



Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

School of Computer Science and Statistics

Comparing Internet Censorship Between Ireland And Saudi Arabia

Ahmed Mahdi

Supervisor: Dr Stephen Farrell

April 15, 2024

A dissertation submitted in partial fulfilment
of the requirements for the degree of
B.A.(Mod) computer Science and German

Declaration

I hereby declare that this dissertation is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at <http://www.tcd.ie/calendar>.

I have completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at <http://tcd-ie.libguides.com/plagiarism/ready-steady-write>.

I consent / do not consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

I agree that this thesis will not be publicly available, but will be available to TCD staff and students in the University's open access institutional repository on the Trinity domain only, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement. **Please consult with your supervisor on this last item before agreeing, and delete if you do not consent**

Signed: _____

Date: _____

Acknowledgements

I express my sincere gratitude to my supervisor, Stephen, whose guidance and encouragement were invaluable throughout my thesis. I am deeply thankful to my family for their unwavering support and belief in my abilities. I also want to thank my friends who stood by me during the difficult times and provided me with companionship and solidarity, especially during the long hours we spent in the glass rooms.

Contents

1	Introduction	1
1.1	Internet censorship	1
1.1.1	User Privacy	2
1.2	Project goals	3
1.3	Report Structure	4
2	State of the Art	5
2.1	Global Context	5
2.2	Censorship Mechanisms and Circumvention Techniques	6
2.2.1	Points of Control	6
2.2.2	Network-Level Filtering	7
2.2.3	Multi-layer and Non-layer Specific Filtering	10
2.2.4	Filtering by Example	13
2.3	Country Specific Practices	14
2.3.1	Ireland	14
2.3.2	Saudi Arabia	16
2.3.3	Effects of Censorship	18
2.4	Circumvention Tools	18
2.4.1	VPN	19
2.4.2	TOR, the onion router	19
2.4.3	Proxy Servers	20
2.4.4	Psiphon	21
2.4.5	Challenges & Limitations	21
2.5	Challenges and Concluding thoughts	21
3	Methodology	23
3.1	Introduction	23
3.1.1	Description of OONI	23
3.2	Real-Time Data Collection	25
3.2.1	Data Collection in Ireland	25

3.2.2	Data Collection in Saudi	26
3.2.3	Process Overview	27
3.3	Challenges and Solutions	30
4	Results and Evaluation	31
4.1	OONI Curated Lists	31
4.1.1	Saudi Arabia	31
4.1.2	Ireland	34
4.2	Comparison	38
4.3	Top 100 Websites Worldwide	41
4.4	Random URLs	42
4.5	Cross-Comparison of Most visited Websites in both countries	43
4.6	Circumvention Tests	44
4.7	Deep analysis of blocking methods	45
4.8	Conclusion	48
5	Conclusion	49
5.1	Conclusion	49
5.2	Future work	49
A1	Appendix	58

List of Figures

2.1	Overview of nation-state censorship methods, adapted from Master and Garman's study on global internet restrictions.	6
2.2	OSI model	8
2.3	The Fundamental Process of a DNS Resolution and Two Main Techniques in DNS Tampering	12
2.4	Block page seen When accessing Russia.tv	16
2.5	Ireland Measurements on the OONI database taken by other people who have used OONIprobe to test Russia.TV	16
2.6	Ireland Measurements on the OONI database taken by other people who have used OONIprobe to test RT.com	16
2.7	A basic architecture of the TOR network	20
3.1	Measurement taken on our VPS based in Riyadh	25
3.2	same measurement on the Ooni explorer showcasing real-time update of our data	25
3.3	Tests ran in both countries.	27
3.4	JSON scraper function	28
3.5	CSV code	28
3.6	Code normalising the domain names for the API to recognise	29
3.7	Function Interacting with the OONI API	30
4.1	Censorship Mechanisms dominant in SA	34
4.2	OONI Data for Vesti.ru	35
4.3	OONI Data for Rt.com	36
4.4	Censorship Mechanisms dominant in Ireland	36
4.5	OONI Curated List for Ireland tested in Saudi Arabia	39
4.6	Blocking Methods used for Saudi URLS in Ireland	40
4.7	Overlap between Ireland and Saudi Arabia	40
4.8	Comparing Ireland and Saudi Worldwide restrictions	42
4.9	OONI Measurement for VK	42

4.10 Graph comparing restrictions on 100 Random URLS	43
4.11 Ireland top 100 URLS tested in Saudi	44
4.12 Snippet from Saudi VPS	45
4.13 Traceroute Snippet	47

List of Tables

4.1	OONI Curated list for Saudi Arabia Tested in Saudi Arabia	32
4.2	OONI Curated list for Ireland Tested in Ireland	37
4.3	OONI Curated List for Saudi Arabia tested in Ireland	41

Abstract

This thesis investigates the contrasting approaches to internet censorship in Ireland and Saudi Arabia. Utilising data gathered through the OONI probe, supplemented by tests conducted on a Riyadh-based Virtual Private Server in Saudi Arabia, Saudi Arabia web filtering targets content related to Gambling, Anonymization and circumvention Tools, as well as Alcohol and drugs. Conversely, Ireland exhibits minimal filtering, primarily focused on preventing illegal file sharing, aligning with its commitment to internet freedom and following EU regulations. This study provides a detailed comparison of the censorship techniques used in both countries.

1 Introduction

1.1 Internet censorship

The primary purpose of this report is to critically compare the methods that Ireland and Saudi Arabia use to censor the Internet. According to Giuseppe Aceto and Antonio Pescapé [1], this phrase refers to the regulation or suppression of content available on the Internet. This filtering and blocking of content can hinder the free flow of information, impacting our ability to share, learn, and interact with one another.

The government may use censorship efforts to supposedly protect their people from sensitive or dangerous content, e.g., political dissent or adult material. The reason for these actions is often laid in the realm of preservation of cultural norms and safeguarding of national security [2]. These actions often violate freedom of speech and exclude significant discussions that might have led to social development [3]. Additionally, the issue of over-blocking represents a significant concern. This phenomenon occurs when censorship tools, designed to protect or exert control, unintentionally prevent access to various content [4]. The majority of this content, which is generally harmless and not sensitive, becomes unfairly caught in the wide reach of these techniques, contributing to an unwarranted restriction on the free flow of information and knowledge.

Censorship is not always about silencing the opposition or hiding sensitive information. Sometimes it's about protecting users, particularly young audiences, from explicit and harmful content. However, the criteria for what is considered "harmful" is subjective and varies across different cultures and legal systems. Deibert et al. [5] advocate for the notion that the internet should be a democratic space, with content regulated by the consensus of its global community of users, rather than by the unpredictable judgements of authorities. They critique the evolution of internet surveillance and control, arguing that this trend is detrimental to the liberty and empowerment that The Internet supposedly promotes.

1.1.1 User Privacy

The way a censor can monitor user traffic is also a significant aspect of censorship. They have the capability to track browsing habits, which raises concerns about the privacy of internet users. Analysing the destination IP addresses of web traffic, ISPs can identify and block websites, even if the content is encrypted, affecting user privacy [6]. It's important to note that ISPs are not the only ones capable of monitoring online behaviour. Prominent technology corporations, such as Google, Meta (previously known as Facebook), Apple, and Microsoft, command even more refined methodologies for tracking user interactions, predominantly via the deployment of HTTP cookies [7]. These mechanisms allow organisations to gather extensive information on users' online activities, highlighting general worries about digital privacy and monitoring.

In this context, Meta's recent conduct, as reported by Human Rights Watch, brings to light the complex nature of content moderation and its implications for censorship. According to the report, Meta has engaged in "systemic and global" censorship, particularly evident during the peak of the Israel-Gaza conflict. The company was found to have removed content and suspended or permanently banned accounts that showed peaceful support for Palestine. These actions were part of six key patterns of undue censorship, including reducing the reach of content, disabling accounts, and shadow banning, which decreases the visibility and impact of user material [8].

The advancement of web hosting technologies does impact the censor's ability to block content. In the realm of web hosting, the notion that websites rely solely on a single server for content delivery is progressively becoming outdated. The modern web hosting landscape is more diverse in architecture, especially when it comes to websites with high amounts of traffic. Advancing rapidly, cloud computing offers more robust and efficient alternatives to old models through server hosting services like Virtual Machines, Docker containers, and Serverless Architectures [9]. A large number of these websites now use cloud-based infrastructure, for instance, Amazon Web Services (AWS) offers distributed hosting solutions. These services play a crucial role in mitigating the hazards associated with centralised hosting paradigms, including server malfunctions or susceptibility to Distributed Denial of Service (DDoS) attacks. Moreover, the use of Content Delivery Networks (CDNs) and load Balancers significantly change the way content is shared online. CDNs help reduce delays by storing content on edge servers closer to users, making websites and media load faster [10]. Load Balancers spread incoming traffic over several servers to prevent any single server from overloading. This improves web service reliability and keeps sites running smoothly [11]. These technologies make the web more robust and scalable, allowing sites to perform well and stay available, even with many users or during cyber-attacks. Cloud computing, distributed hosting solutions, CDNs, and load balancers create a more complex

web architecture that can offer some resistance to censorship efforts. With content distributed across various servers and locations, it becomes more challenging for a censor to block or filter content, since there isn't a single point of failure, and in a shared hosting environment, multiple websites can be hosted on a single server and share the same IP address, leading to over-blocking. also, the widespread adoption of HTTPS over HTTP presents challenges for censorship. The encryption in HTTPS, provided by the TLS protocol, ensures that data integrity and privacy are maintained, making censorship attempts more complex. While plaintext HTTP traffic is less common and easier to intercept or block, modern censors must now employ advanced techniques such as Deep Packet Inspection (DPI) for encrypted traffic analysis. The shift towards a more diverse web architecture strengthens web service reliability and complicates the censor's task of filtering content in an HTTPS-dominated internet.

1.2 Project goals

The objective of this study is to compare internet censorship, its implementation, and its impacts in Ireland and Saudi Arabia using sophisticated tools that provide insights into the political and technological factors influencing information availability and actual evidence on internet restriction.

- **OONI (Open Observatory of Network Interference):** : OONI is a global observation network aimed at detecting censorship, surveillance, and traffic manipulation. The OONI Probe, a tool developed by OONI, conducts various internet measurement tests in over two hundred countries. The data, a collection of over a billion measurements from 26 thousand networks in 241 countries and territories, has been contributed by users of the OONI Probe worldwide. These users conduct censorship measurement tests within their local networks, sharing their results as open data in real time. One of the key tests, known as the "web connectivity" test, is designed to identify instances of internet censorship by detecting DNS tampering, TCP/IP protocol interference, or the use of a transparent HTTP proxy to block or filter web content [12].
- **Censored Planet:** Censored Planet is a remote censorship measurement platform that collects data on internet blocking from around the world without requiring users to install software. it offers a web dashboard that shows the results of internet censorship from a global perspective, covering more than 200 countries. It employs various remote sensing techniques to assess internet blocking and filtering. Their methodology includes using TCP/IP side channels to check connectivity between internet locations without a central observation point, leveraging public DNS resolvers to see how webpages are being resolved and identifying interference [13].

- **Virtual Private Server:** We selected a VPS provided by LightNode [14] that is based in Saudi Arabia for our investigation. we can employ the Open Observatory of Network Interference (OONI) Probe to run measurements across a curated selection of websites. These websites will span various categories to provide a comprehensive overview of accessibility within the region. The empirical data obtained from these experiments will supplement our research and provide a solid empirical foundation for our knowledge of internet restriction and digital rights in Saudi Arabia.

There are ethical considerations in studying internet censorship, especially when dealing with sensitive topics. This project will adhere to rigorous ethical guidelines, ensuring the research does not unintentionally harm individuals or groups by exposing sensitive information. For this project, we won't be looking into the blocking of content that's not appropriate for work (NSFW) or any material related to harming children. We know these topics are very important and need careful handling, but the censorship or availability of such content raises ethical issues that would require more effort than is available for this study. By leveraging these tools and methodologies, this study aims to comprehensively understand the strategies, objectives, and impacts of internet censorship across both countries.

1.3 Report Structure

This report is laid out in a way that will take you step by step through our study of internet censorship in the chosen countries. Here's what each chapter covers:

- **Chapter 2:** state of the art – We start off by looking at what's already known about internet censorship. This sets the stage for our own work.
- **Chapter 3:** Methodology – Next, we talk about how we planned our research, including what we wanted to find out and how we decided to go about it.
- **Chapter 4:** Results and Evaluation – This chapter gives you a look at the data we collected. We'll go over the initial findings and what they might mean.
- **Chapter 5:** Conclusion – In the last chapter, we take a moment to reflect on the study and think about what could be studied next.

2 State of the Art

This chapter delves into the evolution and current practices of internet censorship and control in Ireland and Saudi Arabi. We also provide a technical overview of censorship mechanisms and the tactics people use to bypass them. We want to understand the different ways these countries manage the online world. We chose these countries because they have varying approaches to internet regulation. Ireland has a relatively open online space, and Saudi Arabia has strict controls on specific categories of content

2.1 Global Context

In the global context, internet censorship is a reality that varies widely from one country to another. Around the world, many governments and corporations have put in place systems to monitor and sometimes restrict what can be seen or shared online. as highlighted by Alexander Master and Christina Garman's comprehensive study, which underscores the all-round nature of state-implemented internet censorship [2]. The study states that some countries exert extensive control to deny citizens free access to internet resources, with tactics ranging from centralised to decentralised censorship. Centralised censorship typically involves government-controlled infrastructure and is often straightforward, where a single entity controls access, like China's Great Firewall. Other countries like Syria, with one government-controlled autonomous system, can implement uniform censoring across its population. Decentralised censorship means restrictions vary regionally within a country. For instance, the Russian Federation's approach allows local ISPs to implement their own content filtering methods, resulting in varied internet access within the country. This decentralised model may deny access to some regions while others remain unrestricted. Examples include localised internet shutdowns in India and localised legal compliance by service providers, which may lead to different parts of the country having different degrees of network throttling or even content filtering [2].

The chart 2.1, derived from the study by Alexander Master and Christina Garman, summarises in simple form the mechanisms used by nation-states in internet censorship and contrasts the present and past data to highlight shifts in practice. It shows the influence of

socio-political and technological factors on implementing internet restrictions worldwide. Looking at patterns of control, the research provides a historical perspective on our understanding of modern-day censorship.

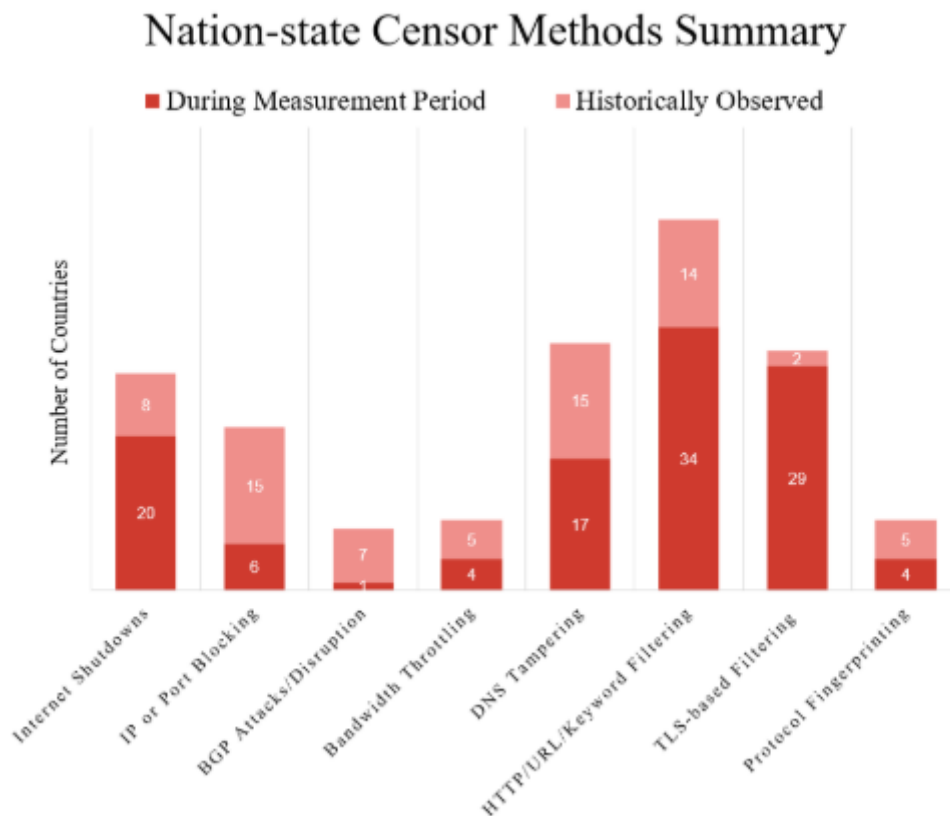


Figure 2.1: Overview of nation-state censorship methods, adapted from Master and Garman’s study on global internet restrictions.

2.2 Censorship Mechanisms and Circumvention Techniques

In this section, we look at two main things: how internet censorship works and how people try to get around it. First, we will discuss how governments and organisations block or filter information online and how this affects society. Then, we will explore what tools are used to bypass these blocks, keeping the internet open and free.

2.2.1 Points of Control

Points of control refer to specific layers within the technological infrastructure where censorship measures can be enforced, ranging from Internet Service Providers to personal devices. These control points are chosen based on the intended scope and effectiveness of the censorship. Common locations include:

Internet Service Providers (ISPs): ISPs can enforce censorship directives by blocking or throttling access to specific IP addresses or domains, effectively controlling the gateway to the internet for users within their network. Targeting specific Internet Protocol (IP) addresses for censorship typically occurs when the censor aims to block access to particular online services or content. For instance, a government might block an IP hosting a news website known for dissenting views, or an IP associated with an online forum discussing banned topics. This method is particularly effective for pinpointing and restricting access to targeted online services [15].

National Gateways: Some countries, such as China [5], Iran [16], Syria [17] and Arab Countries [18] implement censorship at the national level, using centralised filtering systems to monitor and control all Internet traffic entering or leaving the country.

IXP's (Internet Exchange points): Internet exchange Points (IXPs) serve as pivotal nodes within the global network infrastructure, enabling a high-capacity interconnection between Autonomous Systems (ASes) for the exchange of internet traffic [19]. Their centrality makes them potential tools for state or corporate control, presenting risks for surveillance and censorship. Such power over IXPs can directly impact internet freedom and digital rights

Local Networks: Organizations and educational institutions may apply censorship within their local networks to restrict access to certain content, using firewalls and content filtering software [20].

2.2.2 Network-Level Filtering

Network-level filtering, a mechanism implemented within layers 3 and 4 of the OSI model 2.2, checks each packet in real time as it traverses the network's filtering devices, for example, routers. This inspection depends on the packet's header content, leading to one of two outcomes: the packet is either transmitted to its destination or silently discarded. This silent type of filtering is common because it is integrated into most networking equipment out there, and it does not require any extra gadgets or new technologies to work so that the censor can use it easily. [21]. One of the big problems, however, is that users are kept in the dark; there is no indication or alert given when their data is being filtered out.

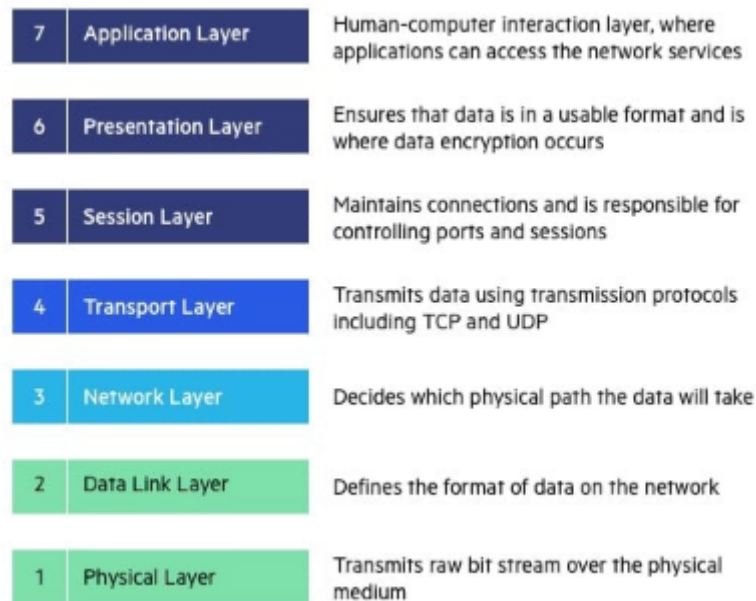


Figure 2.2: OSI model

Layer 3 filtering (Network Layer)

Layer 3 blocking, corresponding to the network layer in the OSI model, involves filtering out data packets based on their IP header information, which includes source and destination IP addresses. This is accomplished by configuring network devices like routers to employ access control lists (ACLs) that enforce rules on which traffic is allowed through. If a data packet's header matches a blocked IP address in the ACL, the packet is dropped, making the target website or service inaccessible. [21]. The process does not require heavy resources and is relatively simple to implement. Rules can be as specific as denying access to individual IP addresses or as broad as blocking a range of IPs. The advantage of this method is its straightforwardness and its capacity to be executed by any routing device in the network path. However, this method may cause problems with performance because routers may take a while to scan through long lists of ACLs. If this happens, the routing devices' processing capability may be exceeded, negatively impacting performance. Layer 3 filtering lacks the capability to isolate blocking to a particular service or port, meaning it cannot differentiate between multiple services that might be using the same IP address. As a result, when attempting to block one unwanted service, it may accidentally block all services associated with that IP address, leading to "over-blocking," where even legitimate services you did not intend to block are also blocked. [22]

Border Gateway Protocol (BGP) is one of the most common routing protocols that the Internet uses; it ensures the control of data flow across large networks. Malicious actors can reroute traffic when exploited for censorship by broadcasting false routing information. This

action can redirect data streams intended for specific IP addresses, effectively silencing targeted online resources or services by diverting or blocking their traffic. The architecture of BGP has the potential to have a wide-ranging effect, as it can affect not only individual networks but entire regions. This was starkly illustrated when Pakistan Telecom (AS17557), under government orders, attempted to block YouTube in 2008 by broadcasting unauthorised BGP announcements. The unintended outcome was a global disruption; all traffic meant for YouTube was misdirected to Pakistan Telecom, leading to widespread inaccessibility to the platform. As BGP routers worldwide accepted the incorrect routing information, a domino effect occurred, with significant portions of YouTube traffic being misrouted, severely affecting accessibility in several regions beyond the intended scope [23]. This incident shows how vulnerable internet infrastructure can be and how manipulating it can have far-reaching effects. Although BGP hijacking is a powerful censorship instrument, it must be closely monitored to avoid collateral damage that can go far beyond the intended targets. This emphasises the need for careful supervision when implementing network-layer filtering approaches.

Layer 4 filtering (Transport Layer)

Layer 4 of the OSI model, also known as the transport layer, controls the flow and exchange of data across the network. It also uses protocols such as TCP and UDP. In internet content blocking, Layer 4 filtering uses the IP addresses from Layer 3 and port numbers found in the transport layer headers of data packets to decide whether to allow or block traffic. Port numbers route packets to specific services running on a server. For instance, web traffic typically uses port 80 for HTTP and port 443 for HTTPS. Blocking at Layer 4, therefore, can be fine-tuned by the IP address and port number, allowing for service-specific restrictions. This means blocks can be executed more accurately, such as denying access to email services or specific web pages, without affecting all traffic to an IP address [22]. However, it blocks all or nothing on these ports and does not offer accurate control over which specific web pages get blocked. If you block a port, you block all the content accessed through that port, potentially including legitimate traffic you should allow. It's a challenge to block only the unwanted content without blocking what you want to keep accessible [21]. In simple terms, Layer 3 filtering is like blocking calls from a specific area code on your phone; you end up blocking everyone from that area. Layer 4 filtering is like blocking calls to and from a specific person's phone extension at a company. However, if that extension is used for multiple departments, you will block all calls to those departments, not just the one person you wanted to avoid.

Performance degradation can be used as a soft censorship or as a purposeful network management method that affects the efficiency of the transport layer by purposefully slowing down traffic to manage network resources. Prioritising some traffic over others or setting

data rate caps are two possible strategies to avoid network congestion, impairing non-prioritised services' operation. [24]

RST packet injection involves the unauthorised insertion of TCP reset packets into a communication stream, forcibly terminating connections. This can be used for censorship, disrupting unauthorised or unwanted network communications, and is considered a security threat when used maliciously. RST attacks are a well-known technique used by the Great Firewall of China [25] to interfere with and block connections to impose Internet censorship.

2.2.3 Multi-layer and Non-layer Specific Filtering

HTTP Keyword filtering: The technique of employing a list of banned keywords for Internet censorship involves a systematic approach where content containing these specified keywords is automatically blocked or flagged for review. This method is usually applied in URL and HTTP filtering. For example, keywords in the URL `www.protest.example` or "protest" in an HTTP header can trigger content blocks. This method is particularly effective in targeting and suppressing specific subjects and is heavily used by Saudi Arabia [26]. However, its effectiveness can be a double-edged sword; broad or commonly used keywords might result in excessive censorship, inadvertently blocking otherwise harmless or irrelevant content to the intended target [27]. This can lead to significant implications for freedom of expression and information accessibility, as seen in cases where generic terms related to political dissent or social movements are filtered, thereby limiting public discourse on these topics [28].

TLS (Transport Layer Security): TLS-based network censorship involves analysing unencrypted elements exchanged during TLS handshakes, such as the Server Name Indication (SNI). This analysis, known as TLS fingerprinting, allows censors to probabilistically identify a user's operating system, browser, or application by comparing specific combinations of TLS versions, cipher suites, and other options sent in the ClientHello message. This technique is beneficial for getting beyond the encryption barrier, allowing censors to identify the kind of services or information accessed even when the data flow itself is encrypted [29]. TLS-based network censorship, while capable of identifying user characteristics through elements like TLS fingerprinting, faces limitations. Its advantage lies in working out details such as the operating system or browser from unencrypted TLS handshake elements, providing insights despite the encryption.

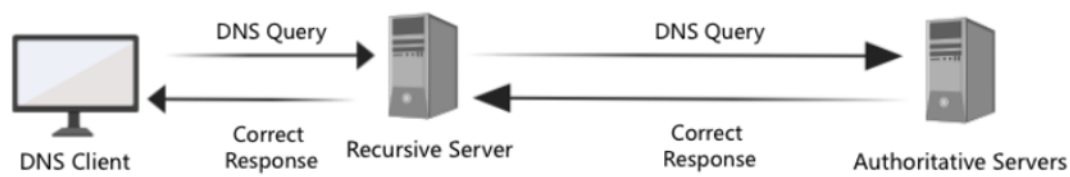
Deep Packet Inspection (DPI): DPI is an advanced censorship tool that goes beyond analysing basic packet headers and examining the application data section of network flows. Deep Packet Inspection (DPI) dives deep into internet traffic. By doing so, DPI can pinpoint

specific content within the traffic, such as particular words or phrases. This is more advanced than just identifying which type of service is being used or the destination of the data; it is about understanding the content of the communication itself. When DPI encounters encrypted traffic, like that protected by Transport Layer Security (TLS), its ability to read the content is significantly reduced. However, even with TLS, certain information remains unencrypted. The Server Name Indication (SNI) is a perfect example. SNI is an extension of TLS used to identify the server's hostname at the start of the handshake process [30]. It allows a server to present multiple certificates on the same IP address and port number and hence enables multiple secure (HTTPS) websites (or any other service over TLS) to be served off the same IP address without requiring all those services to use the same certificate. DPI tools can leverage this unencrypted information in the SNI field to determine which website a user is trying to access, even though the rest of their traffic is encrypted. For instance, if a user is accessing sensitive websites that a censor blocks, the user's request will still reveal the website's name via the SNI despite the encryption. The censor can then block the request based on this information. This can be particularly effective in regimes where censorship is heavily implemented, for example, in the Great Wall of China, as it allows the censor to block access to specific sites without decrypting the traffic.

DNS Manipulation

The Domain Name System (DNS) is a widely used, structured system designed to convert domain names, like 'example.com' into associated IP addresses, such as '192.0.2.1'. Although its original purpose was not for blocking access, it has become a popular method for this due to how easily and effectively DNS queries can be altered or redirected [22].

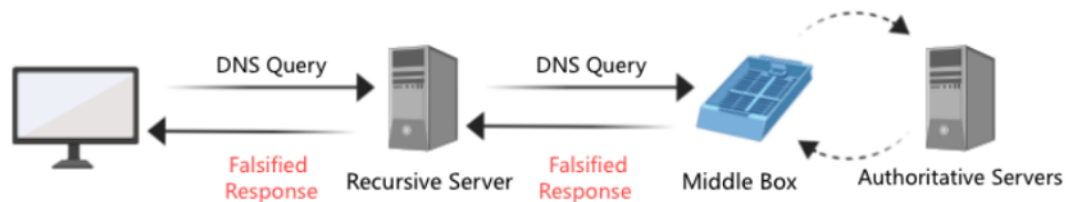
DNS tampering: DNS tampering is a more sophisticated censorship technique involving Domain Name System (DNS) manipulation. This method can disrupt the fundamental Internet infrastructure or hinder users' ability to locate specific websites. It involves rerouting or obfuscating requests, rendering websites inaccessible or misleadingly redirecting users with an incorrect address [31]. We will focus on introducing the two main techniques in DNS tampering, namely DNS hijacking and DNS injection (Figure 2.3)



(a) The fundamental process of a DNS resolution



(b) DNS Hijacking



(c) DNS Injection

Figure 2.3: The Fundamental Process of a DNS Resolution and Two Main Techniques in DNS Tampering

DNS Hijacking: In DNS Hijacking, a deviation from the standard DNS protocol stated above occurs when a DNS recursive server, rather than fetching resource records from authoritative servers or cached entries, is manipulated to return falsified responses [32]. This change is facilitated by gaining administrative control over the DNS server, allowing the response to a type A query, typically a legitimate Resource Record, to be replaced with a different, intentionally misleading result [1]. For example, a user attempting to visit a news site may be redirected to a site with government-approved content. This method effectively disrupts the expected behaviour of the DNS server, redirecting users away from authentic web resources.

DNS Injection: DNS Injection involves injecting false DNS information into the user's query responses, leading to incorrect IP address resolutions. It is a more subtle method of censorship, making it appear as if the website itself is faulty or inaccessible. This technique can occur at various network locations and requires a surveillance device along the network path. It works by replying to the client's DNS query faster than the legitimate resolver, with

the client accepting this first well-formed message and ignoring subsequent legitimate ones. The technique's impact can be broader than intended, affecting even those outside the censor's jurisdiction. This approach weakens simple circumvention methods against DNS Hijacking, as it can intercept and alter queries directed to any DNS resolver. This method is commonly used in China and Iran [33].

Choking the Band-width: also known as band-width throttling. It is a subtle form of internet censorship. the speed of internet access to specific websites or services is intentionally slowed down by the censor. This method only partially blocks content but makes access frustratingly slow, discouraging users from accessing certain information or services. Bandwidth throttling can target specific online platforms, such as video streaming services, social media, or news websites, especially those hosting content deemed undesirable by the censor. By reducing the quality of service, this technique can subtly manipulate online behaviour without the appearance of censorship. For instance, the Russian government initiated large-scale, targeted throttling of Twitter to pressure compliance with content removal requests, marking a significant shift towards centralised control of internet censorship [34].

Network Outages: Network outages as a method of censorship means deliberate interruptions of internet or telecommunications services by governments or authoritative bodies to control the flow of information and prevent the spread of communication. In extreme cases, the techniques employed for implementing network outages differ, including prefix hijacking attacks, DDos, and cutting off internet access entirely to a country or specific area. If the outages are decided over a state level, it is typically executed at the ISP level [35]. Egypt's internet blackout in 2011 during the Arab Spring is one of the most famous instances. During the anti-government demonstrations demanding the resignation of Egyptian President Hosni Mubarak, the Egyptian government ordered a complete shutdown of the Internet [36]. The longest internet blackout in a democratic nation occurred in Jammu and Kashmir, India, after the region's special autonomy status was revoked in 2019 [37]. Internet restriction in Myanmar reached previously unheard-of heights, including stringent website blocking, the suspension of cellular data in multiple networks, and total disconnection from the Internet after a military takeover in February 2021 [38].

2.2.4 Filtering by Example

The Great Firewall of China (GFW), officially known as the Golden Shield Project, was developed and implemented in 2008. The Great Firewall of China utilises a multi-layered Internet censorship and surveillance method. Key technologies employed include URL filtering, DNS tampering, IP blocking, and DPI. Maintaining the Chinese Communist Party's

(CCP) control over the nation's information flow is the central goal of the Great Firewall, which also serves to protect national security [39]. However, the Great Firewall also has weaknesses, and other techniques, like the usage of virtual private networks (VPNs), Tor browser, alternative DNS services, and proxy servers, have been created to get around its censorship measures [40]. Furthermore, because the system relies on pre-established lists of forbidden words and phrases, Chinese internet users develop new slang, euphemisms, and metaphors to circumvent censorship. The Great Firewall of China (GFW) implemented a new censorship method in November 2021 that passively identifies and blocks fully encrypted communications in real time. Many well-known censorship circumvention protocols, such as Shadowsocks, VMess, and Obfs4, are impacted by the GFW's new censorship capability [41].

2.3 Country Specific Practices

2.3.1 Ireland

In Ireland, the approach to Internet censorship is characterised by a Selective Filtering Approach, focusing on a more targeted strategy. This method primarily targets illegal content, such as child exploitation material, steering clear of the broad censorship of political or social topics. This is complemented by a strong Voluntary ISP Involvement, whereby the Irish Internet Service Providers are at the centre of content management [42]. These providers voluntarily comply with EU rules and national legislation, emphasising safeguarding user privacy. Moreover, the strategy involves Public-Private Collaborative Regulation, which sees the Irish government working with the technology industry. This collaboration is vital in applying censorship measures while ensuring privacy rights are upheld and respected. Ireland's model thus reflects a balanced approach, blending targeted content regulation with a commitment to preserving individual privacy rights.

In response to Russia's invasion of Ukraine, according to the journal, the EU has banned the state-controlled Russian news networks RT (previously Russia Today), Russia.tv, and Sputnik [43]. This prohibition includes the suspension of satellite signals, internet access, and app availability across EU territory and the cancellation of broadcasting licences. The decision aims to curb the spread of disinformation propagated by these channels. Following the EU ruling, RT broadcasting in Ireland ended, with major service providers like Eir and Virgin Media Television removing the station from their network. Sky Ireland has previously removed the channel from its platform; therefore, no more action was required. The efforts to limit the effect of RT and Sputnik are part of a larger initiative to combat disinformation and maintain the integrity of news broadcast within the EU and its partner nations. It is not only Ireland taking part in this mass censoring of Russian media but also global tech giants, including YouTube, TikTok, Meta (parent company of Facebook and Instagram), and

Google, have taken actions to restrict RT and Sputnik's reach on their platforms in Europe. Google and Apple have removed their mobile apps from stores. Twitter has initiated warnings on tweets linking to Russian state-affiliated media, aiming to reduce the circulation of their content.

When attempting to access Russia.tv from an Eir network, I was presented with a message stating access to this website has been blocked 2.4. When I attempted to access RT.com, the page reloaded without fully loading, meaning that the TCP handshake, the initial step to establishing a network connection, was being disrupted. Recent measurements conducted by OONI In the past month confirm that these websites are being censored in Ireland. Russia.tv (2.5) appears to be censored through DNS-based filtering, which means that when the computer tries to discover the IP address associated with Russia.tv domain, the ISP is simply routing my request to a block page, effectively preventing access. We used the "dig" command-line tool to test this. Access to Russia.tv is blocked when the domain is resolved through the ISP's DNS server, which produces an IP address linked to a local block page. However, queries made with publicly available DNS resolvers, such as 1.1.1.1 and 8.8.8.8, get a reliable IP address that seems to be the official one for Russia.tv. This consistency in public resolver results and the discrepancy when using the ISP's DNS point to the possibility that the ISP is blocking Russia.tv through DNS-based censorship. On the other hand, RT.com (2.6) is being blocked through TCP/IP interference, which is a deeper level of interference that involves the blocking or resetting TCP connections to RT.com's servers. This can be done by the ISP dropping the packets so they never reach rt.com's servers or by the servers themselves if they have been instructed to reject connections from certain IP ranges. The result is that my browser will keep trying to load the page, attempting the handshake repeatedly, but it will not succeed, leading to an endlessly reloading page. Essentially, any attempt to communicate with the server is cut off, rendering the website inaccessible. Using Censored Planet data reinforces the findings that Ireland blocks certain Russian media outlets, with the example of Sputnik News showing a 100% unexpected rate on a specific network AS546.



ACCESS TO THIS IP ADDRESS RELATING TO THIS WEBSITE HAS BEEN BLOCKED

If you have questions or concerns that you would like to discuss please contact:
eir Customer Care on 1901.
eir broadband support on 1890 260 260.

Customers of other operators

If you are a customer with another operator please contact your operator directly regarding this message.

eir and open eir are trading names of eircom Limited, Registered as a Branch in Ireland Number 907674, Incorporated in Jersey Number 116389. 24, Ireland.
2015 eir. All rights reserved.

Figure 2.4: Block page seen When accessing Russia.tv

IE		AS 41230	2024-03-24 22:04 UTC	Web Connectivity Test	http://russia.tv/	dns
IE		AS 41230	2024-03-24 19:22 UTC	Web Connectivity Test	http://russia.tv/	dns
IE		AS 41230	2024-03-24 16:55 UTC	Web Connectivity Test	http://russia.tv/	dns
IE		AS 41230	2024-03-24 15:40 UTC	Web Connectivity Test	http://russia.tv/	dns

Figure 2.5: Ireland Measurements on the OONI database taken by other people who have used OONIprobe to test Russia.TV

IE		AS 5466	2024-03-24 22:52 UTC	Web Connectivity Test	https://www.rt.com/	tcp_ip
IE		AS 5466	2024-03-24 22:02 UTC	Web Connectivity Test	https://www.rt.com/	tcp_ip
IE		AS 5466	2024-03-24 02:21 UTC	Web Connectivity Test	https://www.rt.com/	tcp_ip

Figure 2.6: Ireland Measurements on the OONI database taken by other people who have used OONIprobe to test RT.com

2.3.2 Saudi Arabia

The Internet censorship landscape in Saudi Arabia is marked by Extensive and Intrusive Censorship Methods, reflecting the nation's commitment to aligning digital spaces with its

cultural and religious norms. The heart of this approach is the Extensive Content Blockade Systems, which block a wide array of content categories, effectively shaping the online narrative in line with state policies [18]. Complementing this broad blockade is the use of Sophisticated Surveillance Techniques. The state leverages advanced technologies for monitoring Internet usage and enforcing its censorship rules, including tracking digital communications and social media interactions.

In a 2004 study done by the OpenNet Initiative, it was observed that Saudi Arabia's approach to internet filtering primarily targeted content related to Adult content, drugs, gambling, religious conversion, and tools for anonymising internet activities. Out of a total of 741 websites tested, 18% were blocked. The most extensively blocked content categories were Adult (98% of sites blocked), gambling (93%), and drugs (86%). The study found the filtering system to be effectively narrow, focusing on these specific categories and successfully blocking many sites within them. However, there was an incidence of overblocking, where some websites that did not fall under these categories were also blocked. Despite these measures, sites outside the stated categories remained broadly accessible, indicating a targeted and somewhat effective filtering regime [44].

The study by Alharbi et al. reveals that Saudi Arabia employs HTTP and TLS filtering techniques to regulate internet content. HTTP filtering involves analysing unencrypted URL strings to identify and block access to undesirable websites, particularly content related to adult material, shopping, and games. This form of filtering operates by inspecting the clear text of HTTP GET requests for specific keywords that might be associated with banned content. The research indicates a reduction in the percentage of sites blocked through HTTP filtering from 2018 to 2019, with adult content sites being filtered at a rate of 82.2%, shopping at 7.6%, and games at 6.2% as of 2019. These percentages suggest a focus on censoring primarily adult content over other categories. TLS filtering, on the other hand, targets encrypted HTTPS traffic. Since HTTPS conceals the specifics of the GET request within encrypted traffic, keyword-based filtering is ineffective. The shift towards HTTPS across the internet means that traditional HTTP keyword filtering is becoming less effective, necessitating different approaches for content filtering, like DNS and IP blocking [26]. The study done by Master [2] observed that during its analysis period, Saudi Arabia employed HTTP keyword filtering alongside TLS-based filtering as primary methods of internet censorship. Historically, the country also utilised IP and port blocking and DNS tampering. However, the move away from these outdated techniques is due to their tendency for over-blocking, which unintentionally makes it harder to access allowed content. This observation that Saudi Arabia uses mainly HTTP-based filtering aligns with Alharbi's [26] findings, and it is anticipated that our results (excluding Adult Content) will similarly reflect and reinforce the patterns of internet censorship outlined in the existing literature.

2.3.3 Effects of Censorship

Underblocking Underblocking refers to the failure of filtering systems to adequately restrict access to all content required for censorship. It occurs when content that should be blocked evades detection by censorship mechanisms [45]. The reasons for underblocking commonly involve inadequate keyword filtering, where reliance on specific keywords fails to identify metaphors, alternative spellings, or visual content without accompanying text. Additionally, websites with dynamic content changes can outpace static filters [46]. Underblocking exposes audiences, especially vulnerable groups, to potentially harmful or unsuitable content [47].

Chilling effect The chilling effect of censorship captures the phenomenon where the threat or fear of laws, regulations and the more general legal system inhibits the free exercise of rights, particularly those related to expression and press. This effect often arises from laws that are either too vague or too wide, creating an atmosphere of uncertainty that leads individuals and entities to self-censor [48]. Such an atmosphere can be exacerbated by government surveillance, which may push individuals to limit their speech to avoid potential legal risks [49]. The chilling effect is particularly evident when journalists conduct self-censor to avoid legal challenges or strong opposition from government agencies. Since the mid-2010s, South Korea has seen widespread anti-press sentiment fueled by emotions like disgust and hatred. This has led to a chilling effect in news organizations [50], potentially stifling public communication and citizen engagement in democracy.

Echo Chamber Effect The echo chamber effect is a phenomenon in digital communication where individuals are exposed mainly to opinions and information that reinforce their beliefs. Algorithm-driven content systems on social media platforms, such as Facebook, Reddit, and Twitter, amplify this effect. This personalized content creates virtual echo chambers, isolating users from contradictory viewpoints and contributing to polarization within discourse [51]. Notable consequences of the echo chamber effect include the spread of misinformation, as unchecked information can spread in closed communities [52]. It also causes societal divisions, as individuals become more ideologically extreme and less tolerant of opposing views [53].

2.4 Circumvention Tools

Circumvention tools are software and technologies designed to bypass internet censorship and surveillance. These tools enable users to access content and communicate online even when various entities impose restrictions, such as governments, corporations, or local internet service providers. The need for such tools arises in environments where information

is hindered, whether for political, cultural, or commercial reasons. While these tools are instrumental in promoting digital rights and ensuring access to information, they also come with legality, ethics, and security complexities. Their development and usage are subjects of ongoing global discourse, reflecting the challenges of balancing open access with regulatory and societal concerns [54].

2.4.1 VPN

VPNs are like secret tunnels for your internet data. Using a VPN hides what you're doing online by sending your data through a secure path to a server somewhere else. That server then sends your data to the internet, so it looks like the data is coming from there, not from you. This keeps your information safe from people trying to snoop by encrypting the data and helps you get around blocks that stop you from seeing certain websites. VPNs use protocols like OpenVPN and WireGuard to define how securely the data is encrypted. In the paper titled "GoHop: Personal VPN to Defend from Censorship" by Yuzhi Wang et al., the authors introduce GoHop [55], a sophisticated VPN designed to counteract internet censorship. The paper underscores the importance of online privacy and how recent advancements like high-speed Deep Packet Inspection (DPI) and statistical traffic analysis can compromise it. Traditional encryption methods discussed above are often insufficient, as censorship systems have evolved to detect and block encrypted traffic. To address these challenges, GoHop employs several methods, including protocol obfuscation and traffic morphing, which enhance the strength and stealth of its encryption. The authors demonstrate how GoHop effectively prevents the detection and blocking of its traffic by censorship mechanisms, ensuring users can maintain high-bandwidth network throughput for their online activities [55]. This tool represents a significant advancement in the fight against censorship, offering users more reliable protection against the suppression of information.

2.4.2 TOR, the onion router

Tor (The Onion Router) is a free software that enables anonymous communication by directing internet traffic through a worldwide, volunteer-operated network consisting of thousands of relays. Each relay encrypts the data and peels off a layer of encryption (2.7), similar to layers of an onion, hence the name. it does this by:

- **Layered Encryption:** Tor uses layered encryption, where each relay (or node) in the network peels away a single layer of encryption, revealing the next relay's address but not the final destination.
- **Relay Network:** Tor traffic passes through a series of relays, each run by volunteers worldwide. The traffic is randomly bounced through a number of these relays before

reaching its final destination. The relays are categorized as entry (guard) nodes, middle (relay) nodes, and exit nodes, each serving a specific function in the anonymization process.

- **Entry and Exit Nodes:** The entry node sees the user's IP address but not the final destination of the data, while the exit node sees the data's final destination but not the user's IP address. This separation ensures that no single relay knows both the origin and the destination of the data.
- **Circuit Construction:** Tor establishes a 'circuit' of encrypted connections through a series of relays, and this circuit changes periodically, making tracking even more challenging. [56].

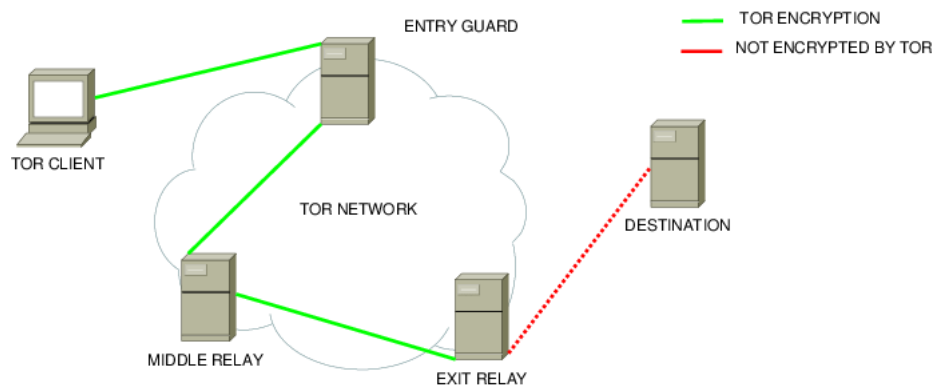


Figure 2.7: A basic architecture of the TOR network

Tor is firmly opposed to any form of information restriction and consequently does not govern user activities; this lack of moderation has led to Tor being utilized for a range of unlawful activities.

2.4.3 Proxy Servers

A proxy server functions as an intermediary relay between a user's device and the broader internet. When a user requests access to a specific online resource, the request is first routed through the proxy. This server then forwards the request to the target website on behalf of the user. Upon retrieving the website's data, the proxy server routes the information back to the user. This process allows the user to bypass regional content restrictions or access blocked websites [57]. While proxy servers can circumvent location-based barriers, they differ from VPNs in that they typically do not encrypt the user's traffic. Consequently, proxies provide a level of anonymity but may not offer comprehensive privacy protection for all user activities online.

2.4.4 Psiphon

Psiphon is a tool designed to help people access the internet freely, even when certain websites are blocked [58]. It is like a special pass that finds open routes on the internet, choosing the ones that work best and are most challenging to block. Psiphon is smart; it tests different paths and picks the ones that deliver information quickly and reliably. This is especially useful when a country's government tries to limit what people can see or do online. For example, during the conflict in Ukraine, Psiphon became a go-to tool for many in Russia when the government restricted access to sites like Twitter and Facebook [59]. It helped over a million people stay connected and informed, proving that there are still ways to reach the wider world online even when barriers go up.

2.4.5 Challenges & Limitations

Circumvention tools like VPNs, proxies, and the Tor network are designed to bypass internet censorship, allowing users to access blocked content and maintain their privacy online. Their effectiveness can vary depending on the strength of the censorship they are facing. These tools can be a lifeline for accessing unrestricted information and communicating freely in places with heavy internet restrictions. However, these tools have their limitations. For instance, they can slow down your internet speed because your data takes a longer and more complex route. There can also be legal issues, as some countries have laws against using such tools, leading to potential consequences for users. Ethically, the use of circumvention tools raises questions. While they can support the freedom of expression and protect privacy, they might also enable access to harmful or illegal content. In regions with oppressive regimes, they can be crucial for activists and journalists, but in more open societies, their usage may need to be weighed against other considerations, like respecting local laws and regulations.

2.5 Challenges and Concluding thoughts

In conducting research for the state-of-the-art section, several challenges were encountered, particularly in sourcing scholarly articles that conduct comparative analyses of internet censorship across different countries. The lack of studies comparing internet censorship practices in different contexts, especially in places like Ireland, where censorship is minimal, points to a significant gap in current research. Ireland's minimal use of internet censorship mechanisms makes it an interesting case study that needs to be better represented in existing studies. Identifying and selecting relevant information about censorship methods and circumvention tools was challenging due to the wide variety of technologies available. This research aims to simplify this information, focusing on the most effective tools and

strategies to offer a thorough understanding of how digital control is practised. In order to produce an in-depth and perceptive analysis within the broad field of internet censorship research, a complete understanding of the implementation of digital control was essential. Determining the precise techniques used in censoring practices in nations like Saudi Arabia and Ireland requires an understanding of these mechanisms and technologies.

3 Methodology

3.1 Introduction

This thesis section will explain the research methodology used to investigate internet censorship in Saudi Arabia and Ireland. The approach involved using two techniques: digital proxy through the Virtual Private Server (VPS) and direct observation with the OONIprobe program. The data collected through these techniques was then processed and categorized using Python scripts before being converted into CSV files.

3.1.1 Description of OONI

OONI's Web Connectivity test is a thorough method for probing internet censorship. It runs a set of operations trying to reach a site to understand if and when a site is reachable and look at the blocking put in place. The test initiates by performing a DNS lookup to translate the website URL into an IP address. Subsequently, it attempts to establish a TCP connection with the server hosting the website. Following a successful TCP handshake, a HTTP GET request is sent to fetch the full webpage content. The test then compares these results with those from a control server typically unaffected by local censorship mechanisms to check for discrepancies. If the results do not match, there is some interference. DNS blocking is inferred if the IP address returned by the DNS query does not match that of the control. TCP/IP blocking is suspected when a TCP connection fails. HTTP blocking is indicated in case there is a failure of the HTTP request or there are noticeable variations in body lengths, HTTP headers, or HTML title tags, between the content control server and that from the tested network. Through this process, the Web Connectivity test can detect the presence of network interference, offering a technical map of censorship practices.

The TLS handshake is an essential part of how tools like OONI measure internet censorship, including the blocking of websites. When OONI's software tries to connect to a website, it initiates a TLS handshake to establish a secure connection. If a website is blocked, the TLS handshake might fail, but not all TLS handshake failures are due to censorship. The reasons for failure could be network issues, server problems, or intentional blocking. OONI's methodology involves running two TLS handshakes as part of its experiments: one with a

control SNI (Server Name Indication) that should not be blocked and one with an SNI that could be blocked. If the control handshake succeeds but the second fails, it can signal network interference or blocking. For example, in the case of SNI-based blocking, the handshake may fail if a network rule is implemented to block any TLS handshake where the SNI extension contains a specific domain known to be censored. However, if the first control handshake fails, the test is inconclusive. Moreover, TLS handshake failures in OONI's tests, such as timeouts or connection resets, can be seen differently. A timeout could mean the server did not respond in time, which might suggest blocking or a network problem. A connection reset, where the handshake is interrupted, is a stronger indication of interference, often pointing to the network actively cutting off the connection attempt. These measurements are critical because they help identify the techniques the censor uses to block or censor content, ranging from DNS tampering to SNI-based filtering or TCP/IP interference.

Circumvention Test: OONI Probe was also utilised in the circumvention test carried out as part of this study to evaluate the effectiveness and accessibility of several circumvention tools across the Saudi Arabian and Irish internet infrastructures. The test specifically targeted the reachability of the Tor network, including directory authorities that guide traffic across the network, OR ports used by Tor bridges that facilitate connection establishment, and obfs4 bridges, which employ the obfuscated obfs4 protocol to conceal Tor traffic from network monitoring tools aiming to block it. The operational steps of the circumvention test were as follows:

- **HTTP GET request:** An HTTP GET request was initiated to the Tor network's directory authorities via The URL path `/tor/status-vote/current/consensus.z` is a resource essential for the functioning of Tor relays.
- **TLS Handshake:** Attempts were made to connect to both OR ports and directory authority ports, followed by a TLS handshake to establish a secure connection channel.
- **Obfs4 Handshake:** A specialised handshake was performed with obfs4 addresses to confirm the functionality of this obfuscation protocol within the network.

The Psiphon test implemented via the OONI Probe is designed to check the functionality of Psiphon within a given network. The test sequence is initiated by using the Psiphon protocol to establish a tunnel. It then verifies the tunnel's operational status by attempting to retrieve a webpage. The outcome of this test is pivotal. It determines not only whether the Psiphon application can be bootstrapped but also if the established tunnel is capable of facilitating web access. The testing procedure has three potential outcomes: (1) successful bootstrap and webpage retrieval indicating that Psiphon is operable and can circumvent censorship, (2) successful bootstrap without webpage access suggesting the tunnel is active

but cannot reach the web resources, likely due to selective blocking within the network, and (3) failure to establish a Psiphon tunnel, pointing towards a network where Psiphon is blocked outright.

3.2 Real-Time Data Collection

OONI relies on a collaborative approach, gathering measurements from a global community of volunteers. The majority of the data OONI analyses is contributed by everyday users who run tests from their devices, allowing for a broad collection of internet censorship incidents. Our testing on March 1, 2024, at 14:22 UTC of the website <http://www.scarleteen.com/> confirms the functionality of real-time data capture (3.2) and that Ooni was uploading the results we collected on their database. The 'uploaded' flag within the measurement log was active, signifying that the data was collected in real time during our assessment (3.1). To differentiate our dataset from the one OONI publicly shares and to facilitate an independent analysis, we turned off the feature that automatically uploads collected data to OONI databases. This adjustment was essential to our methodology, allowing us to independently accumulate and curate our dataset, guaranteeing that our results would be distinct from the results that OONI openly disclosed.

```
#1933
http://www.scarleteen.com/ (XED)
web_connectivity          ok: ✗
success: ✓                uploaded: ✓
{"accessible": false,
 "blocking": "http-diff"}
```

Figure 3.1: Measurement taken on our VPS based in Riyadh



Figure 3.2: same measurement on the Ooni explorer showcasing real-time update of our data

3.2.1 Data Collection in Ireland

The Windows Subsystem for Linux (WSL) was used to deploy the OONIprobe programme, which provided a suitable environment for its command-line interface and facilitated the data-gathering procedure in Ireland. The inability of OONIprobe, to support Windows natively, made this technique necessary. OONIprobe was configured after WSL installation to start the web connectivity tests.

To conduct testing, an extensive list of URLs representing potentially censored information in Ireland was tested. This list, which OONI carefully selected in cooperation with community members, non-governmental organisations, and other stakeholders, included about 1300 websites. Additionally, to gain a global perspective, we curated our own list of URLs, including the top 100 global websites as well as the top 100 websites in Saudi Arabia in the testing to draw comparative insights. SimilarWeb [60] was used to get the list of most visited websites in each country and the top 100 Worldwide lists. The assessment was further expanded to incorporate our own 100 randomly selected URLs across key categories, including news media, social & entertainment media, political criticism & government sites, LGBTQ content, and gambling platforms. For the random selection of websites, 100 were identified from each category under consideration. Utilising a random wheel generator, all 100 websites from each category were entered into the system. With each spin of the wheel, 20 websites were randomly selected from the pool of 100, ensuring a fair and unbiased selection process. This diverse set aimed to paint a detailed picture of internet accessibility and highlight content-specific blocking practices. Finally, the OONI-curated list for Saudi Arabia, consisting of around 2300 websites, was also tested to establish a baseline for censorship in the region. The results of these tests, which recorded the sites' accessibility and blocking, were methodically gathered and ready for further analysis. Lastly, As the introduction states, these lists do not include adult content websites.

3.2.2 Data Collection in Saudi

To emulate internet access within Saudi Arabia, we bought an Ubuntu-based Virtual Private Server (VPS) hosted in Riyadh. This provides a platform to conduct tests with the same network conditions as users in Saudi Arabia would experience.

Secure Shell (SSH) was used to gain remote access to the VPS. SSH provides a secure channel for remote computer access and command execution. It was used for its robust security features, like encrypted communication and authentication. The process began by generating a pair of cryptographic keys: a private key on the local machine and a public key placed on the VPS. This setup allowed for passwordless authentication, which is more secure than traditional passwords. With the public key on the VPS, only someone with the corresponding private key could gain access, significantly reducing the risk of unauthorised entry. Once the keys were in place, remote access was as simple as launching an SSH client and connecting to the VPS using its IP address and the specified user account. The convenience of SSH came from its ability to provide a secure command-line interface to the remote server, enabling the installation and operation of OONIProbe from the local machine as if one were physically present at the server's location. This remote capability was crucial for conducting internet censorship research, as it allowed for real-time, secure management of the VPS without the need for physical intervention.

Following the same strategy used in Ireland. The first test was the OONI-curated list, which included about 2300 websites specifically chosen for Saudi Arabia by the OONI organisation. These tests aimed to detect any web filtering specific to the country. To create a comparative framework, we also ran connectivity tests for the top 100 global websites, alongside the top 100 Irish websites, to understand the difference in access between global, Irish, and Saudi domains. Additionally, the same 100 random URLs from the different categories, like those tested in Ireland, were tested to uncover any targeted censorship within Saudi Arabia. Finally, to round off the cross-regional analysis, we tested the same 1300 websites from the OONI-curated Irish list. This step was crucial in contrasting the internet censorship landscape between the two countries. The VPS systematically documented All test outcomes in JSON format, ensuring a structured and analysable dataset. The JSON files recorded details of the testing process, such as the success or failure of connections, along with any anomalies that suggest filtering.

Saudi Arabia	Ireland
OONI curated list Saudi	OONI curated list Ireland
Top 100 Ireland websites tested in Saudi	Top 100 Saudi websites tested in Ireland
Test OONI-curated list Ireland in Saudi	Test OONI-curated list Saudi in Ireland
Top 100 worldwide websites in Saudi	Top 100 worldwide websites in Ireland
Random 100 URLs in Saudi	Random 100 URLs in Ireland

Figure 3.3: Tests ran in both countries.

3.2.3 Process Overview

Once Ooni was installed, the “ooniprobe run websites” command was run .This tests the Ooni curated list for that country. Once the measurements are finished. A directory is created inside the measurements folder, which houses JSON files of each measurement. In order to run our own curated list, we first had to create a file with the list of URLs we want to test in the server, and ooni has a command for that which is “ooniprobe run --input-file “name of the file”.

The data extraction process from the JSON files generated by OONIProbe was automated using a Python script. These scripts are parsed through each JSON file in the specified directory, extract crucial details, and record them in a structured format. Specifically, the script opened each file using the JSON library to load the data, emphasising the UTF-8 encoding to ensure the correct handling of international characters.

The core of the JSON scraper involved a function designed to process individual JSON files, extracting relevant fields such as the URL, the accessibility of the website and the method of blocking if applicable. This function determines the accessibility status by checking for the presence of test keys within the JSON structure, which directly indicates whether a site was blocked and by what method.

```

# CSV headers
headers = ['URL', 'Blocked', 'Blocking Method']

# Function to process each JSON file and extract the data
def process_json_file(file_path):
    with open(file_path, 'r', encoding='utf-8') as file:
        data = json.load(file)
        url = data.get('input', '')
        test_keys = data.get('test_keys', {})
        accessible = test_keys.get('accessible', '')
        blocking = test_keys.get('blocking', '')

        # Convert boolean to string for CSV output
        blocked = 'Yes' if not accessible and blocking else 'No'
        blocking_method = blocking if blocking else ''

        return [url, blocked, blocking_method]

```

Figure 3.4: JSON scraper function

Upon extraction, the data was converted into a CSV format for easy analysis. The script created a CSV file with headers reflecting the key data points: 'URL', 'Blocked', and 'Blocking Method'. It utilised the CSV writer from Python's built-in CSV module (3.5), ensuring the data was written in line with standard data analysis tools' expectations. The script iterated over each JSON file within the designated directory, invoking the extraction function on each and writing the results row by row into the CSV file. The try-except blocks within the script ensured that any errors encountered during processing did not halt the execution; instead, they were printed to the console for debugging purposes. Upon completion, a message confirmed the successful creation of the output CSV file. The script effectively transformed the raw OONI data into a rich dataset ready for analysis, which in turn enhanced the efficiency of data handling but also ensured consistency and accuracy in the dataset.

```

# Write to CSV
with open(csv_file_path, 'w', newline='', encoding='utf-8') as csvfile:
    writer = csv.writer(csvfile)
    writer.writerow(headers)

# Walk through the directory and process each JSON file
for file_name in os.listdir(json_directory_path):
    if file_name.endswith('.json'):
        try:
            row = process_json_file(json_directory_path / file_name)
            writer.writerow(row)
        except Exception as e:
            print(f"Error processing {file_name}: {e}")

print(f"Data extraction complete. CSV file created at {csv_file_path}")

```

Figure 3.5: CSV code

Categorisation of URLs

The JSON files generated by the OONIprobe tests did not have category information, making it challenging to analyse and categorise the URLs. To tackle this issue, a second Python script was created that utilised the OONI API to include the missing categorisation data in the dataset. The script used a function called 'extract domain(url, domain data)', which parsed each URL and extracted its domain name. This normalisation process was necessary to accurately match each URL with its corresponding category using the OONI API.

```
# Function to extract domain from URL
def extract_domain(url, domain_data):
    parsed_uri = urlparse(url)
    domain = '{uri.netloc}'.format(uri=parsed_uri).replace('www.', '')
    if 'www.' + domain in domain_data:
        return 'www.' + domain
    return domain
```

Figure 3.6: Code normalising the domain names for the API to recognise

Another function, “get all domains(api URL)” (3.7), was implemented to interact with the OONI API endpoint “/API/ /domains”, which lists all domains in the test-lists along with their measurement count. By querying this endpoint, the script could retrieve a comprehensive list of categorised domains. The script performed a loop to request data from the OONI API until all pages of results were fetched and processed. If the response status was successful (HTTP 200), the script parsed the JSON response and updated a dictionary of “all domains” with the domain name and corresponding category code. Once the list of categorised domains was compiled, the script read through the list of domains we tested, which were stored in a text file 'URLlist.txt'. The domains were then matched with their categories from the “all domains” dictionary. If a domain was not found in the dictionary, it was assigned a default category of 'Unknown'. The script concluded by writing the normalised domain names and their associated categories to a CSV file named 'categorised domains'.

```

def get_all_domains(api_url):
    all_domains = {}
    page_url = api_url # Start with the initial API URL

    while page_url:
        response = requests.get(page_url)
        if response.status_code == 200:
            data = response.json()
            results = data.get('results', [])
            # Update all_domains with new data
            all_domains.update({item['domain_name']: item['category_code'] for item in results})
            page_url = data.get('next')
        else:
            print(f'Failed to fetch page: {page_url}')
            break

    return all_domains

# Start the pagination process
categorised_domains = get_all_domains('https://api.ooni.io/api/_/domains')

```

Figure 3.7: Function Interacting with the OONI API

3.3 Challenges and Solutions

During the process of collecting data, I faced several challenges, each requiring a different solution to ensure the project's success.

Challenge 1: Data Volume and API Rate Limiting The vast volume of data from the OONIProbe results presented a challenge, mainly due to rate limiting on the OONI API, which restricted the frequency and volume of data that could be fetched in a given time.

Solution: Implementing pagination in the Python script solved this challenge. It allowed for sequential data retrieval, ensuring the script collected data in manageable chunks while adhering to the API's rate limits.

Challenge 2: Inconsistent Data Formats JSON files lacked uniformity in structure, which complicated the data parsing process. Moreover, some files had missing fields or were formatted differently, posing a challenge for automated extraction. **Solution:** The Python script was refined to include error handling and data validation. Functions were designed to check the presence and type of data fields before extraction and to handle exceptions, ensuring that one malformed file did not disrupt the entire dataset's processing.

Challenge 3: Compatibility and Environment Configuration The testing for Ireland faced a compatibility challenge as OONIProbe's command line interface did not support the Windows operating system. **Solution:** The compatibility issue was resolved by using the Windows Subsystem for Linux (WSL). This allowed for a Linux-compatible environment to be set up on a Windows machine, enabling the use of OONIProbe's command-line interface. The WSL served as a bridge, providing the Unix-like environment to run the tests effectively without needing a separate Linux machine or dual-boot setup.

4 Results and Evaluation

4.1 OONI Curated Lists

4.1.1 Saudi Arabia

Based on the data collected from Saudi Arabia using the VPS, we can perform a detailed analysis of the Internet censorship landscape in the country. The dataset in the table (4.1) categorises different types of content and indicates the percentage of sites that have been blocked within each category. The highest rates of censorship are seen in categories such as Gambling, Anonymization & Circumvention Tools, and Alcohol & Drugs, suggesting a focus on restricting content that is against social norms and legal constraints.

Category	Total Sites	Blocked (%)
Gambling	64	54.69
Anonymization and circumvention tools	141	41.84
Alcohol and Drugs	46	34.78
Terrorism and Militants	7	28.57
Political Criticism	69	26.09
LGBT	108	25.00
News Media	301	21.93
Provocative Attire	40	20.00
Media Sharing	78	17.95
Religion	115	17.39
Online Dating	55	12.73
Sex Education	52	11.54
Human rights issues	187	10.16
Gaming	38	7.89
File Sharing	94	6.38
Search Engines	51	5.88
Communication tools	138	5.07
Hosting and Blogging Platforms	149	4.03
Government	44	2.27
Social networking	89	2.25
Culture	96	2.08
Hacking Tools	61	1.64
Economics	64	1.56
Public Health	71	1.41
E-commerce	31	0.00
Hate Speech	7	0.00
Control content	45	0.00
Environment	57	0.00
Intergovernmental Organisations	14	0.00

Table 4.1: OONI Curated list for Saudi Arabia Tested in Saudi Arabia

The graph 4.1 detailing the count of blocking methods used reveals significant reliance on HTTP failures and HTTP diffs, indicating two primary censorship tactics. These refer to different outcomes of HTTP-based tests that are designed to detect web censorship:

HTTP Failure: A HTTP failure occurs when an HTTP request to access a website is not completed successfully. This could mean that the website did not return a response at all or that the connection was interrupted. It could indicate that the website is down, or more relevant in the context of censorship, that access to the website has been blocked or interfered with in a way that prevents the completion of the HTTP request. A HTTP failure as a result of censorship might be due to:

IP blocking: The censor blocks the IP address of the website, so requests to the site do not reach it at all.

DNS tampering: The DNS response is manipulated so that the domain name does not resolve to the correct IP address, leading to a failure in the HTTP request.

TCP/IP blocking: The TCP connection is not established due to the ISP filtering or network interference that disrupts the TCP handshake.

HTTP Diff: A HTTP differential test (HTTP diff) involves comparing the content of a website accessed from different networks, typically one where censorship is suspected and another known to be free of censorship. If the website's content is different when accessed from the tested network compared to the uncensored network, this is termed "HTTP diff". A HTTP diff could indicate the following types of censorship:

Content tampering: The website's content is altered by the censor, which could include the injection of additional content, removing or altering existing content, or redirecting to a different page.

Keyword filtering: Specific keywords trigger censorship mechanisms, and pages containing those keywords may be altered or blocked.

Selective content manipulation: Only some aspects of a website is blocked or altered, such as specific images, scripts, or style sheets. These indicators can help understand the methods the censor uses to control access to information online. Identifying the precise blocking technique can sometimes be complex because different methods can produce similar outcomes, and censors often use a combination of techniques to enforce content restrictions.

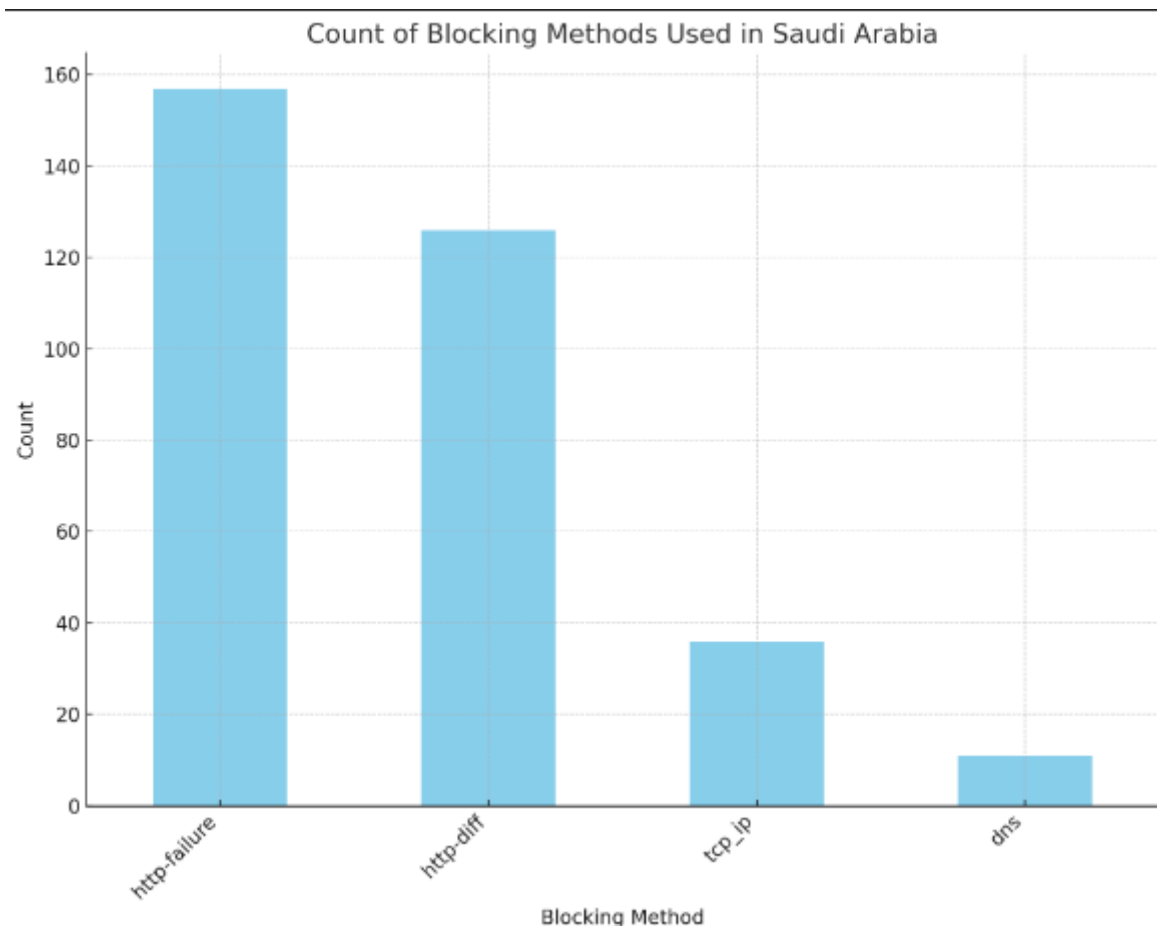


Figure 4.1: Censorship Mechanisms dominant in SA

4.1.2 Ireland

Ireland's internet filtering statistics paints a different picture than Saudi Arabia's. Data gathered from Ireland's Ooni Curated List indicates that the country has comparatively low censoring rates in several areas. (Table. 4.2). With a 17.33% blocking rate, File Sharing is the most restricted category, reflecting efforts to prevent unauthorised downloads and safeguard copyright. There is no indication of filtering in areas like social networking, provocative attire, LGBT content, and several others, in contrast to Saudi Arabia, indicating a more free internet environment.

Looking at the graph 4.4 on the blocking methods used in Ireland, DNS blocking appears to be the most used technique. DNS blocking intercepts the resolution of domain names to IP addresses, preventing users from reaching certain websites. The count for TCP/IP blocking is also significant, though to a lesser extent than DNS blocking. As previously said, TCP/IP blocking hinders the connection process, which may be a sign of censorship efforts or, less likely, network management procedures. The HTTP differential tests (HTTP diff) and HTTP failures are much less prevalent in Ireland, indicating fewer instances where the content is manipulated or a HTTP request fails altogether. These instances of HTTP diff

and HTTP failure are minimal, which could imply that when websites are inaccessible, it is more likely due to technical reasons rather than deliberate blocking. The key takeaway from the Irish dataset is the limited use of censorship, with most content categories showing 0% blocking. This indicates a strong commitment to maintaining internet freedom, contrasting sharply with the practices observed in Saudi Arabia. The minimal use of HTTP diffs and HTTP failures further suggest that when censorship does occur, it is less about content manipulation and more about access restriction, potentially for legal and regulatory compliance. The OONI dataset presents an interesting picture of internet filtering in Ireland, which is characterised by a varied practice among different Autonomous Systems (AS). For example, the AS5466, which operates for Eir (the ISP used for our Ireland testing), does not disrupt access to vesti.ru (4.2). This may be a leading indicator that they have an open policy. This contrasts AS6830, which blocks the same site and demonstrates a tendency to follow content management strategies selectively. Interestingly, both AS5466 and AS6830 uniformly do not allow access to rt.com (4.3). This uniform approach of blocking rt.com contrasts the varied treatment of vesti.ru, bringing out the complex web of content control within the country. Further discrepancies arise when the broader OONI data is compared with my testing on AS5466, in which Vesti.ru was found blocked over the DNS. This further illustrates how different similar censorship results can be even within the same network. These differences underscore the utmost importance that individual ISPs bring to internet governance in Ireland, and very detailed research is needed to tell the factors underpinning these inconsistencies. This disparity suggests that while some ISPs in Ireland proactively engage in blocking practices, others opt for a more open approach, leading to a varied internet experience for users depending on their specific internet provider. In conclusion, Ireland presents an interesting case study in internet governance, where the freedom of access to information appears primarily respected.

IE		AS 5466	2024-03-31 18:53 UTC	Web Connectivity Test	https://www.vesti.ru/	Accessible
IE		AS 5466	2024-03-31 18:53 UTC	Web Connectivity Test	http://www.vesti.ru/	Accessible
IE		AS 6830	2024-03-31 10:58 UTC	Web Connectivity Test	https://www.vesti.ru/	dns

Figure 4.2: OONI Data for Vesti.ru



IE		AS 5466	2024-04-01 13:54 UTC	Web Connectivity Test	https://www.rt.com/	tcp_ip
IE		AS 6830	2024-04-01 05:11 UTC	Web Connectivity Test	https://www.rt.com/	dns

Figure 4.3: OONI Data for Rt.com

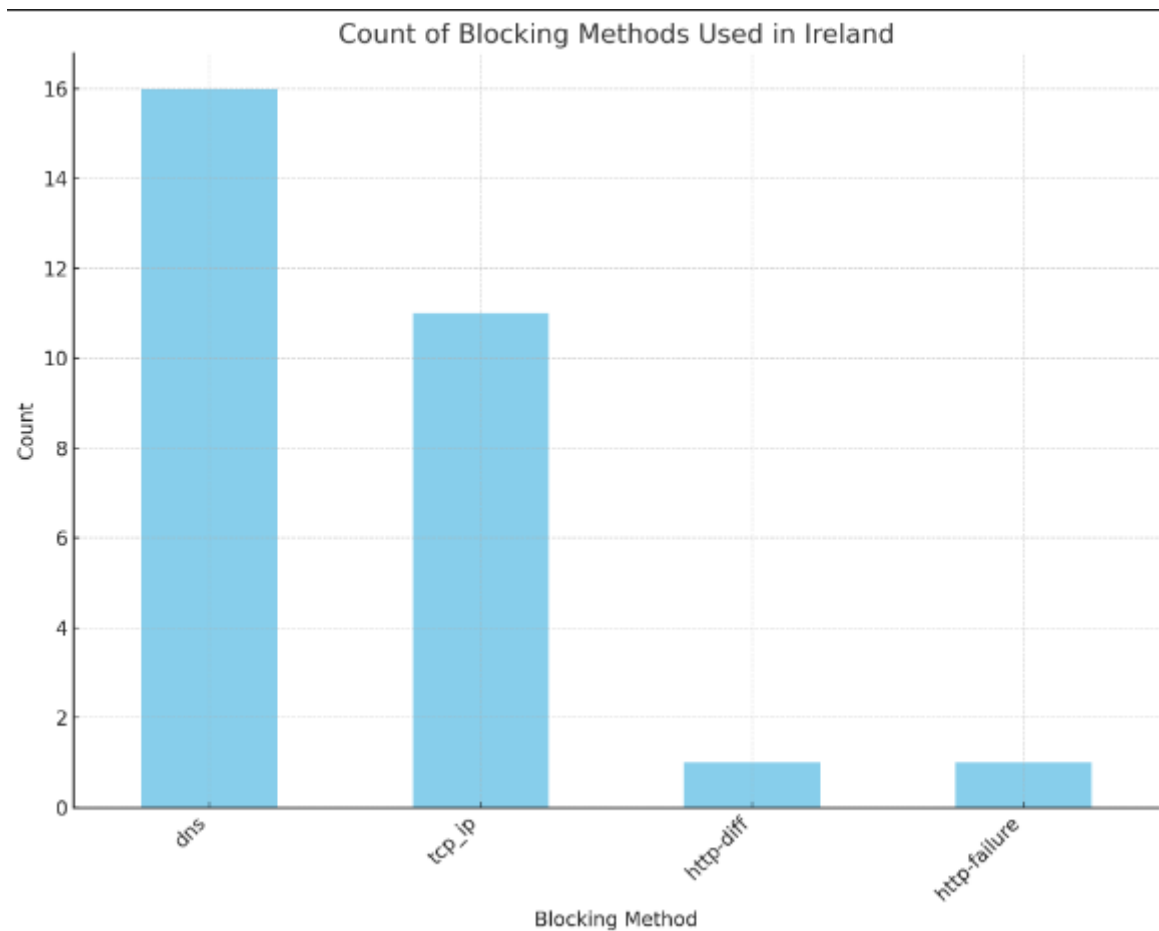


Figure 4.4: Censorship Mechanisms dominant in Ireland

Category	Total Sites	Blocked (%)
File Sharing	75	17.33
News Media	137	4.38
Public Health	53	3.77
Political Criticism	31	3.23
Sex Education	40	2.50
Search engines	41	2.44
Religion	67	1.49
Communication Tools	135	1.48
Anonymization and circumvention tools	121	0.83
Hosting and Blogging Platforms	136	0.74
Intergovernmental Organisations	14	0.00
Social Networking	84	0.00
Provocative Attire	17	0.00
Online Dating	17	0.00
Media Sharing	52	0.00
LGBT	83	0.00
Alcohol and Drugs	36	0.00
Human rights issues	154	0.00
Hate Speech	7	0.00
Hacking Tools	47	0.00
Government	33	0.00
Gaming	29	0.00
Gambling	28	0.00
Environment	45	0.00
Economics	37	0.00
E-Commerce	30	0.00
Culture	86	0.00
Control Content	24	0.00
Terrorism and militants	4	0.00

Table 4.2: OONI Curated list for Ireland Tested in Ireland

4.2 Comparison

Ireland List tested in Saudi In a cross-comparison of internet censorship using OONI-curated lists, the differences between Saudi Arabia and Ireland become pretty distinct. When the list curated for Ireland is tested in Saudi Arabia, it reveals that the latter implements a more extensive content-blocking regime. Saudi Arabia's censorship includes a wide range of categories that are not blocked in Ireland, such as Alcohol and Drugs, Culture, Gambling, Gaming, Human Rights Issues, LGBT, Media Sharing, Online Dating, Provocative Attire, Social Networking, and Terrorism and Militants.

One fascinating feature of this comparative investigation is the various geopolitical details of censorship between the two countries. For instance, while Ireland blocks certain Russian media outlets, Saudi Arabia does not. However, it does block Turkish, Lebanese, and Iranian news platforms such as Daily Sabah, Al Manar, and IRNA. This reflects each country's respective political climates and relationships, with censorship serving as a tool to shield citizens from what may be deemed as "disinformation" or to align with national political and security interests. Another notable element is the two countries' agreement prohibiting illegal file-sharing websites. Both Saudi Arabia and Ireland have blocked access to sites such as LibGen and The Pirate Bay, demonstrating a shared commitment to intellectual property laws and the worldwide anti-piracy movement.

The analysis of the blocking methods used for Irish URLs when tested in Saudi Arabia, as indicated by the graph 4.5, shows a reliance on HTTP failures as the primary method of censorship. This method, which involves the complete failure to retrieve content from a server, is a blunt but effective means of preventing access to information. Its prevalence suggests that, for the Saudi authorities, rendering a website completely inaccessible is often preferable to more sophisticated methods that might allow some access or provide users with information on why a site is blocked. In contrast to Ireland, where DNS blocking is more prominent, the use of HTTP failures could be due to a combination of factors, including a preference for absolute control over the content and a desire for less transparency in the censorship process. HTTP failures do not always disclose the reason for the blocking, leaving users uncertain whether a site is down due to censorship, technical issues, or other reasons.

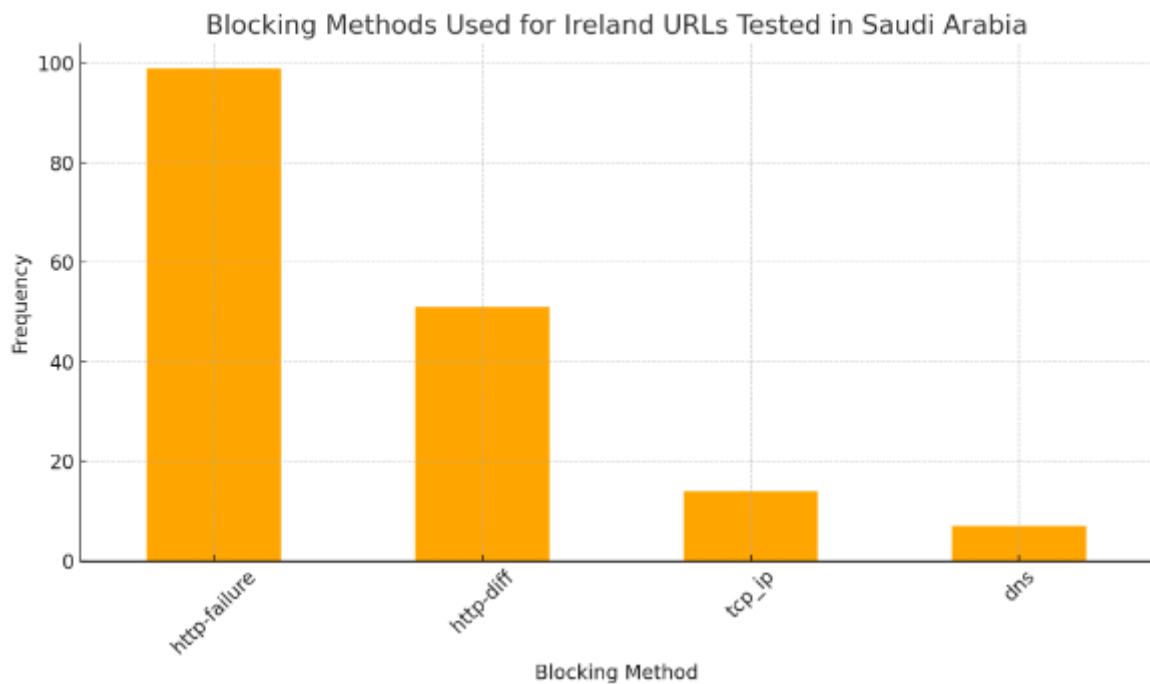


Figure 4.5: OONI Curated List for Ireland tested in Saudi Arabia

Saudi List In Ireland Conversely, when tested in Ireland, the list tailored for Saudi Arabia does not highlight unique categories blocked within Irish borders. A graph depicting the number of URLs blocked in both countries by category shows overlap (4.7) in censorship practices, focusing on news media and file sharing. Notably, Viber, a VoIP service, remains blocked in Saudi Arabia, aligning with its policies on communication tools, which have historically restricted access to such services to control information flow and maintain telecom monopolies [61].

In Ireland, the censorship reflected in the Saudi curated list is minimal, with only 56 out of 2300 URLs blocked, contrasting with nearly 300 in Saudi Arabia. Moreover, the OONI data and the analysis from Censored Planet support our findings, especially regarding the interference with LGBT content in Saudi Arabia. This is evident in the Censored Planet dashboard data, which indicates that a majority of networks in Saudi Arabia engage in some form of interference with LGBT sites, underscoring the country's strict stance on content that diverges from its cultural and religious norms.

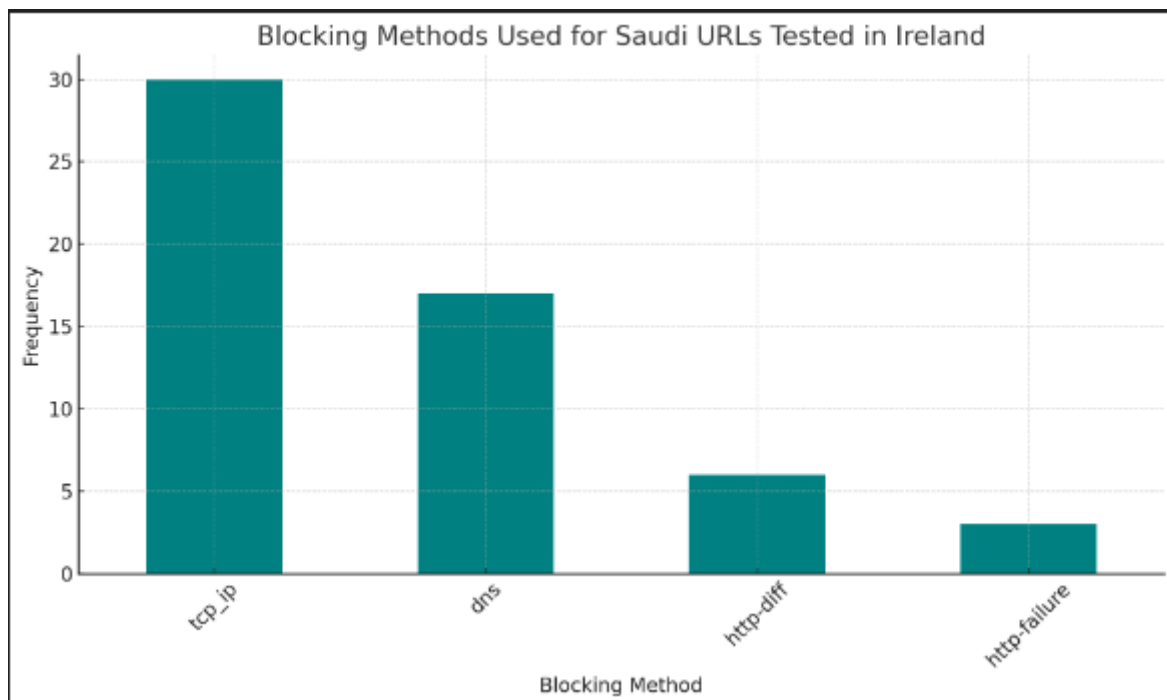


Figure 4.6: Blocking Methods used for Saudi URLS in Ireland

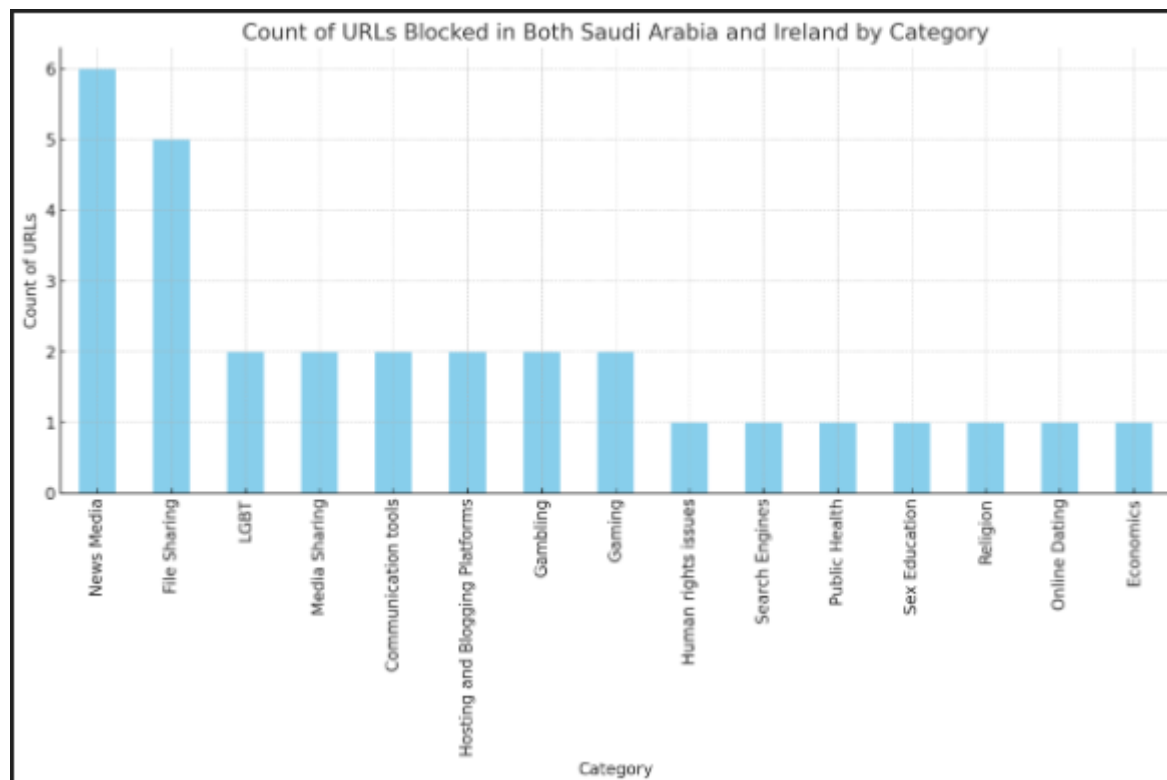


Figure 4.7: Overlap between Ireland and Saudi Arabia

Category	Total Sites	Blocked Percentage
Alcohol and Drugs	46	0.0%
Anonymization and Circumvention Tools	141	0.71%
Communication Tools	138	1.45%
Control Content	45	0.0%
Culture	96	0.0%
E-commerce	31	0.0%
Economics	64	1.56%
Environment	57	0.0%
File Sharing	94	15.96%
Gambling	64	4.69%
Gaming	38	5.26%
Government	44	0.0%
Hacking Tools	61	0.0%
Hate Speech	7	0.0%
Hosting and Blogging Platforms	149	1.34%
Human Rights Issues	187	0.53%
Intergovernmental Organisations	14	0.0%
LGBT	108	1.85%
Media Sharing	78	2.56%
News Media	301	4.65%
Online Dating	55	5.45%
Political Criticism	69	1.45%
Provocative Attire	40	0.0%
Public Health	71	1.41%
Religion	115	2.61%
Search Engines	51	1.96%
Sex Education	52	0.92%
Social Networking	89	1.02%
Terrorism and Militants	7	0.0%

Table 4.3: OONI Curated List for Saudi Arabia tested in Ireland

4.3 Top 100 Websites Worldwide

The contrast between Saudi Arabia and Ireland becomes clear in an examination of the top 100 most visited websites worldwide. Ireland is committed to open access, with no instances of blocking across various categories in the data set. On the other hand, Saudi Arabia's approach is more restrictive, with a particular focus on Corporate and Business Solutions and Social Networking categories. This is shown by the blocking of VK (4.9), a popular Russian social networking site, which, according to OONI's database, has consistently faced HTTP failure when accessed from within Saudi Arabia.

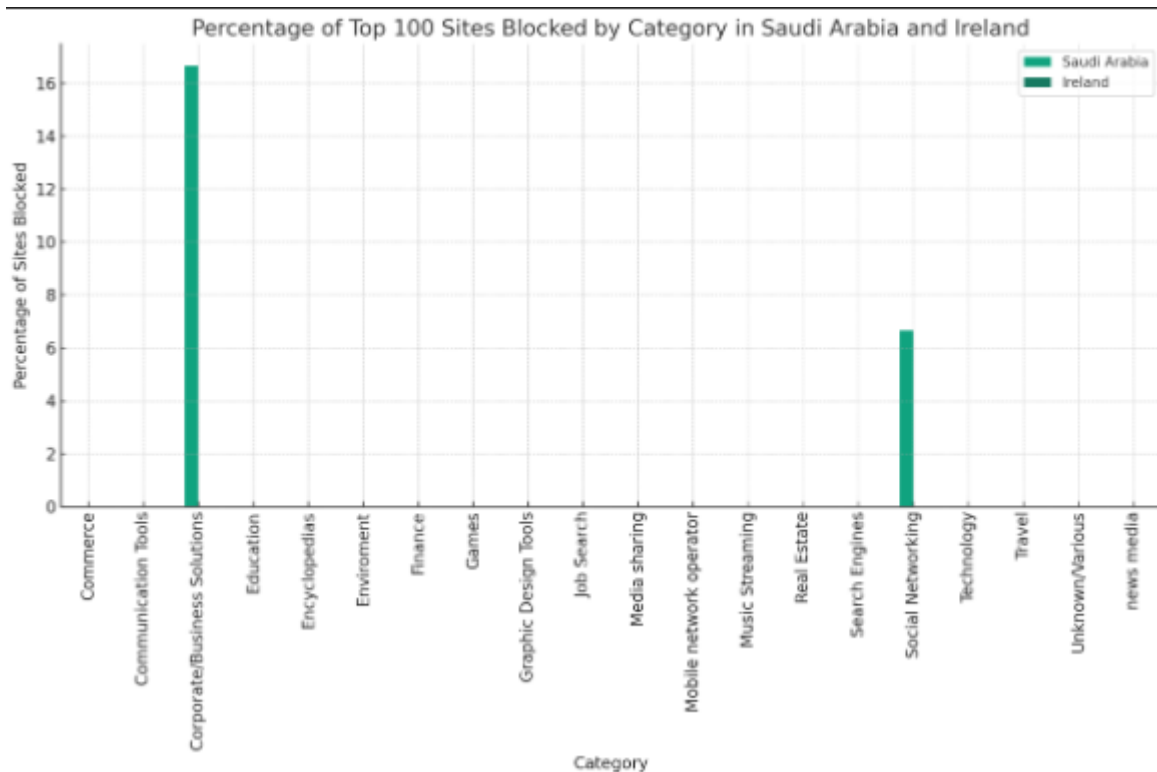


Figure 4.8: Comparing Ireland and Saudi Worldwide restrictions

SA		AS 43766	2024-04-07 23:48 UTC	Web Connectivity Test	https://vk.com/	http-failure
SA		AS 43766	2024-04-07 23:21 UTC	Web Connectivity Test	https://vk.com/	http-failure
SA		AS 43766	2024-04-07 22:53 UTC	Web Connectivity Test	https://vk.com/	http-failure
SA		AS 43766	2024-04-07 22:50 UTC	Web Connectivity Test	https://vk.com/	http-failure

Figure 4.9: OONI Measurement for VK

4.4 Random URLs

In categories such as Gambling, Government/Political Criticism, LGBTQ+, social networking and news media, 100 random websites were chosen and tested in both countries for comparison; Saudi Arabia exhibits substantial levels of blocking, with the most pronounced censorship seen in Gambling, where over 80% of URLs are blocked. Conversely, Ireland shows minimal to no blocking in these categories, reflecting a higher tolerance for content diversity and freedom of expression. The graph also indicates that both countries enforce some censorship on News Media, though Saudi Arabia's blocking rate is significantly higher.

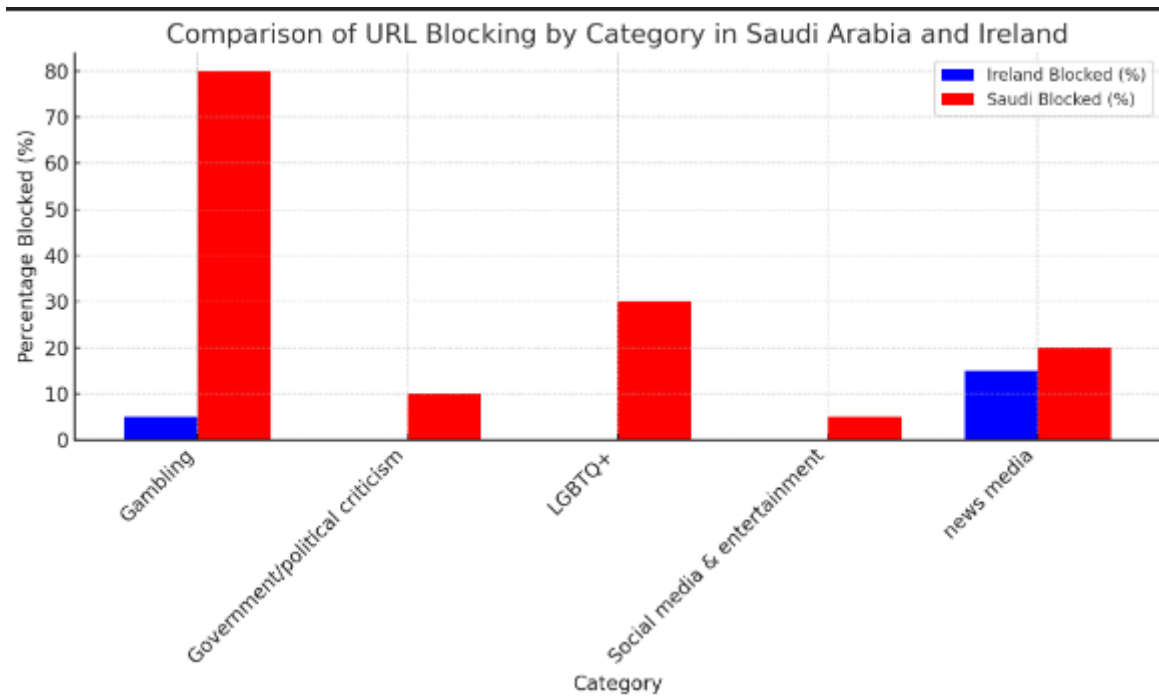


Figure 4.10: Graph comparing restrictions on 100 Random URLs

4.5 Cross-Comparison of Most visited Websites in both countries

Saudi: The graph illustrating the percentage of the top 100 sites in Ireland blocked by category in Saudi Arabia shows a significant result in one category: Gambling, with 100% of such sites being blocked (4.11). The sites include Paddy Power, Bet365, and the Lottery. All other categories in the dataset show no instances of blocking.

Ireland: In a cross-national test of web accessibility, Ireland's open internet policies were once again underscored. When the top 100 Saudi websites were tested in Ireland, only one site, "ejar.sa" was inaccessible. Ejar.sa is an official Saudi Arabian portal designed to facilitate real estate rentals and safeguard the rights of all parties involved in leasing transactions. The blocking of this singular site in Ireland is an outlier in what is otherwise a pattern of unrestricted access. It is attributable to specific operational reasons rather than a broad censorship policy.

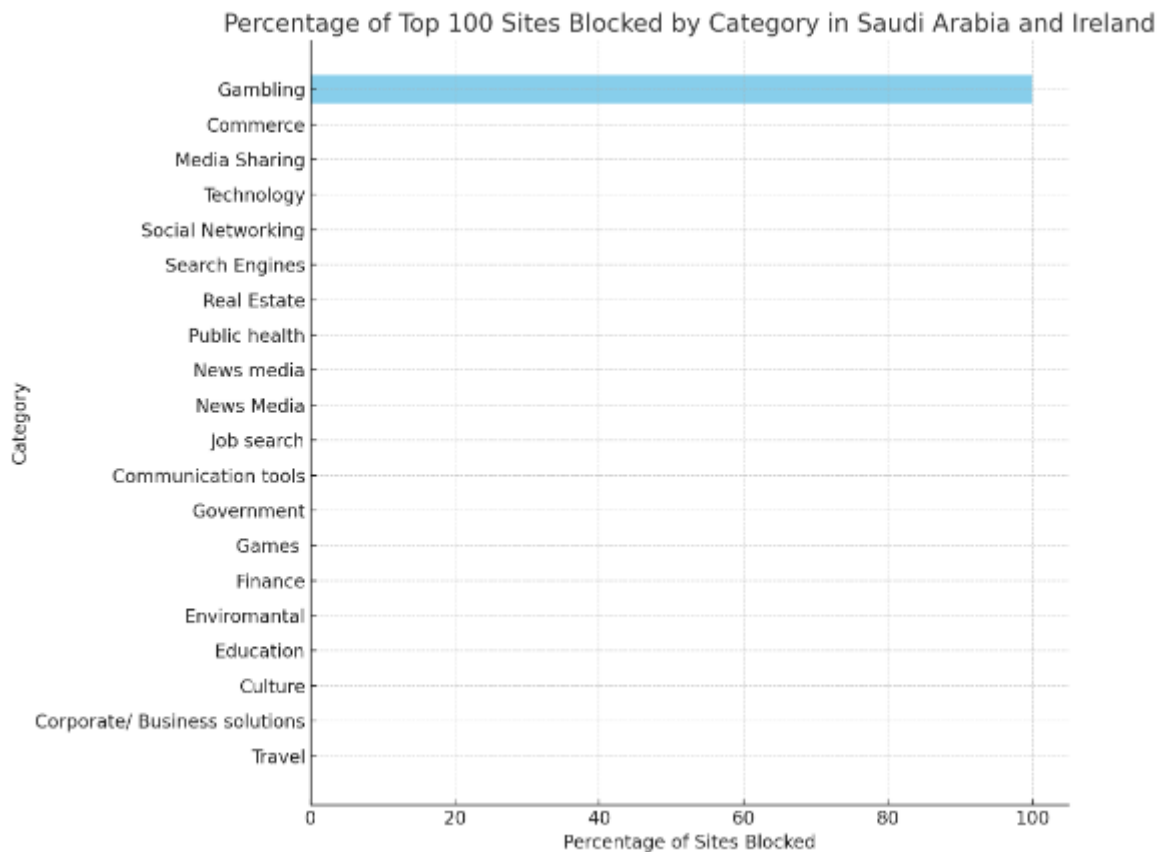


Figure 4.11: Ireland top 100 URLs tested in Saudi

4.6 Circumvention Tests

It is interesting to observe that both Ireland and Saudi Arabia do not impose censorship on circumvention tools like Tor and Psiphon. This contrasts with Saudi Arabia's approach to the majority of conventional VPNs. Typically, these tools are targeted because they allow users to bypass state-imposed content restrictions and surveillance measures. The exemption of tools like Tor and Psiphon from this censorship is interesting. Tor and Psiphon are designed to obfuscate user activities, making it difficult for authorities to monitor and block them without also compromising general internet traffic, which could explain their availability. This could reflect an intentional decision by some countries to allow certain anonymizing services, possibly to retain the appearance of openness. Alternatively, it could be due to the technical challenges involved in blocking them effectively. Furthermore, it may indicate an acknowledgement of the critical role these technologies play for some users in preserving their online privacy and freedom, or it could indicate an oversight or a gap in the implementation of censorship regulations. implementation of censorship regulations.

```
circumvention          2 tested
Kaopu Cloud HK Limited 0 blocked
AS138915 (SA)
```

Figure 4.12: Snippet from Saudi VPS

4.7 Deep analysis of blocking methods

The occurrence of TCP/IP blocking does not definitively indicate an act of censorship. While it might suggest that access to a particular website is being restricted, TCP/IP blocking can sometimes result from technical issues unrelated to deliberate content suppression. For instance, if a website is no longer active on the internet, attempts to connect to it would result in TCP/IP blocking. Similarly, network configuration issues or server downtime could also manifest as TCP/IP blocks. Misinterpreting these incidents as censorship without further investigation could lead to misleading conclusions about a country's internet freedom status. Therefore, while TCP/IP blocking is a valuable signal worth monitoring, it should be understood within a broader analysis that considers various potential causes—ranging from technical glitches to intentional restrictions—to accurately assess the presence and extent of web censorship.

To understand the censorship mechanisms further, we did a deep analysis of exactly how the blocking is being done, whether it is through DNS tampering or is it through the SNI. we chose 1 website for each country to do further analysis. For Ireland, we chose (<https://thepiratebay.org/>) which, from our collected Ooni data, is blocked via DNS. We took several steps to test this and see exactly how it was being done:

- **DNS Resolution Tests:** We used the `dig` command to compare DNS resolutions from my local DNS resolver against public DNS resolvers (Google's 8.8.8.8 and Cloudflare's 1.1.1.1).
- **Direct IP Access Test:** We entered the IP address returned by my local DNS resolver directly into a web browser.
- **curl Test with Cloudflare IPs:** We executed a `curl` command to simulate a web browser request using the IP addresses returned by the public DNS resolvers.

```
curl -v -H "Host: thepiratebay.org" https://162.159.137.6
```

- **Bypassing SSL Certificate Verification:** We used `curl` with the `-k` option to bypass SSL verification, and `-L` to follow redirects.

```
curl -v -k -L -H "Host: thepiratebay.org" https:// 162.159.137.6
```

- **Traceroute Test:** We used to traceroute to map the packet's path to the destination. This traceroute command maps the route data packets take from the local machine to the domain "thepiratebay. org". It lists each hop the packets take across the network to reach the final destination, which should be the server associated with "thepiratebay.org"

By doing these tests we found out:

DNS Inconsistency: the local DNS resolver returned an IP address that, when accessed with the web browser, leads to an ISP block page. Public DNS resolvers returned different IP addresses, indicating they are not being tampered with and are likely giving the correct addresses managed by Cloudflare.

Successful Connection to Cloudflare: The curl command with the -k option successfully made a connection to Cloudflare's network, but the request was ultimately redirected back to an IP address that the ISP was blocking.

The traceroute successfully traced the path through my local network and my ISP's network up to a point in Dublin's Eircom network infrastructure. However, it wasn't able to complete the trace to thepiratebay.org because it encountered a block or filter somewhere along the way.

ISP Block Confirmation: The ISP's block appears to be enforced after the initial DNS resolution, at the connection or content delivery stage. The use of a specific block page indicates a deliberate blocking mechanism.

The tests indicate that the ISP is be redirecting "thepiratebay.org" queries to a block page, a potential sign of DNS tampering. Discrepancies in DNS results where only the local resolver leads to a block, while public resolvers don't, support this. It seems like the ISP specifically targets "thepiratebay.org" for blocking rather than the whole Cloudflare IP.

For Saudi Arabia, we decided to test (<https://nordvpn.com/>) which we know from our data that it is being blocked via HTTP-failure. These are the steps we took: We did a "dig" to see if the domains are resolved correctly and since we knew this is not DNS and IP based blocking, as expected they were resolved.

Traceroute: next we did a traceroute test (4.13) and the traceroute to nordvpn.com completed, indicating that the network path to the server hosting nordvpn.com is reachable. There are no interruptions in the path, as seen by the trace completing and reaching an IP address associated with nordvpn.com.

```

root@ubuntu:~# traceroute nordvpn.com
traceroute to nordvpn.com (104.16.208.203), 30 hops max, 60 byte packets
 1 38.54.114.1 (38.54.114.1) 12.250 ms 12.258 ms 12.241 ms
 2 10.8.8.1 (10.8.8.1) 0.455 ms 0.519 ms 0.502 ms
 3 178.86.50.12 (178.86.50.12) 1.106 ms 1.121 ms 178.86.50.108 (178.86.50.108) 0.593 ms
 4 10.188.195.73 (10.188.195.73) 13.247 ms 12.728 ms 13.190 ms
 5 * * *
 6 * * *
 7 * * *
 8 * 104.16.208.203 (104.16.208.203) 70.162 ms 69.732 ms

```

Figure 4.13: Traceroute Snippet

OpenSSL Connection Test: The OpenSSL command attempts to establish a TLS connection directly with nordvpn.com on port 443. However, it does not retrieve a server certificate ("no peer certificate available"), which is unusual as it should normally receive a certificate used for establishing a secure connection. • **Curl HTTPS Request:** The curl command with the -v verbose flag and -k to bypass SSL certificate verification attempts to connect to nordvpn.com. The connection is initiated, but the server resets the connection. This is indicated by the message "Connection reset by peer," which suggests that something in the network path is interrupting the connection.

Curl with Different User-Agents: The last curl command uses a custom user-agent string to mimic Googlebot. The server also resets these connections, just as with the standard curl attempt. This consistent connection reset, regardless of user-agent, suggests that the blocking mechanism is indifferent to the type of client making the request.

```

curl -v -A "Mozilla/5.0 (compatible; Googlebot/2.1;
+http://www.google.com/bot.html) "https://nordvpn.com

```

The fact that connections are being reset during the TLS handshake phase implies that there is a filtering mechanism that interrupts connections specifically when trying to negotiate HTTPS. This could be due to:

Deep Packet Inspection (DPI) employed by the network to prevent access to VPN services.

Network policies that specifically target VPN-related traffic. The test results from the VPS in Riyadh indicate that the connection to nordvpn.com is consistently interrupted during the TLS handshake process. The evidence points towards an active network-level interference. This pattern of connection resets, especially when they occur right after initiating a secure connection attempt, points to a deliberate disruption likely implemented by the network provider (AS), suggesting that access to nordvpn.com is being actively blocked.

4.8 Conclusion

This chapter compares and contrasts internet filtering policies in Saudi Arabia and Ireland, highlighting the key distinctions between their strategies and underlying goals. A vast range of material categories are heavily restricted in Saudi Arabia, including gambling, methods for anonymization, and subjects that go against social norms. HTTP diffs and HTTP failures are the standard censoring techniques, indicating a thorough control plan for information access.

Ireland, on the other hand, demonstrates very little censorship, mainly focusing on file sharing in order to uphold copyright regulations. The most popular technique is DNS blocking, suggesting a focused and less invasive approach to internet governance. This is evidence of a more significant commitment to preserving internet freedom, as the majority of material categories exhibit little to no filtering.

The sharp differences between the two nations are highlighted by the cross-national analysis that makes use of OONI-curated lists, with Saudi Arabia implementing a more extensive content filtering policy. We also discussed how geopolitical factors affect censoring policies and how important ISPs are to internet governance in both nations. A thorough analysis was undertaken to understand the technological components of censorship better.

5 Conclusion

5.1 Conclusion

This study examined the scope, workings, and effects of internet censorship in Ireland and Saudi Arabia in an attempt to shed light on these differences. This thesis has offered thorough empirical research highlighting the difference between the two countries' censorship strategies through the use of the OONI probe, a virtual private server located in Riyadh, and the Windows Subsystem for Linux.

The main conclusions show that Saudi Arabia has a thorough censorship policy that targets a wide range of content categories and is consistent with its strict socio-political goals. The main method of HTTP-based blocking suggests a preference for direct control over the flow of information. On the other hand, Ireland takes a less restrictive stance while strongly emphasising copyright enforcement. This implies a kind of governance prioritising protecting internet freedom while adhering to intellectual property regulations. The report also clarifies the impact of certain ISPs in Ireland, suggesting a more decentralised internet governance environment that differs according to the ISP. The wide variation in censoring policies across the nation implies that ISP independence and regional regulatory compliance are important factors in determining how user experience is shaped.

Regarding attempts at circumvention, the study indicates that although programs such as VPNs and Tor continue to be useful in some situations, their effectiveness depends on how sophisticated the censorship strategies used by states are. This highlights the ongoing cat-and-mouse game between censorship mechanisms and circumvention technologies.

5.2 Future work

Future research could incorporate longitudinal studies on censorship trends and expand comparative analyses to include a broader range of countries. They would track the evolution of internet censorship over time, providing insights into how political shifts, technological advancements, and changes in social norms influence censorship practices. This approach would offer an understanding of the censorship landscape, helping predict

future trends and their impacts on society. Additionally, Expanding the scope of comparative research to include numerous countries with varying degrees of internet freedom can improve our understanding of global internet governance.

Bibliography

- [1] G. Aceto and A. Pescapè, "Internet censorship detection: A survey," *Computer Networks*, vol. 83, pp. 381–421, 2015. [Online]. Available: <https://censorbib.nymity.ch/pdf/Aceto2015b.pdf>
- [2] A. Master and C. Garman, "A worldwide view of nation-state Internet censorship," in *Free and Open Communications on the Internet*, 2023. [Online]. Available: <https://www.petsymposium.org/foci/2023/foci-2023-0008.pdf>
- [3] J. Hyland, "Internet censorship: An integrative review of technologies employed to limit access to the internet, monitor user actions, and their effects on culture," 2020. [Online]. Available: <https://digitalcommons.liberty.edu/honors/997/>
- [4] N. Koumartzis and A. Veglis, "Internet censorship and regulation systems in democracies: Emerging research and opportunities: Emerging research and opportunities," 2020. [Online]. Available: <https://books.google.ie/books?id=aETHDwAAQBAJ&lpg=PR1&pg=PR1#v=onepage&q&f=false>
- [5] R. Deibert, J. Palfrey, R. Rohozinski, and J. Zittrain, *Access controlled: The shaping of power, rights, and rule in cyberspace*. the MIT Press, 2010. [Online]. Available: <http://library.oapen.org/handle/20.500.12657/26076>
- [6] D. Katira, G. Grover, K. Singh, and V. Bansal, "CensorWatch: On the implementation of online censorship in India," in *Free and Open Communications on the Internet*, 2023. [Online]. Available: <https://www.petsymposium.org/foci/2023/foci-2023-0006.pdf>
- [7] A. Cahn, S. Alfeld, P. Barford, and S. Muthukrishnan, "An empirical study of web cookies," in *Proceedings of the 25th international conference on world wide web*, 2016, pp. 891–901. [Online]. Available: <https://dl.acm.org/doi/abs/10.1145/2872427.2882991>
- [8] Human Rights Watch, "Meta: Systemic censorship of palestine content," December 2023, accessed: 2024-03-25. [Online]. Available: <https://www.hrw.org/news/2023/12/20/meta-systemic-censorship-palestine-content>

- [9] P. Jain, Y. Munjal, J. Gera, and P. Gupta, "Performance analysis of various server hosting techniques," *Procedia Computer Science*, vol. 173, pp. 70–77, 2020. [Online]. Available: <https://doi.org/10.1016/j.procs.2020.06.010>
- [10] M. Taha and A. Ali, "Redirection and protocol mechanisms in content delivery network-edge servers for adaptive video streaming," *Applied Sciences*, vol. 13, no. 9, p. 5386, 2023. [Online]. Available: <https://doi.org/10.3390/app13095386>
- [11] D. E. Eisenbud, C. Yi, C. Contavalli, C. Smith, R. Kononov, E. Mann-Hielscher, A. Cilingiroglu, B. Cheyney, W. Shang, and J. D. Hosein, "Maglev: A fast and reliable software network load balancer," in *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*, 2016, pp. 523–535. [Online]. Available: <https://www.usenix.org/sites/default/files/nsdi16-paper-eisenbud.pdf>
- [12] Open Observatory of Network Interference, "Ooni: Measuring internet censorship," <https://ooni.org/>, 2023, accessed: 2024-02-01.
- [13] Censored Planet, "Censored planet: A global censorship observatory," <https://censoredplanet.org/>, 2023, accessed: 2024-02-01.
- [14] LightNode, "Lightnode - flexible and scalable cloud computing service," 2024, accessed: 2024-04-12. [Online]. Available: <https://lightnode.com/en-US/product>
- [15] P. P. Tsang, A. Kapadia, C. Cornelius, and S. W. Smith, "Nymble: Blocking misbehaving users in anonymizing networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 2, pp. 256–269, 2009. [Online]. Available: <https://doi.org/10.1109/TDSC.2009.38>
- [16] R. Raman, A. Stoll, J. Dalek, R. Ramesh, W. Scott, and R. Ensafi, "Measuring the deployment of network censorship filters at global scale," 2020. [Online]. Available: <https://doi.org/10.14722/ndss.2020.23099>
- [17] C. Abdelberi, T. Chen, M. Cunche, E. D. Cristofaro, A. Friedman, and M. Kâafar, "Censorship in the wild: Analyzing internet filtering in syria," *Proceedings of the 2014 Conference on Internet Measurement Conference*, 2014. [Online]. Available: <https://doi.org/10.1145/2663716.2663720>
- [18] A. Shishkina and L. Issaev, "Internet censorship in arab countries: Religious and moral aspects," *Religions*, vol. 9, no. 11, p. 358, 2018. [Online]. Available: <https://doi.org/10.3390/rel9110358>
- [19] K. Xu, Z. Duan, Z. Li-Zhang, and J. Chandrashekar, "On properties of internet exchange points and their impact on as topology and relationship," in *Networking 2004*, ser. Lecture Notes in Computer Science, vol. 3042. International Conference on Research in Networking, 2004, pp. 284–295.

- [20] E. Al-Shaer, H. Hamed, R. Boutaba, and M. Hasan, "Conflict classification and analysis of distributed firewall policies," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 2069–2084, 2005. [Online]. Available: <https://doi.org/10.1109/JSAC.2005.854119>
- [21] S. Wolfgarten, "Investigating large-scale internet content filtering," Master's thesis, Dublin City University, Dublin, Ireland, Aug 2006, m.Sc. in Security & Forensic Computing 2005/2006.
- [22] M. Dornseif, "Government mandated blocking of foreign web content," Available at <http://md.hudora.de/>, 2004, [Accessed 2 April 2004].
- [23] P.-W. Tsai, A. C. Risdianto, M. H. Choi, S. K. Permal, and T. C. Ling, "Sd-brov: An enhanced bgp hijacking protection with route validation in software-defined exchange," *Future Internet*, vol. 13, no. 7, p. 171, 2021. [Online]. Available: <https://doi.org/10.3390/fi13070171>
- [24] T. Lin, T. Alpcan, and K. Hinton, "A game-theoretic analysis of energy efficiency and performance for cloud computing in communication networks," *IEEE Systems Journal*, vol. 11, pp. 649–660, 2017. [Online]. Available: <https://doi.org/10.1109/JSYST.2015.2451195>
- [25] R. Ensafi, P. Winter, A. Mueen, and J. R. Crandall, "Analyzing the great firewall of china over space and time." *Proc. Priv. Enhancing Technol.*, vol. 2015, no. 1, pp. 61–76, 2015. [Online]. Available: <https://petsymposium.org/popets/2015/popets-2015-0005.pdf>
- [26] F. Alharbi, M. Faloutsos, and N. Abu-Ghazaleh, "Opening digital borders cautiously yet decisively: Digital filtering in saudi arabia," in *Proc. of: Free and Open Communications on the Internet*. USENIX, 2020.
- [27] A. Chaabane, T. Chen, M. Cunche, E. De Cristofaro, A. Friedman, and M. A. Kaafar, "Censorship in the wild: Analyzing internet filtering in syria," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, 2014, pp. 285–298. [Online]. Available: <https://doi.org/10.1145/2663716.2663720>
- [28] W. Ye and L. Zhao, "'i know it's sensitive': Internet censorship, recoding, and the sensitive word culture in china," *Discourse, Context & Media*, vol. 51, p. 100666, 2023. [Online]. Available: <https://doi.org/10.1016/j.dcm.2022.100666>
- [29] S. Satija and R. Chatterjee, "Blindtls: Circumventing tls-based https censorship," in *Proceedings of the ACM SIGCOMM 2021 Workshop on Free and Open Communications on the Internet*, 2021, pp. 43–49. [Online]. Available: <https://doi.org/10.1145/3473604.3474564>

- [30] J. L. Hall, M. D. Aaron, S. Adams, A. Andersdotter, B. Jones, and N. Feamster, "A survey of worldwide censorship techniques," Internet-Draft draft-irtf-pearg-censorship-04. Internet Engineering Task . . . , Tech. Rep., 2020. [Online]. Available: <https://www.ietf.org/archive/id/draft-irtf-pearg-censorship-10.html#name-technical-prescription>
- [31] S. Ren, B. Liu, F. Yang, X. Wei, X. Yang, and C. Wang, "Blockdns: enhancing domain name ownership and data authenticity with blockchain," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9013817?casa_token=ZcF0e_il-30AAAAA:xfGkvmZCFMMFOINXUjvmlg3XScWDleqSVa5LB9cZ9z3luGIRf9D7O3rQsEoWG0XNyjqDJwAqOfcV
- [32] R. Houser, S. Hao, Z. Li, D. Liu, C. Cotton, and H. Wang, "A comprehensive measurement-based investigation of dns hijacking," in *2021 40th International Symposium on Reliable Distributed Systems (SRDS)*. IEEE, 2021, pp. 210–221. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9603621?casa_token=F5Q4qxPPN70AAAAA:dLkx0q_PsOvdX2F0XOHftJWiRHm7pij7EpjInx_zOqUbptUbiy6eg2UZf7_c_h5QDbJNpU9zBjK
- [33] M. Wander, C. Boelmann, L. Schwittmann, and T. Weis, "Measurement of globally visible dns injection," *IEEE Access*, vol. 2, pp. 526–536, 2014. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/6814824>
- [34] D. Xue, R. Ramesh, L. Evdokimov, A. Viktorov, A. Jain, E. Wustrow, S. Basso, and R. Ensafi, "Throttling twitter: an emerging censorship technique in russia," in *Proceedings of the 21st ACM Internet Measurement Conference*, 2021, pp. 435–443. [Online]. Available: <https://doi.org/10.1145/3487552.3487858>
- [35] G. Aceto, A. Botta, P. Marchetta, V. Persico, and A. Pescapé, "A comprehensive survey on internet outages," *Journal of Network and Computer Applications*, vol. 113, pp. 36–63, 2018. [Online]. Available: <https://doi.org/10.1016/j.jnca.2018.03.026>
- [36] J. Curran, "The internet of history: Rethinking the internet's past," in *Misunderstanding the internet*. Routledge, 2016, pp. 48–84. [Online]. Available: <https://books.google.ie/books?id=4GeFCwAAQBAJ&lpg=PA48&ots=R9bAFUA7J2&dq=censorship%20Egypt's%20internet%20blackout%20&lr&pg=PA48#v=onepage&q&f=false>
- [37] M. N. Momen and D. Das, "Mediated democracy and internet shutdown in india," *Journal of Information, Communication and Ethics in Society*, vol. 19, no. 2, pp. 222–235, 2021. [Online]. Available: <https://www.emerald.com/insight/content/doi/10.1108/JICES-07-2020-0075/full/pdf?title=mediated-democracy-and-internet-shutdown-in-india>

- [38] R. Padmanabhan, A. Filastò, M. Xynou, R. S. Raman, K. Middleton, M. Zhang, D. Madory, M. Roberts, and A. Dainotti, "A multi-perspective view of internet censorship in myanmar," in *Proceedings of the ACM SIGCOMM 2021 Workshop on Free and Open Communications on the Internet*, 2021, pp. 27–36. [Online]. Available: <https://doi.org/10.1145/3473604.3474562>
- [39] B. AKDUMAN, "From the great wall to the great firewall: A historical analysis of surveillance," *Uluslararası Sosyal Bilimler Dergisi*, vol. 7, no. 28, pp. 442–469, 2023. [Online]. Available: https://www.sobider.net/FileUpload/ep842424/File/21.from_the_great_wall_to_the_great_firewall.pdf
- [40] W. C. Burghard, "Covertnet: Circumventing web surveillance using covert channels," 2019. [Online]. Available: https://digitalcommons.bard.edu/senproj_f2019/38
- [41] M. Wu, J. Sippe, D. Sivakumar, J. Burg, P. Anderson, X. Wang, K. Bock, A. Houmansadr, D. Levin, and E. Wustrow, "How the great firewall of china detects and blocks fully encrypted traffic," in *32nd USENIX Security Symposium (USENIX Security 23)*, 2023, pp. 2653–2670. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity23/presentation/wu-mingshi>
- [42] A. Shirokanova and O. Silyutina, "Internet regulation media coverage in russia: Topics and countries," in *Proceedings of the 10th ACM Conference on Web Science*, 2018, pp. 359–363. [Online]. Available: <https://doi.org/10.1145/3201064.3201102>
- [43] The Journal. (2022, mar) Russian state-controlled news channel rt removed from irish screens; russia blocks twitter and facebook. [Online]. Available: <https://www.thejournal.ie/russian-news-channel-rt-banned-eu-ireland-5698500-Mar2022/>
- [44] OpenNet Initiative, "Internet filtering in saudi arabia in 2004," 2004, accessed: 2024-03-27. [Online]. Available: <https://opennet.net/studies/saudi>
- [45] A. Monea, "3. overblocking," *The Digital Closet*, 2022. [Online]. Available: <https://assets.pubpub.org/000jvjs0/21b37f5b-4427-4f75-815a-4ced85a18291.pdf>
- [46] A. Ramanujan and B. A. Varghese, "Internet censorship based on bayes learning model," in *Second International Conference on Networks and Advances in Computational Technologies: NetACT 19*. Springer, 2021, pp. 49–60. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-49500-8_5
- [47] B. R. Turner, "An investigation into the efficacy of url content filtering systems," 2021. [Online]. Available: <https://ro.ecu.edu.au/theses/2409/>
- [48] J. Penney, "Internet surveillance, regulation, and chilling effects online: A comparative case study," *Regulation, and Chilling Effects Online: A Comparative Case Study (May*

- 27, 2017), vol. 6, no. 2, 2017. [Online]. Available:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2959611
- [49] B. Canes-Wrone and M. C. Dorf, "Measuring the chilling effect," *NYUL Rev.*, vol. 90, p. 1095, 2015. [Online]. Available:
<https://heinonline.org/HOL/P?h=hein.journals/nylr90&i=1113>
- [50] C. Kim and W. Shin, "Harassment of journalists and its aftermath: Anti-press violence, psychological suffering, and an internal chilling effect," *Digital Journalism*, pp. 1–17, 2022. [Online]. Available: <https://doi.org/10.1080/21670811.2022.2034027>
- [51] M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, "The echo chamber effect on social media," *Proceedings of the National Academy of Sciences*, vol. 118, no. 9, p. e2023301118, 2021. [Online]. Available:
<https://doi.org/10.1073/pnas.2023301118>
- [52] P. Törnberg, "Echo chambers and viral misinformation: Modeling fake news as complex contagion," *PLoS one*, vol. 13, no. 9, p. e0203958, 2018. [Online]. Available:
<https://doi.org/10.1371/journal.pone.0203958>
- [53] L. T. L. Terren and R. B.-B. R. Borge-Bravo, "Echo chambers on social media: A systematic review of the literature," *Review of Communication Research*, vol. 9, 2021. [Online]. Available: <https://www.rcommunicationr.org/index.php/rcr/article/view/16>
- [54] Y. Mou, K. Wu, and D. Atkin, "Understanding the use of circumvention tools to bypass online censorship," *New Media & Society*, vol. 18, no. 5, pp. 837–856, 2016. [Online]. Available: <https://doi.org/10.1177/1461444814548994>
- [55] Y. Wang, P. Ji, B. Ye, P. Wang, R. Luo, and H. Yang, "GoHop: Personal VPN to defend from censorship," in *International Conference on Advanced Communication Technology*. IEEE, 2014. [Online]. Available:
<https://censorbib.nymity.ch/pdf/Wang2014a.pdf>
- [56] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," DTIC Document, Technical Report, 2004.
- [57] F. Douglas, Rorshach, W. Pan, and M. Caesar, "Salmon: Robust proxy distribution for censorship circumvention," *Privacy Enhancing Technologies*, vol. 2016, no. 4, pp. 4–20, 2016. [Online]. Available: <https://censorbib.nymity.ch/pdf/Douglas2016a.pdf>
- [58] "Psiphon - access the entire internet," <https://www.psiphon.ca>, accessed: 2024-04-14.
- [59] R. Ramesh, R. S. Raman, A. Virkud, A. Dirksen, A. Huremagic, D. Fifield, D. Rodenburg, R. Hynes, D. Madory, and R. Ensafi, "Network responses to Russia's invasion of Ukraine in 2022: A cautionary tale for Internet freedom," in *USENIX*

Security Symposium. USENIX, 2023. [Online]. Available:
<https://censoredplanet.org/assets/russia-ukraine-invasion.pdf>

[60] SimilarWeb, "Similarweb - access behind-the-scenes analytics for every site online,"
<https://www.similarweb.com>, accessed: 2024-04-14.

[61] Freedom House, "Freedom on the net 2022 – saudi arabia," 2022,
<https://freedomhouse.org/country/saudi-arabia/freedom-net/2022>.

A1 Appendix

Code

<https://github.com/ahmed-mahdi18/Capstone-23-24-Results-Code> - this repo contains all the data that was collected for Ireland and Saudi Arabia as well as the Python programmes that were used for the data extraction.

<https://github.com/citizenlab/test-lists/tree/master/lists> - this is an Ooni repo that includes the test lists for Ireland and Saudi Arabia as well as a legend for what websites are tested in each category.