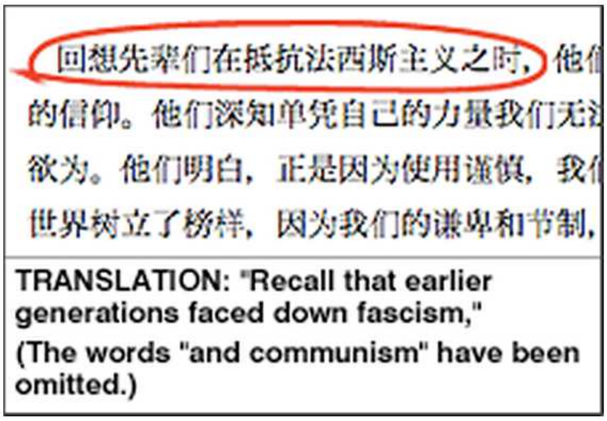
4. Having determined if there is no IP address blocking, the agent tries to find out if there is any kind of filtering at the application level (HTTP). The agent tries to detect URL keyword filtering. Inspired by Park's [90] approach, we have set up a seperate web server serving null (empty) content so as to avoid HTML response keyword filtering. The agent contacts our webserver requesting our webserver's default URL concatenated with the URL requested from CensMon server. This is, if the requested URL is initially `http://www.cnn.com`, then our agent's HTTP request becomes `http://www.ourdomain.com/www.cnn.com`. This way the agent detects if any kind of URL keyword filtering is present or the expected *404 Not Found* Status Code is returned. Again, if the agent detects URL keyword filtering it reports it to the central server.

5. Finally, the agent attempts to access the initial URL (using HTTP 1.1 as described in [90]) and sends the received status code, HTML code and information gathered from the previous steps to the central server. In case of redirection the agent additionally reports the final URL and the final IP address visited.

6. Since all the agents have reported their findings for a specific probe, the central server starts its post-mortem analysis of the agents' reports. CensMon can detect filtering that uses DNS Name hijacking (probable redirection to a block-page) by correlating the resolved IP addresses (matching IP prefixes) returned by DNS servers to all our agents concerning a specific URL. Nevertheless, if a domain name has more than one associated IP addresses, then our data are not enough to determine DNS manilupation with precision.

7. Next, CensMon tries to identify censorship of partial content in a web page.

   Back to 2009, China has censored parts of the new US president's inauguration speech that have appeared on a number of websites [28]. On the website of state-run Xinhua news agency the Chinese-language version, the word “communism” was taken out from the phrase “Recall that earlier generations faced down fascism and communism ... ” as seen in Figure 3.3.



回想先辈们在抵抗法西斯主义之时，他

的信仰。他们深知单凭自己的力量我们无

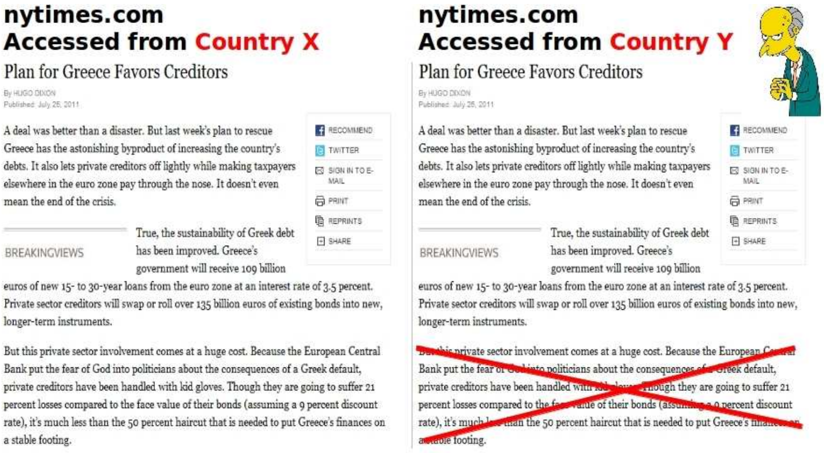
欲为。他们明白，正是因为使用谨慎，我

世界树立了榜样，因为我们的谦卑和节制，

TRANSLATION: "Recall that earlier generations faced down fascism,"
(The words "and communism" have been omitted.)

FIGURE 3.3: Obama speech censored in China (found in [28]).

Our motivation for this partial filtering detection was that we believe that in a near future scenario a smart and powerful (in terms of infrastructure) censor could filter just a part of an article and not the whole website or the URL. This could be accomplished by using on-the-fly filtering by the routers which respectively have very high hardware requirements. An example of partial content filtering can be seen in Figure 3.4.



nytimes.com
Accessed from Country X

Plan for Greece Favors Creditors

By HUGO DIXON
Published: July 25, 2011

A deal was better than a disaster. But last week's plan to rescue Greece has the astonishing byproduct of increasing the country's debts. It also lets private creditors off lightly while making taxpayers elsewhere in the euro zone pay through the nose. It doesn't even mean the end of the crisis.

BREAKINGVIEWS

True, the sustainability of Greek debt has been improved. Greece's government will receive 109 billion euros of new 15- to 30-year loans from the euro zone at an interest rate of 3.5 percent. Private sector creditors will swap or roll over 135 billion euros of existing bonds into new, longer-term instruments.

But this private sector involvement comes at a huge cost. Because the European Central Bank put the fear of God into politicians about the consequences of a Greek default, private creditors have been handled with kid gloves. Though they are going to suffer 21 percent losses compared to the face value of their bonds (assuming a 9 percent discount rate), it's much less than the 50 percent haircut that is needed to put Greece's finances on a stable footing.

nytimes.com
Accessed from Country Y

Plan for Greece Favors Creditors

By HUGO DIXON
Published: July 25, 2011

A deal was better than a disaster. But last week's plan to rescue Greece has the astonishing byproduct of increasing the country's debts. It also lets private creditors off lightly while making taxpayers elsewhere in the euro zone pay through the nose. It doesn't even mean the end of the crisis.

BREAKINGVIEWS

True, the sustainability of Greek debt has been improved. Greece's government will receive 109 billion euros of new 15- to 30-year loans from the euro zone at an interest rate of 3.5 percent. Private sector creditors will swap or roll over 135 billion euros of existing bonds into new, longer-term instruments.

~~But this private sector involvement comes at a huge cost. Because the European Central Bank put the fear of God into politicians about the consequences of a Greek default, private creditors have been handled with kid gloves. Though they are going to suffer 21 percent losses compared to the face value of their bonds (assuming a 9 percent discount rate), it's much less than the 50 percent haircut that is needed to put Greece's finances on a stable footing.~~

FIGURE 3.4: Partial content filtering in a near future scenario (article found in [39]).

Since the HTML code of a web page that was successfully accessed by one of our agents is stored at the server, CensMon analyses the HTML code of the web pages returned by all agents and are associated with the same URL.

CensMon uses Arc90's Readability functionality [5], when possible, in order to extract the content that is most likely to be the stuff a user wants to read (and what the censor wants to filter). Python port of arc90's readability traverses the DOM and uses a scoring function that rewards an element for containing text, punctuation, and class or id attributes typically associated with the main content of a site. This way CensMon deals with automatic changing/updated contents (such as news sites, e.g. nytimes.com) or contents that are localized (depending on where the user is coming from). CensMon uses fuzzy hashing [78] for comparing the URLs' *readable* HTML code and detecting partial filtering.

8. Lastly, when an inaccessibility event is reported, CensMon marks it as suspicious for filtering and begins to track the specific URL with the agent that reported the inaccessibility. This tracking is mandatory for CensMon to be able to differentiate between filtering, in cases where inaccessibility is repeatedly reported, network errors, if the URL finally becomes accessible, or change in censor's policy.

4
Experimental Evaluation

In this chapter we present the results of a preliminary evaluation of CensMon . First, we describe the experimental setup and implementation and then we present all the experimental results.

4.1 Implementation

4.1.1 Central Server

CensMon central server is somewhat the headquarters of the whole system. The web server as well as the database server are placed here. The former is an is Apache mod python web server while the latter is using MySQL with the appropriate relation schema that facilitates our efforts for effective storage and retrieval of the queries' data. Moreover, CGI scripts written in Python are responsible for forwarding queries to the agent network as well as for looking up in the database.

4.1.2 Agent Network

For our testbed we used nodes from PlanetLab. We run CensMon agents on PlanetLab nodes forming the CensMon network and we use the paramiko python module [37] that implements the SSH2 protocol so as to connect to these agents. Moreover, we have deployed 174 agents in 33 different countries (141 distinct ASes, 130 distinct cities) in PlanetLab. Table 4.1 shows the countries, the number of deployed agents in each one of them as well as the number of the corresponding ASes.

Country Code	#Agent Nodes	#ASes	Country Code	#Agent Nodes	#ASes	Country Code	#Agent Nodes	#ASes
AR	1	1	GR	2	1	NZ	2	2
BE	1	1	HK	2	2	PL	5	3
BR	5	3	HU	1	1	PT	2	1
CA	7	7	IE	1	1	RU	2	2
CH	5	1	IL	3	1	SE	2	2
CN	1	1	IT	2	1	SG	2	2
DE	15	4	JO	1	1	SI	1	1
ES	4	3	JP	10	8	TR	1	1
FI	1	1	KR	3	3	TW	5	4
FR	7	4	NL	3	2	US	72	72
GB	3	2	NO	1	1	UY	1	1

TABLE 4.1: Number of CensMon agents and number of ASes per country.

4.1.3 CensMon Input

In order to evaluate our system we used input from different sources. We now discuss how these sources provide CensMon with URLs to test.

User Input

CensMon has a front-end which enables users to insert URLs in the system. Users should specify the URL as well as the agent they want the system to forward the request to. After a successful request, users get the respective response and the HTML code of the requested URL. Another option is to insert a URL to be forwarded to all the CensMon agent network.

Google Alerts

Apart from user input, we use Google Alerts [17], a service provide by Google, for automatically inserting URLs of interest in our system. Google Alerts are email updates of the latest relevant Google results based on a topic of choice. The characteristics of an alert is the topic that we are interested, the frequency of receiving alerts (we choose to receive web alerts as they happen) and finally the type of the alert. Google Alerts offer five types of alerts for a specific topic: News alerts (related URLs from news sites), Blog alerts (URLs from Blogs), Real-time alerts (latest related Tweets), Discussion alerts (related threads from various fora) and Video alerts (newly published related videos).

Using Google Alerts as an input source we can check web content that may be censored and test how CensMon responds to this newly published content. We registered a Gmail account and added 4 topics to our alert services. These topics was *internet censorship*, *net neutrality*, *freedom of speech* and *human rights*. For each of these topics our email account receives alerts for all four type of alerts presented above. Using an IMAP client we fetch and insert all alerts to CensMon .

Internet Trends

In parallel, we want to test URLs that are associated with current trends discussed over the Internet. For this reason, we use the popular social network Twitter [48] and Google Hot Trends [15] for extracting periodically popular trends. Google Hot Trends [15] is a service provided by Google, where one can see a snapshot of people's interests. Nevertheless, since Twitter trends and Google Hot Trends do not include necessarily URLs, we extract characteristic keywords associated with these trends and then, by using the Google Search Engine, we feed CensMon with the top-10 URLs returned by Google for a given trend.

Herdict's Web Reported URLs

Since the OpenNet Initiative [29] is the best source of information for Internet censorship, we use as input the latest reported URLs by web users from the Herdict Web [35] site to test them with our infrastructure. We were periodically visiting Herdict Web site and automatically extracting the URLs that were reported by web users.

ONI's Categories for Internet Censorship

ONI [29] has released a list of categories in the global URL list for Internet censorship research. We chose ten of them in order to find related URLs and insert them to CensMon . The ten categories that we selected are: news outlets, freedom of speech, entertainment, government, terrorism, porn, gambling, religion, net neutrality and human rights. We then proceed and search through Google Search Engine to find the top-100 results for each category. The resulting 1000 URLs of the above categories were inserted to CensMon so as to be tested through the agent network.

4.2 Experimental Results

4.2.1 Filtering Detected

All evaluation measurements were conducted during a 14-day period in April 2011. At this period CensMon tested 4950 unique URLs from 2500 domains. Moreover, CensMon detected 951 unique URLs from 193 domains as filtered. During this period CensMon was able to detect censored content in 8 countries at different protocol levels. Table 4.2 depicts the number of unique domains where URLs have been found as censored by CensMon during the evaluation period.

AR	0	GR	0	NZ	0
BE	0	HK	2	PL	0
BR	0	HU	1	PT	0
CA	0	IE	1	RU	0
CH	0	IL	0	SE	0
CN	176	IT	0	SG	0
DE	1	JO	5	SI	0
ES	0	JP	1	TR	6
FI	0	KR	0	TW	0
FR	0	NL	0	US	0
GB	0	NO	0	UY	0

TABLE 4.2: Domains where censored URLs have been found per country by CensMon.

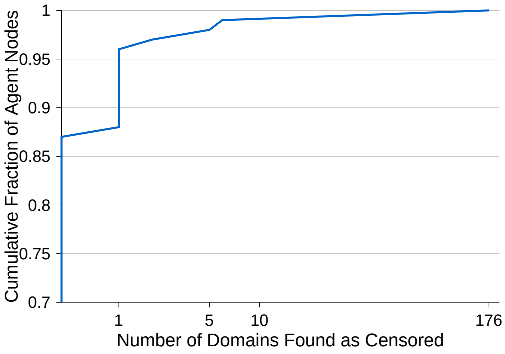


FIGURE 4.1: Cumulative distribution of the unique censored domains found during evaluation period.

Figure 4.1 shows the distribution of the unique domains that CensMon has detected as censored for all the agent nodes. As we can see, about 86% of the agent nodes have not reported any filtering event to be categorized as censored by CensMon . Moreover, about 10% of the agent nodes have found 1 to 6 domains as censored. The Chinese agent node was by far the one that reported filtering in 176 domains marked as censored.

CensMon can detect whether an inaccessibility was reported due to temporary failure or filtering after a number of tracking attempts. Whenever our system gets information by one of the agents that a specific URL is inaccessible, it tracks it in order to spot the differentiation between filtering and a possible network error events. As Figure 4.2 depicts, 21% of all the URLs that CensMon started to track were accessible after the first tracking attempt and during *all* the rest of the evaluation period, concluding that the initial inaccessibility had been caused due to network failure. Moreover, the decrease of the percentage after the first tracking attempt can be explained due to the fact that the very first attempt is done by all

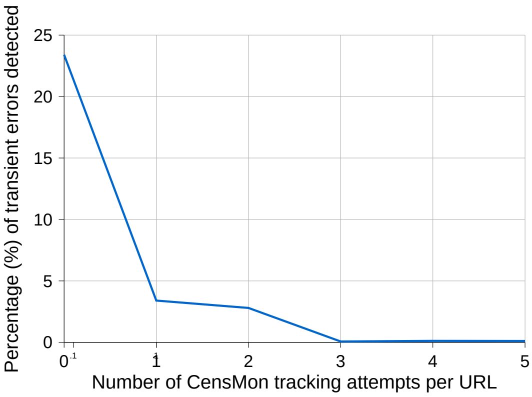


FIGURE 4.2: Percentage of temporary network failures detected after a varying number of CensMon tracking attempts.

agent nodes trying to reach the same web server, while tracking attempts are more lightweight. Therefore, after 3 tracking times we can be confident that what CensMon tracks is rather a filtering event than an early and temporary network failure. However, we have tested manually these 193 domains that were reported as censored in order to check for false positives and we have found that 3 of them were falsely marked.

Finally, 123 distinct URLs that were inserted as input to CensMon through ONI's Herdict Web site. CensMon started to track 245 cases that an inaccessibility was reported coming from 75 agent nodes in 20 countries. At last, 192 of the 245 cases of reported inaccessibility were at least one time accessible from our system while the rest 53 cases were reported as inaccessible during all our evaluation period.

4.2.2 Partial Content Censorship Analysis

During our partial content filtering analysis we analyzed 259 unique URLs from 192 unique domains. The aforementioned domains were imported from Google Alerts since the Arc90's Readability functionality [5] is mainly suitable for news articles. Using fuzzing hashing techniques [78] for comparing the URLs' *readable* HTML code we could be able not only to detect any difference in content but also the percentage of the similarity/difference of the content. Finally, in this analysis CensMon did not detect any partial filtering of a web page. More specifically we found that :

- In 95% of the URLs tested the similarity was 100%, indicating that the *readable* HTML code was identical.

- In 3% of the URLs tested the similarity was 97-99%. This happened because of very small differences in the *readable* HTML code. Examples of this case are the number of views of an article as well as “Last updated at 08:43 ET” contrary to “Last updated at 12:43 GMT”. These minor differences can be avoided in the future by adjusting the readability module to ignore them.

- In 2% of the URLs tested Arc90’s Readability was inappropriate. For example, URLs such as `http://www.bbc.co.uk/world` as well as `http://www.telegraph.co.uk/news/` could not be analysed by the module.

4.2.3 China Case Study

The People’s Republic of China operates an Internet filtering system which is widely considered to be one of the most sophisticated in the world [34]. Observing the Great Firewall of China’ss technical behaviour was quite challenging so during our 14-day evaluation and through our PlanetLab node located in China we noticed some interesting characteristics of the GFC. Our observations, while preliminary are listed below:

- We have observed what Joss Wright states in his blog post [42] “... of the redirected (faked) results received from DNS servers, we observed that these are often redirected to a small pool of sink IP addresses”. The IP addresses that we have found belonging to this pool are the ones shown in Table 4.3. This mean that whenever the GFC conducts a DNS redirection filtering attempt the faked returned IP address is one of the IP addresses shown in Table 4.3. Finally, by conducting our measurements from a single node inside China, we are not sure whether these sinks are consistent across other regions of China.

“sink” IP address
59.24.3.173
46.82.174.68
8.7.198.45
159.106.121.75
37.61.54.158
78.16.49.15
203.98.7.65
243.185.187.39

TABLE 4.3: Pool of “sink” IP addresses found in China.

- We have also noticed that there are multiple layers (levels) of filtering attempts concerning specific (major) webistes. For example, for Facebook’s webiste we have find out that apart from blocking some legitimate Facebook's IP addresses from 69.63.0.0/16 subnet, the keyword *www.facebook.com* is also filtered through URL filtering. What is more, although we noticed some legitimate DNS replies, we have observed DNS redirection attempts (resolved IP addresses came from the "sink" ones) concerning Facebook's website.

5
Related Work

5.1 Internet Censorship Monitoring Applications

The Open Net Initiative [29] is an organization that investigates and analyses Internet filtering. They provide country profiles that describe the social background and the reasons why censorship is employed, and release reports on different countries that censor the Internet. Moreover, they provide statistics about internet usage at each country, as well as a service called HerdictWeb [35] that informs web site visitors about what is censored in each one of these countries. Herdict Web allows one to see what is inaccessible, where it is inaccessible, and for how long. It uses crowd-sourcing to get information about censorship and present a real-time view of the experiences of users around the globe. Unfortunately, this service relies heavily on web users that can sometimes falsely report the inaccessibility of a site, raising a need for differentiation between censorship and network error. CensMon does not have access to as many nodes as Herdict, but it can work in a complementary fashion since it uses an automated mechanism.

The Alkasir project [1] combines user-based reporting of blocked content with an anti-censorship tool that attempts to penetrate such filtering. It has the ability to track censorship of reported URLs to periodically verify if they continue to be blocked or got unblocked. Such a function enables it to measure the trends of Internet filtering in a practical way. As for the censorship circumvention is concerned, Alkasir uses proxy servers to allow users to circumvent censorship of URLs. Thus, it is predominantly used by persons in countries where there is censorship of political content such as news, opinion articles, blog entries, forum discussions, political videos etc.

Finally, there a few other existing applications that monitor Internet censorship events which are either focused on a specific country [19] or their agent network is still quite limited to have an accurate global view [27].

5.2 Measuring Censorship and Detecting Filtered Web Content

Zittrain and Edelman [107] have found a number of blocked websites in China associated with sensitive material. They used URLs from search results from web searches as input to test for blocking, making their probing more efficient and more targeted. In order to evaluate our system we used their methodology as one source of our system's inputs.

In 2007, Crandall et al. [62] proposed ConceptDoppler, an approach based on Latent Semantic Analysis (LSA) [82] to semi-automatically extract filtered keywords in China. In ConceptDoppler, words that were related to concepts that are deemed sensitive were extracted using LSA and then active measurements were conducted to evaluate their results. Moreover, Park and Crandall's latest work [90] is focusing on HTML response filtering and the discontinuation of this technique in China. Both these works influenced our methodology in order to spot URL keyword filtering and differentiate it from HTTP response keyword filtering and HTTP Header keyword filtering.

Mathrani et al. [84] tried to get a snapshot of censorship in 10 countries. Their design methodology of their probing mechanism influenced us in enabling us to detect the root cause of the filtering that was reported.

In August 2011, during USENIX workshop on Free and Open Communications on the Internet (FOCI'11) [50], Wright et al. [102] presented their work on examining the problem of mapping Internet censorship at a finer-grained level than the national. They are based on the fact that users accessing the internet through different providers or services, may experience differences in the filtering applied to their internet connectivity. Moreover, Wright et al. seek to stimulate discussion concerning the potentially serious legal and ethical concerns that are intrinsic to this form of research.

During *CAIDA Workshop on BGP and Traceroute data* [6], recent works concerning Internet censorship were presented among others. Gupta [74] pinpoints the lack of ongoing projects to measure censorship in a technically sound way. Gupta's goal is to measure who censors what, how and when on an ongoing basis. Throughout her preliminary experiments, which are concentrated in China an Iran, she used 20 free proxies in each of these two country. By conducting accessibility tests through these proxies for known blocked sites (and sites resulting from Google searches), she found that what is blocked changes over time, supporting the need for ongoing measurements despite the fact that some sites are blocked throughout her measurement period of about a month. As far as Dainotti's work [63] is concerned, Dainotti used combined different measurement sources

(BGP data, active traceroute probing and Internet background radiation) in order to analyze country-wide Internet outages caused by censorship attempts.

5.3 Studies of Deployed Filtering Mechanisms

Dornseif [68] and Clayton [60] both give detailed deconstructions of particular implementations of Internet censorship in Germany and the United Kingdom, respectively. These two studies have looked at the effectiveness of Internet Service Providers filtering out web sites that are known to contain child pornography (UK case) and Nazi related content (Germany case). Clayton in his study of the content blocking system *cleanfeed* they discovered that forbidden content could be trivially accessed. Furthermore, the blocking mechanism had precisely the opposite of its intended effect in that it could be used as an oracle for interested parties to discover sites with illegal material.

The methods of China's HTTP keyword filtering were first published by the Global Internet Freedom Consortium [52]. Howerver, Clayton et al. [61] give a detailed presentation of how web content blocking works. The main focus of their work is to reveal the mechanisms behind the Great Firewall of China. Moreover, they propose a naive but effective way to circumvent the Chinese firewall by just ignoring the injected TCP RST packets that the firewall generates. This work further motivates our effort and provide us with information about characteristics of the specific content blocking mechanism.

In 2007, Crandall et al. [62] with their work disprove the notion that GFC keyword filtering is a firewall strictly at the border of China's Internet. They suggested that the GFC's keyword filtering is more a panopticon than a firewall, i.e., it need not block every illicit word, but only enough to promote self-censorship. China's largest ISP, ChinaNET, performed 83.3% of all filtering of their probes, and 99.1% of all filtering that occurred at the first hop past the Chinese border. Filtering occurred beyond the third hop for 11.8% of their probes, and there were sometimes as many as 13 hops past the border to a filtering router. Approximately 28.3% of the Chinese hosts they sent probes to were reachable along paths that were not filtered at all.

Villeneuve [96] demonstrated that the chat functionality of TOM-Skype [47] triggers on certain keywords, preventing their communication and uploading messages to a server in China. He provided some high-level analysis of what is censored and how this mechanism works. Knockel et al. [77] provide a more detailed analysis of TOM-Skype, including the algorithms for protecting the keywords that trigger censorship and surveillance. Based on their data, Knockel et al. present five conjectures that they believe to be formal enough to be hypotheses that the Internet censorship research community could potentially answer with more data and appropriate computational and analytic techniques.

Xu et al. [104] use low-level network characteristics in order to detect the location of the filtering devices. To accomplish that, they use PlanetLab nodes for their probes, an approach that we follow as well.

The analysis and the reverse engineering of these previous deployed filtering systems gave us insight of how these systems work in order to design our system.

5.4 Cersorship Circumvention and Anti-Censorship Systems

There has been sustained interest in this decade in both the research and development of circumvention tools. Works here are divided in two major categories. In the first we can find simple web based anonymous proxies (like Proxify [41], DynaWeb [8], Guardster [20], Anonymouse [4] and Anonymizer [3]). In the second category we can find standalone applications that either enable Internet users to view websites blocked taking advantage of a range of open proxies or dedicated nodes to access remote content (like Freegate [12], GPass [18] and UltraSurf [49]) or voluntarily turn users' machines in encrypted web proxies and enable web users from censored countries connect in order to access filtered content (like Psiphon [51]). Finally, Peacefire [38] provides Internet user with information on how to bypass web filter and gives out through a email list new web circumvention sites.

Tor [66] aims to provide end-user anonymity with constraints such as low-latency, deployability, usability, flexibility, and simple design. Tor is a distributed circuit-switching overlay network consisting of over two-thousand volunteer-run Tor routers operating around the world. Tor clients achieve anonymity by source-routing their traffic through three Tor routers using onion routing [73]. Thus, it is complex and resource intensive for eavesdropping attackers and malicious nodes within the network to link the originator of a circuit to the destination.

A major problem with these approaches is how to distribute the IP addresses of proxies to users without their falling into the hands of the censors [85] [72] [83] [94]. Over time, it is expected that the censors can enumerate all proxy IP addresses [86], allowing them to block new users as well as identify past users in traces. As a result, some services cycle the hosts through a range of IP addresses to avoid blacklists. The effectiveness of this evasion technique depends upon the number of available IP addresses and their distribution across IP address space.

Numerous Censorship Resistant Systems have been developed during the last decade. One attempt at censorship resistant web publishing is the Publius system [98]. Publius makes use of many cryptographic elements and uses Shamir's threshold secret sharing scheme [93] to split the shares amongst many servers. Because servers do not store the entire key for a particular document, and documents are stored encrypted, Publius suggests that it achieves server deniability, the ability for a server to deny knowledge of the hosted documents' contents.

GNUnet [13] is an anonymous file-sharing network in which files are broken into fixed-size blocks and each block is stored under its own key. Since this creates a large index relative to the amount of data, a Bloom filter [57] is used to reduce the size of the in-memory index. Kugler [80] describes a predecessor attack against reader anonymity in GNUnet. If a file is held permanently on a single node rather than being inserted into the network, a similar attack can be used to find the publisher.

Free Haven [65] is a peer-to-peer network that provides anonymity and document persistence. In Free Haven, each stored document is divided into shares that are signed with the document's private key. The shares are stored on a server along with the hash of the corresponding public key. Clients retrieve documents with a get(name) interface where name is the hash of the document's public key. A request is broadcast to the entire network; servers holding the matching key hash respond with the stored shares. The client recreates the file upon receiving a sufficient number of shares.

Freenet [59] consists of volunteer servers that provide a document store. A document in Freenet is encrypted with a descriptive name as its key and requested using Freenet's get(name) interface where name is its content key. Queries are forwarded to servers hosting names which offer the closest match to name. If the document is found, the server reverse-routes doc back to the client, and each server on the return route caches a copy of the document.

Tangler [97] is a network of servers that provide data storage. Censorship resistance is provided by "entangling" documents such that the removal of one document will result in the removal of other documents. In Tangler, each document is divided into blocks, each of which is entangled (using Shamir secret sharing [93]) with two arbitrary blocks in the system. Each entanglement creates two new blocks in addition to the two existing ones. A threshold number of entangled blocks reconstruct the original block.

Feamster et al. proposed Infranet [71], a framework to use covert channels in HTTP to circumvent censorship. Web servers participating in Infranet receive covert requests for web pages encoded as a sequence of HTTP requests to harmless web pages and return the content hidden inside harmless images using steganography. The client uses a covert channel based on the sequence of HTTP requests to communicate what true destination it wants to reach; the proxy then fetches the data and uses steganography to embed it inside images that it serves back to the client. However, as with other proxy approaches, it relies on the censor not knowing the proxy address.

In 2010, Burnett et al. [58] proposed a system, called Collage, that hides parts of the steganogram in user-generated content on the Internet. Rather than relying on a single system or set of proxies to circumvent censorship firewalls, Burnett proposes the vast deployment of sites that host user-generated content to breach these firewalls and filtering mechanisms. Client requests are, likewise, sent via such images, improving unobservability, but significantly reducing the interactive performance of the client. A second problem is that a censor might discover which sites are used for such communication and block them entirely; for example, Burnett et al. consider sending user generated content via Twitter [48], which has already been blocked by several countries.

In 2011, 3 research groups proposed a new approach to circumvent state-level Internet censorship which is markedly different from past anticensorship efforts. Telex [103], Decoy Routing [76] and Cirripede [75] circumvention systems embed the proxy within the network itself making the blocking of individual web sites ineffective. Through cooperation from large ISPs, anticensorship technology is placed into the Internet's core network infrastructure and it is easy to distribute and very difficult to detect and block. Finally, these systems intercept connections from clients to innocent-looking destinations and redirects them to the true destination requested by the client.

The aforementioned systems are more sophisticated that the previous ones and more or less their goal is to provide to their users the ability to publish or to gain access to censored material. Most of these systems aim to offer their service based on design goals such as anonymity, deniability, confidentiality and unlink-ability. We believe that it is necessary for the users of these systems to know what exactly is censored hence, a service like the one provided from CensMon will be useful to them.

5.5 Net Neutrality

Lately, an intense and wide-ranging policy debate on network neutrality and ISP traffic management has begun. It have been reported that some certain ISPs do not permit their costumers to upload data when they use BitTorrent. This phenomenon motivated Dischinger et al. [67] to develop Glasnost, a tool that imitates BitTorrent's protocol message exchange in order to conduct a measurement study that examined if BitTorrent uploads are blocked throughout many different ISPs.

Hereupon, Tariq et al. [95] proposed NANO a system that its purpose is to infer in what extend an ISP policy is responsible for the performance degradation of a particular service. Beverly et al. [56] leveraged the “referral” feature of Gnutella to conduct TCP port reachability tests from 72,000 unique Gnutella clients, finding that Microsoft's network filesharing ports are frequently blocked, and that email-related ports are more than twice as likely to be blocked as other ports.

Reis et al. [91] used JavaScript-based “web tripwires” to detect modifications to HTTP-borne HTML documents. Reis et al. have shown that a nontrivial number of modifications occur to web pages on their journey from servers to browsers.

NetPolice [105] measured traffic differentiation in 18 large ISPs for several popular services in terms of packet loss, using multiple end points inside a given ISP to transmit application-layer traffic to destinations using the same ISP egress points. They found clear indications of preferential treatments for different kinds of service.

Huang et al. [45] released a network tester for smartphones to detect hidden proxies and service blocks.

In 2010, Krebich et al. [79] proposed Netalyzr, a network measurement and debugging service that evaluates the functionality provided by people's Internet connectivity. Netalyzr can act as an ongoing service for illuminating edge network neutrality, security, and performance.

ISPs try to grow their profit margins by employing "error traffic monetization", the practice of redirecting customers whose DNS lookups fail to advertisement-oriented Web servers. In 2011, Weaver et al. [101] have conducted a technical analysis of DNS error traffic monetization evident in 66,000 Netalyzr sessions, including fingerprinting derived from patterns seen in the resulting ad landing pages. Weaver et al. identified major players in this industry, their ISP affiliations over time, and available user opt-out mechanisms.

We believe that web censorship is a facet of network neutrality violation so the previous works are very related to ours . In spite of that, the focus of our work is not to provide the means for detecting ISP misbehavior but information about filtered web content throughout the world.

5.6 Censorship of Search Results

CenSEARCHip [7] is a project where one can have comparisons concerning Web search and image search functions of four national versions of Google and Yahoo!: the United States, China, France, and Germany.

Wang [99], compared the search return results from China with the results from New Zealand with 200 English keywords. Finally, Zhu et al. [106] conducted a set of experiments on major search engines employed by Internet users in China, issuing queries against a variety of different words. While their results don't offer any fundamental insight into how to defeat or work around Chinese internet censorship, they are still helpful to understand the structure of how censorship duties are shared between the Great Firewall and Chinese search engines.

6
Future Work

One limitation of our work for now is the fact that CensMon can only monitor the extent of censorship of a country within which we are able to access a PlanetLab node. However, we can overcome this issue by using web proxies worldwide as agents conducting simple accessibility tests, Tor exit nodes or even by developing a software client or even a Firefox add-on that will act as a CensMon agent (having previously informed the user about the potential risks and having got his consent).

Another direction that we plan to explore further is monitoring specific news sites via RSS feeds, and measure Internet censorship concerning realtime news events. Moreover, we can use our agent network to measure filtering of non-HTTP traffic and ports (e.g. P2P, SMTP, VPN etc.) or execute network level probes in order to test network infrastructure in terms of censorship.

As Rogers proposed in [92], we can make use of CensMon's infrastructure to find known blocked content on unblocked sites within a country. This way we could study how the content at censored websites is redistributed in uncensored sites.

Moreover, when CensMon detects an IP blocking event we could detect overblocking by looking for other websites that are resolved to the same blocked IP address. This can easily be achieved by using free reverse IP lookup services such as MyIP-Neighbors [26].

Finally, by conducting long-term measurements and since in CensMon each URL is forwarded to all its agents, we could detect common domains censored by different countries. As a result, we can extract information about common worldwide web filtering trends among countries.

7 Discussion

7.1 Approaches

There are limitations from the nature of Internet censorship itself. As Warf [100] stated in his work, “there is a highly uneven topography of Internet censorship around the globe, one that reflects the geographies of the world’s diverse political systems, the extent of Internet penetration rates, the social, cultural, and economic constitutions of various societies, and the degree of political opposition. Such complexity means that patterns of Internet censorship do not lend themselves readily to pat characterizations but require a more detailed, case-by-case analysis”.

OpenNet Initiative uses volunteers and direct means to examine filtering around the world. This way it publishes frequent per country reports concerning filtering in this specific country. Moreover, HerdictWeb crowdsources filtering information from volunteer web users. Wright et al. [102] from Oxford Internet Institute study filtering across a state, as opposed to our approach of studying censorship at a national scale (assuming to some extent homogeneous national filtering) but varying over time. They expect that censorship at the national level need not be applied equally across a country. Their approach as well as their initial results in China have shown that there is indeed differences in filtering patterns across China. Finally, The Alkasir project combines user-based reporting of blocked content with an anti-censorship tool that attempts to penetrate such filtering. The key points of the above approaches can be seen in Table 7.1.

Censorship Study Approach	Key Points
HerdictWeb	Uses crowdsourcing Homogeneous national filtering
OpenNet Initiative	Uses volunteers and direct means Per country reports
Oxford Internet Institute	DNS server probing Localised filtering Input from HerdictWeb
Alkasir Project	User-based reporting of blocked content Censorship tracking Anti-censorship tool
CensMon	Uses direct investigation Detect filtering technique Input from a plethora of sources

TABLE 7.1: Existing Censorship Study Approaches.

7.2 Legal and Ethical Implications

The aforementioned approaches raise some topics/challenges concerning filtering that should be taken into consideration. These topics concerning legality and ethics where introduced by Wright et al. [102] during the USENIX Free and Open Communications Workshop (FOCI'11).

- *Is it legal to access blocked sites?* Web users who request blocked content are likely to face repercussions since such attempts may be logged. It would be impractical for a state to take note of every blocking action taken by their filtering mechanism. It is possible, however, that sufficiently high-volume requests for banned content may raise suspicions and be considered worthy of further action.

- *Is direct investigation needed so as to study censorship?* The simplest way to learn how an individual computer's connection is filtered is by contacting a remote web user and asking them to run censorship detection code themselves. Despite of the fact that such an approach is dufficult to scale, researchers can get a great amount of information from such an experiment. Moreover, through direct investigation researchers can find not only the blocking status but also the type of blocking. Nevertheless, the problem in such cases of direct investigation is that these direct services are expensive as well as rare.

- *Is it ethical to ask someone else to access blocked websites?*

  Volunteers that participate in Internet censorship research by running a filtering detection tool must having been fully informed for the nature of the tool and the potential risks involved. As a result, there is an important added burden on the researcher to state to the participant, who may well not have any significant level of technical expertise, what the tool will do and what particular risks they run. The critical point here is the *user's consent* since otherwise it would be considered as unethical.

8 Conclusion

Censorship on the world wide web appears to be taking place more than ever before. Nevertheless, most existing resources of Internet censorship do not provide detailed technical information. They are mainly based on accessibility tests and a lot of current censorship measurement efforts are mostly conducted by journalists without technical background. As a result, there is a lack in measuring who censors what, how, when on an ongoing basis since what is blocked changes over time. Moreover, by knowing how filtering is technically conducted, one can design practical anti-censorship evasion techniques.

In this thesis we presented CensMon, an Internet censorship monitoring infrastructure. The aforementioned increase of global web censorship motivated us to design and build a system that can detect filtering characteristics and also be capable of differentiating between censorship and network failures. We implemented our design and evaluated it on the PlanetLab testbed using information streams automatically extracted from a plethora of sources. Based on our experience with using CensMon, as well as on the experimental results presented in this thesis, we believe that CensMon can be a valuable resource for Internet censorship detection, and can provide useful information for both researchers and regular web users.

Bibliography

[1] Alkasir - Mapping and Circumventing cyber-Censorship. https://alkasir.com.

[2] An Open Letter From Internet Engineers to the U.S. Congress. Electronic Frontier Foundation. https://www.eff.org/deeplinks/2011/12/internet-inventors-warn-against-sopa-and-pipa.

[3] Anonymizer. http://www.anonymizer.com.

[4] Anonymouse. http://anonymouse.org.

[5] Arc90's Readability. http://www.readability.com.

[6] CAIDA Workshop on BGP and Traceroute data. http://www.caida.org/workshops/bgp-traceroute.

[7] CenSEARCHip. http://carl.cs.indiana.edu/censearchip.

[8] DynaWeb. http://www.dongtaiwang.com/home_en.php.

[9] Egypt leaves the Internet. http://www.renesys.com/blog/2011/01/egypt-leaves-the-internet.shtml.

[10] Egypt's Internet Shut Down, According To Reports. http://www.huffingtonpost.com/2011/01/27/egypt-internet-goes-down-_n_815156.html.

[11] Freedom House. Freedom on the Net 2011. http://www.freedomhouse.org/template.cfm?page=664.

[12] Freegate, Dynamic Intenet Technology. http://www.dit-inc.us/freegate.

[13] GNUnet. http://www.gnunet.org.

[14] Google excluding controversial sites. http://news.cnet.com/2100-1023-963132.html.

[15] Google Hot Trends. https://www.google.com/trends/hottrends.

[16] Google Transparency Report. Libya Traffic Divided by Worldwide Traffic and Normalized. http://www.google.com/transparencyreport/traffic/?r=LY&l=EVERYTHING&csd=1296862200000&ced=1299281400000.

[17] Google Web Alerts. http://www.google.com/alerts.

[18] GPass. http://gpass1.com/gpass.

[19] GreatFirewall.biz. http://www.greatfirewall.biz.

[20] Guardster. http://www.guardster.com.

[21] Homeland Security Request to Take Down MafiaaFire Add-on. http://lockshot.wordpress.com/2011/05/05/homeland-security-request-to-take-down-mafiaafire-add-on.

[22] International Telecommunication Union. Measuring the Information Society. http://www.itu.int/ITU-D/ict/publications/idi/2010/Material/MIS_2010_without_annex_4-e.pdf.

[23] Libya Protests: Anti-Government Protesters Killed During “Day Of Rage”. http://www.huffingtonpost.com/2011/02/17/libya-protests-antigovern_0_n_824826.html.

[24] MAFIAAFire Redirector Mozilla Add-on. https://addons.mozilla.org/en-US/firefox/addon/mafiaafire-redirector.

[25] Mozilla resists US government request to nuke ”MafiaaFire” add-on. http://arstechnica.com/tech-policy/news/2011/05/mozilla-resists-us-govt-request-to-nuke-mafiaafire-add-on.ars.

[26] MyIPNeighbors : Reverse IP Lookup. http://www.my-ip-neighbors.com.

[27] Net Neutrality Monitor. http://www.neumon.org.

[28] Obama Speech Censored in China. http://news.bbc.co.uk/2/hi/7841580.stm.

[29] Open Net Initiative. http://www.opennet.net.

[30] OpenNet Initiative. Global Filtering Map. Conflict and Security Content. http://http://map.opennet.net/filtering-consec.html.

[31] OpenNet Initiative. Global Filtering Map. Internet Tools Content. http://http://map.opennet.net/filtering-IT.html.

[32] OpenNet Initiative. Global Filtering Map. Political Content. http://http://map.opennet.net/filtering-pol.html.

[33] OpenNet Initiative. Global Filtering Map. Social Content. http://http://map.opennet.net/filtering-soc.html.

[34] OpenNet Initiative. Internet Filtering in China. http://opennet.net/sites/opennet.net/files/ONI_China_2009.pdf.

[35] OpenNet Initiative's Herdict Web. http://www.herdict.org/web.

[36] Overview of the Egyptian Internet Shutdown. http://www.pch.net/resources/misc/Egypt-PCH-Overview.pdf.

[37] paramiko : SSH2 protocol for python. http://www.lag.net/paramiko.

[38] Peacefire.org. http://www.peacefire.org.

[39] Plan for Greece Favors Creditors. http://www.nytimes.com/2011/07/26/business/plan-for-greece-favors-creditors.html?_r=1.

[40] PlanetLab. http://www.planet-lab.org.

[41] Proxify. https://proxify.com.

[42] Pseudonymity. Freedom of Communication on the Internet Workshop (FOCI): Fine-Grained Censorship Mapping - Information Sources, Legality and Ethics. http://www.pseudonymity.net/?p=74.

[43] Pulling the Plug: A Technical Review of the Internet Shutdown in Burma. http://opennet.net/research/bulletins/013.

[44] Slashdot. Your Rights Online. http://yro.slashdot.org.

[45] The UMich Smartphone 3G Test. http://www.eecs.umich.edu/3gtest.

[46] The Web Goes On A SOPA Strike. http://www.forbes.com/sites/kashmirhill/2012/01/18/the-web-goes-on-a-sopa-strike-with-the-oatmeal-doing-it-best/.

[47] TOM Skype. http://skype.tom.com.

[48] Twitter. http://twitter.com.

[49] UltraSurf. `http://ultrasurf.us`.

[50] USENIX workshop on Free and Open Communications on the Internet (FOCI'11). `https://db.usenix.org/events/foci11`.

[51] U.Toronto Citizen Laboratory. Psiphon. `http://psiphon.ca`.

[52] The great firewall revealed. *Global Internet Freedom Consortium*, 2002.

[53] PROTECT IP Technical Whitepaper. `http://www.circleid.com/pdf/PROTECT-IP-Technical-Whitepaper-Final.pdf`, 2011.

[54] R. Atkinson, S. Ezell, S. Andes, D. Castro, and R. Bennett. The internet economy 25 years after. com. *Washington, DC: Information Technology and Innovation Foundation*, 2010.

[55] M. Bailey and C. Labovitz. Censorship and co-option of the internet infrastructure. *Ann Arbor*, 1001:48104, 2011.

[56] R. Beverly, S. Bauer, and A. Berger. The internet is not a big truck: toward quantifying network neutrality. *Passive and Active Network Measurement*, pages 135–144, 2007.

[57] B. Bloom. Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13(7):422–426, 1970.

[58] S. Burnett, N. Feamster, and S. Vempala. Chipping away at censorship firewalls with user-generated content.

[59] I. Clarke, O. Sandberg, B. Wiley, and T. Hong. Freenet: A distributed anonymous information storage and retrieval system. In *Designing Privacy Enhancing Technologies*, pages 46–66. Springer, 2001.

[60] R. Clayton. Failures in a hybrid content blocking system. In *Privacy Enhancing Technologies*, pages 78–92. Springer, 2006.

[61] R. Clayton, S. Murdoch, and R. Watson. Ignoring the great firewall of china. In *Privacy Enhancing Technologies*, pages 20–35. Springer, 2006.

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

[63] A. Dainotti. Analysis of country-wide internet outages caused by censorship. In *CAIDA Workshop on BGP and Traceroute data August 22nd, 2011-San Diego (CA), USA*.

[64] A. Dainotti, C. Squarecella, E. Aben, K. Claffy, M. Chiesa, M. Russo, and A. Pescape. Analysis of Country-wide Internet Outages Caused by Censorship. In *Internet Measurement Conference (IMC)*, Berlin, Germany, Nov 2011. Internet Measurement Conference (IMC).

[65] R. Dingledine, M. Freedman, and D. Molnar. The free haven project: Distributed anonymous storage service. In *Designing Privacy Enhancing Technologies*, pages 67–95. Springer, 2001.

[66] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th conference on USENIX Security Symposium-Volume 13*, pages 21–21. USENIX Association, 2004.

[67] M. Dischinger, A. Mislove, A. Haeberlen, and K. Gummadi. Detecting bittorrent blocking. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, pages 3–8. ACM, 2008.

[68] M. Dornseif. Government mandated blocking of foreign Web content. *Arxiv preprint cs/0404005*, 2004.

[69] I. Dubrawsky. Firewall evolution-deep packet inspection. *Security Focus*, 29, 2003.

[70] R. Faris and N. Villeneuve. Measuring global internet filtering. *Access denied: The practice and policy of global Internet filtering*, pages 5–28, 2008.

[71] N. Feamster, M. Balazinska, G. Harfst, H. Balakrishnan, and D. Karger. Infranet: Circumventing web censorship and surveillance. In *Proceedings of the 11th USENIX Security Symposium*, pages 247–262. USENIX Association, 2002.

[72] N. Feamster, M. Balazinska, W. Wang, H. Balakrishnan, and D. Karger. Thwarting web censorship with untrusted messenger discovery. In *Privacy Enhancing Technologies*, pages 125–140. Springer, 2003.

[73] D. Goldschlag, M. Reed, and P. Syverson. Hiding routing information. In *Information Hiding*, pages 137–150. Springer, 1996.

[74] M. Gupta. Preliminary experiments on measuring web censorship around the world. In *CAIDA Workshop on BGP and Traceroute data August 22nd, 2011- San Diego (CA), USA*.

[75] A. Houmansadr, G. Nguyen, M. Caesar, and N. Borisov. Cirripede: Circumvention infrastructure using router redirection with plausible deniability. 2011.

[76] J. Karlin, D. Ellard, A. Jackson, C. Jones, G. Lauer, D. Mankins, and W. Strayer. Decoy routing: Toward unblockable internet communication. In *USENIX Workshop on Free and Open Communications on the Internet*, 2011.

[77] J. Knockel, J. Crandall, and J. Saia. Three researchers, five conjectures: An empirical analysis of tom-skype censorship and surveillance.

[78] J. Kornblum. Identifying almost identical files using context triggered piecewise hashing. *digital investigation*, 3:91–97, 2006.

[79] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson. Netalyzr: Illuminating the edge network. In *Proceedings of the 10th annual conference on Internet measurement*, pages 246–259. ACM, 2010.

[80] D. Kugler. An analysis of gnunet and the implications for anonymous, censorship-resistant networks. In *Privacy Enhancing Technologies*, pages 161–176. Springer, 2003.

[81] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. In *ACM SIGCOMM Computer Communication Review*, volume 40, pages 75–86. ACM, 2010.

[82] T. Landauer, P. Foltz, and D. Laham. An introduction to latent semantic analysis. *Discourse processes*, 25(2):259–284, 1998.

[83] M. Mahdian. Fighting censorship with algorithms. In *Fun with Algorithms*, pages 296–306. Springer, 2010.

[84] A. Mathrani and M. Alipour. Website Blocking Across Ten Countries: A Snapshot. 2010.

[85] D. McCoy, J. Morales, and K. Levchenko. Proximax: A measurement based system for proxies dissemination.

[86] J. McLachlan and N. Hopper. On the risks of serving whenever you surf: vulnerabilities in tor's blocking resistance design. In *Proceedings of the 8th ACM workshop on Privacy in the electronic society*, pages 31–40. ACM, 2009.

[87] P. Mockapetris. Rfc 1035–domain names-implementation and specification. *Internet Engineering Task Force*, 1987.

[88] S. Murdoch and R. Anderson. Tools and technology of internet filtering. *Access Denied: The Practice and Policy of Global Internet Filtering (Information Revolution and Global Politics Series)*, pages 57–72, 2008.

[89] W. Noonan and I. Dubrawsky. *Firewall fundamentals*. Cisco Press, 2006.

[90] J. Park and J. Crandall. Empirical Study of a National-Scale Distributed Intrusion Detection System: Backbone-Level Filtering of HTML Responses in China. In *2010 International Conference on Distributed Computing Systems*, pages 315–326. IEEE, 2010.

[91] C. Reis, S. Gribble, T. Kohno, and N. Weaver. Detecting in-flight page changes with web tripwires. In *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*, pages 31–44. USENIX Association, 2008.

[92] R. Rogers. A New Media Approach to the Study of State Internet Censorship. `http://www.govcom.org`, 2009.

[93] A. Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, 1979.

[94] Y. Sovran, A. Libonati, and J. Li. Pass it on: Social networks stymie censors. In *Proceedings of the 7th international conference on Peer-to-peer systems*, pages 3–3. USENIX Association, 2008.

[95] M. Tariq, M. Motiwala, and N. Feamster. Nano: Network access neutrality observatory. In *Proceedings of ACM HotNets*. Citeseer, 2008.

[96] N. Villeneuve. Breaching trust: an analysis of surveillance and security practices on china's tom-skype platform. 2008.

[97] M. Waldman and D. Mazieres. Tangler: a censorship-resistant publishing system based on document entanglements. In *Proceedings of the 8th ACM conference on Computer and Communications Security*, pages 126–135. ACM, 2001.

[98] M. Waldman, A. Rubin, and L. Cranor. Publius: A robust, tamper-evident, censorship-resistant web publishing system. In *Proceedings of the 9th conference on USENIX Security Symposium-Volume 9*, pages 5–5. USENIX Association, 2000.

[99] N. Wang. Control of internet search engines in china–a study on google and baidu. *Unpublished thesis submitted in partial fulfillment of the degree of Master of Computing, Unitec Institute of Technology, New Zealand*, 2008.

[100] B. Warf. Geographies of global Internet censorship. *GeoJournal*, pages 1–23.

[101] N. Weaver, C. Kreibich, and V. Paxson. Redirecting dns for ads and profit.

[102] J. Wright, T. de Souza, and I. Brown. Fine-grained censorship mapping information sources, legality and ethics.

[103] E. Wustrow, S. Wolchok, I. Goldberg, and J. Halderman. Telex: Anticensorship in the network infrastructure. In *proceedings of the 20th USENIX Security Symposium*, 2011.

[104] X. Xu, Z. Mao, and J. Halderman. Internet Censorship in China: Where Does the Filtering Occur? In *Passive and Active Measurement*, pages 133–142. Springer, 2011.

[105] Y. Zhang, Z. Mao, and M. Zhang. Detecting traffic differentiation in backbone isps with netpolice. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 103–115. ACM, 2009.

[106] T. Zhu, C. Bronk, and D. Wallach. An analysis of chinese search engine filtering. *Arxiv preprint arXiv:1107.3794*, 2011.

[107] J. Zittrain and B. Edelman. Internet filtering in China. *Internet Computing, IEEE*, 7(2):70–77, 2003.