**The Politics of Routing**

How Malicious Routing Behaviour Correlates with Freedom

Ben Kendall

BSc Hons Computer Science

[DATE]

Declaration

I certify that the material contained in this dissertation is my own work and does not contains unreferenced or unacknowledged material. I also warrant that the above statement applies to the implementation of the project and all associated documentation. Regarding he electronically submitted version of this submitted work, I consent to this being stored electronically and copied for assessment purposes, including the School’s use of plagiarism detection systems in order to check the integrity of assessed work.

I agree to my dissertation being placed in the public domain with my name explicitly included as the author of the work.

(not sure about that last one…)

Date:

Signed:

Project Supervisor: Dr Vasileios Giotsas

**Abstract.**

In an era of hidden mass surveillance programs that violate human rights, it has become unclear whether a person’s real-world individual freedoms translate into the digital world. The design of the Internet as a global communication network without centralized control, creates challenges for governments who try to restrict freedom of expression. Consequently, governments that suppress individual freedoms tend to exercise tight control over Internet communications to reduce their citizen’s freedom of speech. This report will investigate the distinctions between online and offline freedoms, how different metrics of freedom correlate with nation states and Autonomous Systems (ASes) that participate in malicious routing behaviour and the impacts of their imposed policies.

Contents

[Introduction 7](#_Toc3389206)

[1.1 Chapter Overview 7](#_Toc3389207)

[1.2 Motivation for Research 7](#_Toc3389208)

[1.3 The Null Hypothesis 8](#_Toc3389209)

[1.4 Aims and Objectives 8](#_Toc3389210)

[1.5 Report Structure 9](#_Toc3389211)

[Background Research 10](#_Toc3389212)

[2.1 Chapter Overview 10](#_Toc3389213)

[2.2 The Politics of Routing: Investigating the Relationship between AS connectivity and Internet Freedom 10](#_Toc3389214)

[2.3 Nation-State Hegemony in Internet Routing 11](#_Toc3389215)

[2.4 Quantifying Information Exposure in Internet Routing 12](#_Toc3389216)

[2.5 Schengen Routing: A Compliance Analysis 12](#_Toc3389217)

[2.6 Politically Motivated DDoS 13](#_Toc3389218)

[2.7 Related Work: Freedom Indexes 14](#_Toc3389219)

[2.7.1 Freedom of the Net 2018 14](#_Toc3389220)

[2.7.2 Freedom of the Press 2017 14](#_Toc3389221)

[2.7.3 The Human Freedom Index (*updated for 2018*) 14](#_Toc3389222)

[2.7.4 State of World Liberty Index 2018 14](#_Toc3389223)

[2.8 Related Work: Data Logs 15](#_Toc3389224)

[2.8.1 Bad Packets Report (Ongoing) 15](#_Toc3389225)

[2.8.2 Mirai Botnet Activity (Data from Jan 2017 – Jan 2019) 15](#_Toc3389226)

[2.8.3 BGPMON (Ongoing) 15](#_Toc3389227)

[2.8.4 Routing Dependencies (Jan 2019) 16](#_Toc3389228)

[Freedom 17](#_Toc3389229)

[3.1 Chapter Overview 17](#_Toc3389230)

[3.2 Defining Freedom 17](#_Toc3389231)

[3.3 Declining Internet Freedom & China 18](#_Toc3389232)

[3.3.1 Social Engineering as a means to an end 18](#_Toc3389233)

[3.4 Freedom Indexes 20](#_Toc3389234)

[3.4.1 Freedom of the Net 20](#_Toc3389235)

[3.4.2 Freedom of the Press 20](#_Toc3389236)

[3.4.3 The Human Freedom Index 20](#_Toc3389237)

[3.4.4 State of World Liberty 21](#_Toc3389238)

[Malicious Routing 22](#_Toc3389239)

[4.1 Chapter Overview 22](#_Toc3389240)

[4.2 What is Malicious Routing 22](#_Toc3389241)

[4.2.1 Mass Surveillance 22](#_Toc3389242)

[4.2.2 Arrests made over social media posts 23](#_Toc3389243)

[4.3.3 Mirai Botnet IP Addresses 23](#_Toc3389244)

[4.3.4 Mirai-like Packet Signatures 24](#_Toc3389245)

[4.3.5 BGP Hijacks 24](#_Toc3389246)

[4.3.6 Information Exposure 24](#_Toc3389247)

[4.3.8 Feature Weightings 25](#_Toc3389248)

[4.3 Nations and Autonomous Systems 25](#_Toc3389249)

[4.4 Methodology 26](#_Toc3389250)

[4.4.1 Data Discovery and Collection 26](#_Toc3389251)

[4.4.2 Data Reformatting 26](#_Toc3389252)

[4.4.3 Data Processing & Weighting 28](#_Toc3389253)

[4.4.4 Calculating Rankings 28](#_Toc3389254)

[4.4.5 Rankings 29](#_Toc3389255)

[4.5 Correlation Methods 31](#_Toc3389256)

[4.5.1 Types of Correlation 31](#_Toc3389257)

[4.5.2 The FOTN Overlap Problem 31](#_Toc3389258)

[Results 32](#_Toc3389259)

[5.1 Chapter Overview 32](#_Toc3389260)

[5.2 Correlations 32](#_Toc3389261)

[5.2.1 Freedom of the Net Index (With Overlapping Features) 32](#_Toc3389262)

[5.2.2 Freedom of the Net Index (Without Overlapping Features) 33](#_Toc3389263)

[5.2.3 Freedom of the Press Index 34](#_Toc3389264)

[5.2.4 The Human Freedom Index 35](#_Toc3389265)

[5.2.5 State of World Liberty Index 36](#_Toc3389266)

[Findings & Analysis 37](#_Toc3389267)

[6.1 Chapter Overview 37](#_Toc3389268)

[6.2 Findings 37](#_Toc3389269)

[6.2.1 The Overall Trend 37](#_Toc3389270)

[6.2.2 Trend Implications 38](#_Toc3389271)

[6.3 Analysis 39](#_Toc3389272)

[6.3.1 There is a low likelihood that freedom causes a reduction in malicious routing. 39](#_Toc3389273)

[6.3.2 There is a low likelihood that a higher level of freedom and a lower level of malicious routing are connected by the same cause 41](#_Toc3389274)

[6.3.3 higher levels of freedom can sometimes contribute to lower levels of malicious routing (and vice versa) in a loop. 42](#_Toc3389275)

[Conclusions 44](#_Toc3389276)

[7.1 Chapter Overview 44](#_Toc3389277)

[7.2 Conclusions 44](#_Toc3389278)

[7.3 Limitations & Future Work 48](#_Toc3389279)

[7.4 Final Thoughts 49](#_Toc3389280)

[References 50](#_Toc3389281)

Chapter 1

# Introduction

## Chapter Overview

This chapter will explain the reasons for conducting this research and the potential benefits that may come of it. An early approximation of what could be achieved by the report’s aims and objectives will also be laid out. These are dynamic objectives and may be added to or adjusted depending on how the results shape up. The report’s general structure will also be outlined, and each section explained.

## Motivation for Research

Global freedom has been in continual decline for 12 consecutive years [4]. It appears that online freedom is mirroring that trend. The internet, as an openly accessible and free network, affords freedom of expression, information, religion, identity and trade as laid out by the Human Freedom Index 2018 [5]. Therefore, a decline in internet freedom must be the result of restrictive policies imposed by government, ASes or Internet Service Providers (ISPs). In June 2013, the colossal power and capability of intelligence agencies such as the National Security Agency (NSA, US) and their Five-Eyes (FVEY) partners was revealed to the world by Edward Snowden. It was shown how mass surveillance tools such as “PRISM” [1] and “X-KEYSCORE” [2] provided the capability to collect and search through millions of people’s digital lives. It was also revealed that many Autonomous Systems (ASes) and Internet Service Providers (ISPs) such as British Telecom, Vodafone Cable, Verizon Business, Level 3, Viatel, Global Crossing and Interoute were collaborating with projects such as “TEMPORA” [3] which tapped fibre optic cables and intercepted dataflows. It seemed that the online world was far less private or free than previously realised. As well as this, new legislation such as the UK’s Investigatory Powers Act of 2016, is increasing the power of intelligence agencies and further restricting internet and personal freedom. The results of which have seen the UK fall 12 places on the Freedom of the Press Index since 2013 [6].

One of the internet pioneers in the design of TCP/IP architecture, David Clark, explained that the preservation of privacy and the protection of anonymity in public political speech were among the intentions of the end-to-end and layered model of the Internet. It was from this, that stemmed the belief that the network should only implement best-effort packet delivery and should not interfere with the application process [31]. At the same time however, it is expected that internet connected networks should try to limit malicious behaviour. This creates the problem whereby protective mechanisms that sometimes require deep packet inspection and middleboxes are violating internet freedom and architectural principles. It also begs the question as to who defines what needs protecting against and how extreme should these protection measures be. Furthermore, given the proclivity of ISPs under the control of oppressive governments to violate best practices of traffic exchange, it is plausible that these ISPs are also violating the same freedom and architectural principles for cyber-attacks rather than cyber-security. Indeed, both of these areas are part of an ongoing debate among Internet engineers and policy makers that questions the very structure and foundations of the internet itself. This work aims to illuminate this debate with actual data and investigate how these malicious routing policies correlate with an individual’s Internet freedom.

## The Null Hypothesis

*Tighter government control of the internet has no effect in reducing malicious routing.*

The assumption of this null hypothesis is that there is a trade-off whereby a citizen of a given nation sacrifices some level of freedom for better security.

## Aims and Objectives

The aims of this report are therefore as follows:

* Accept or reject the null hypothesis.
* Determine what the term ‘Internet freedom’ entails, and what positive and negative freedoms make up internet freedom.
* Discover the extent to which a correlation exists between a nation’s malicious routing behaviours and the freedom of their citizens.

From the report’s aims, the formal objectives are:

* Form and collect information about each nation’s malicious routing behaviours.
* Analyse and rank nations based on their malicious routing activity.
* Compare each nation’s propensity to engage in malicious routing activity with various measures of their freedom to determine the degree to which they are correlated.
* Consider whether the findings are significant and determine whether propensity to engage in malicious routing activity is a good predictor of freedom and vice versa.

## 1.5 Report Structure

This report will begin with a review of the existing research surrounding malicious routing. This will help gain a more comprehensive understanding of the current academic reasoning, background and issues in this area of research. The existing research associated with defining and measuring freedom will also be investigated to aid with narrowing the selection of freedom indexes to the most relevant for this topic.

A nation’s freedom can be measured in a wide variety of different ways, each resulting in a different freedom index. Before any research is conducted into malicious routing, Chapter 3 of the report will cover the various definitions of freedom and determine which indexes are most applicable for the correlation. Due to the similar features being analysed, the primary index to be considered will be the *Freedom House, Freedom of the Net Index* [9]. It has been predicted that this index will be the most closely correlated index and should help to determine whether the best predictor of internet freedom is a nation’s propensity to conduct malicious routing activity.

Chapters 4 and 5 will make up the main section of the report and will focus on the technical aspects of malicious routing, defining its component features, examining the nations that partake in it and finding any correlations with freedom. This will begin by accurately defining what nation’s features can be considered as part of malicious routing behaviour. Then, a table of nations that partake in malicious routing will be built, holding information about the activities of each nation, as well as a standardised numerical score for each nation based on the amount of malicious routing behaviour that said nation conducts. Once the table has been created and the ranks calculated, it will then be possible to compare malicious routing behaviour with the freedom indexes to determine if any correlations are present. It should be noted that while freedom and internet freedom are closely related, they are considered distinct from one another in the context of this report. The features of each are explored further in Chapter 3.

Depending on the results of the correlations, there may be clear connections between certain kinds of malicious routing behaviour and freedom. Chapters 6 and 7 will examine trends seen across the data as well as stand-out and edge cases. This will contain the ‘findings’ chapter that will careful analyse the results and critically review the methods used. Finally, the report will end with conclusions based on the research and a discussion of the limitations and lessons learnt, as well as any possible future work.

Chapter 2

# Background Research

## 2.1 Chapter Overview

This chapter will briefly summarise what has been learnt from existing research papers about closely related topics. The focus is on malicious routing, however some papers regarding the definitions of freedom have also been included to help form a better understanding of how the varying ideas of freedom can be combined to create an index.

Due to the nature of this project, some information, such as the Snowden leaks which predominately came from news outlets like The Guardian, will be found outside of academic papers. In such a case, only reputable sources of information have been used.

## 2.2 The Politics of Routing: Investigating the Relationship between AS connectivity and Internet Freedom

This study, conducted by Rachee Singh et al, aims to find the relationship between the topological characteristics of Autonomous System connectivity within a certain country, and that country’s Internet Freedom [7]. A network topology graph was constructed for each nation using a Border Gateway Protocol (BGP) path simulator to compute AS paths that are compliant with well-known BGP routing policies, such as valley-free routing. Features of this graph were then extracted and applied to various machine learning techniques that tried to predict the freedom of the press for a given country.

In particular, the paper measures the Internet freedom of a nation using the Freedom House, Freedom of the Press Index [8]. However, while closely related in legislation, the freedom of individuals online is fundamentally different from the freedom of the press, since there are fewer social and economic constraints acting on a single internet user. Where the vast majority of the press operate in a profit-driven environment, in which straying from the overton window could damage sales, an individual internet user operating under perceived anonymity is somewhat relieved from such pressures, and therefore is more likely to share nonconforming or illegal material. Therefore, this report will primary focus on measuring internet freedom with the newer and more closely related Freedom House, Freedom of the Net Index [9]. The actual differences between these indexes will be investigated in more detail later in this report.

The study found that, of all the AS graph’s features, IP density and path length were the best predictors of internet freedom. Their machine learning algorithms were able to predict a nation’s freedom category (Free, Partially Free, Not Free) with a 95% accuracy. This report will follow a similar method when defining the features that malicious routing comprises of. It will also make an interesting comparison to see whether the correlations between AS connectivity and internet freedom mirror the correlations between malicious routing and internet freedom.

## 2.3 Nation-State Hegemony in Internet Routing

This paper researches the extent to which various countries are dependent on the United States and other Western nations to connect to popular internet destinations [10]. In general terms, their results show that underserved regions of the globe are dependent on North America and Europe. In some cases, over 85% of traffic from an independent nation was routed through another country such as the United States. However, they also found that by using Region Aware Networking, they were able to route traffic away from these hegemonies with middle to high success rates.

An example extract of their findings is shown below in table 2.1:



Table 2.1: Percentage of paths terminating in each country

For example, in Brazil, 77% of internet paths terminated in the United States. This is significant for this report as it showed that the level of malicious routing that a nation’s citizens are exposed to extends beyond just the country they live in and includes nearby hegemonic nations where traffic is likely to be routed through. It will be worth considering the impact and measurability of this when defining what constitutes as malicious routing. It also brings up the question of whether nations should be judged based on solely their own policies, or whether neighbouring hegemonic entities need to be taken into consideration.

## 2.4 Quantifying Information Exposure in Internet Routing

In close relation to the last paper, this paper measures the extent to which communications between pairs of countries are exposed to other countries [11]. It was found that, even when sending data between physically adjacent countries, there is a high level of information exposure. Typically, even short international paths route through at least 3 separate countries. It was shown that the better connected a country is, the more their information is exposed and there is a trade-off between internet robustness and information exposure. An important feature when measuring a nation’s internet robustness was the number of independent routing paths to surrounding nations, however it was shown that as the number of paths increases the level of information exposure proportionally increases as well. This is because internet routing is ‘best effort’ and if there are many alternate paths available, then there is a higher chance one of those paths will be chosen.

This is significant for this report because when defining exposure to malicious routing, it will be important to include whether the nation in question has many independent paths. With a greater number of paths to choose from, the likelihood of a packets taking an alternative path increases. If these alternative paths route through a nation that performs malicious routing, then the likelihood that the original data packet could be intercepted or observed increases. To quantify this chance, a function that takes a nation’s proximity with countries that perform malicious routing, and that nation’s number of independent paths will be created. This will then be used as one of the features that makes up malicious routing.

## 2.5 Schengen Routing: A Compliance Analysis

As explored earlier, due to the internet’s best effort approach to packet routing, there is a chance that packets will unnecessarily pass through additional nations before reaching the destination. To try to counteract this, European intelligence agencies developed Schengen Routing. The aim of which is to make sure that when the sender and receiver are both within the Schengen area, their packets are entirely routed within the Schengen area.

This paper [12] investigated weaknesses of Schengen Routing, specifically looking into the compliance of available routes and the amount that any nation uses Schengen compliant routes. The paper found that up to 39.7% of routes within the EU were Schengen compliant and compliance levels for individual countries ranged from 0% to 80%. The choropleth map shown in *figure 2.1* has been extracted from the report and shows compliance levels by nation within the EU.



Figure 2.1: Schengen Routing compliance levels from [12]

This legislation helps reduce the exposure to malicious routing. However, this only affects a small proportion of the world, so when quantifying malicious routing in this report, some nations will be subject to unique features that are not applicable to others. Therefore, it may be necessary to assess the impact of legislation such as this for each nation individually. The implications of this will be assessed in *Chapter 6: Findings and Analysis*.

## 2.6 Politically Motivated DDoS

In this paper, [13] Jose Nazario documents and analyses the reasons behind many of the recent denial of service attacks directed by governments. It covers the nations involved, their expected reasons and the result of the attack. The records from this paper will be used alongside the DDoSDB [14], a database that records global denial of service attacks, to gauge an idea of the nations involved in orchestrating DDoS attacks. This information will partially represent a nation’s propensity to conduct offensive cyber-attacks. The combined dataset will form one of the features of malicious routing that, alongside the Mirai-like signatures [18] cover the attacking side of malicious routing. The victims of the attacks will also be recorded and will be an important feature in determining the exposure of surrounding nations to malicious routing activity.

## 2.7 Related Work: Freedom Indexes

Measurements of freedom will provide a necessary backbone from which any correlations with malicious routing will be found. There are many various measurements of freedom and freedom indexes available online. For this project, the primary freedom index that will be contrasted is the Freedom on the Net Index 2018 [15]. This is because we predict that malicious routing behaviour will correlate closest with internet freedom. Multiple other indexes, as listed below, will also be correlated with malicious routing. If all the indexes show a similar degree of correlation, then there is a lower chance that the results were caused by random luck.

### 2.7.1 Freedom of the Net 2018

[15] This will be the main index that malicious routing in contrasted with. The index covers 65 countries and 87% of the world’s internet users. It’s divided into 3 main features that are analysed for each country individually:

* **Obstacles to access**

Referring to how difficult it is for an individual to get online.

* **Limits on content**

Referring to top down restrictions on what can and cannot be shared online.

* **Violation of user rights**

Measures legal protections and restrictions on online activity.

### 2.7.2 Freedom of the Press 2017

As used in 2.2 [7], (The Politics of Routing) Freedom of the Press [16] has been used in the past as a good measure of individual freedom and freedom of speech. While it has its limitations, as outlined earlier, its rankings are not equal to the Freedom of the Net Index and may provide an additional interesting correlation to explore.

### 2.7.3 The Human Freedom Index (*updated for 2018*)

This index [17] focus’ on overall freedom of a nation’s citizens, taking over 100 features divided into personal and economic freedoms. It will be interesting to compare the results from this very broad index, with the more specific indexes mentioned above. It may also provide an interesting analysis to try to find the single feature from this index that correlates best with malicious routing propensity.

### 2.7.4 State of World Liberty Index 2018

The State of World Liberty index [30] is a combination of freedom indices including 2.7.2 FOTP and 2.7.3 HFI. It aims to create the most accurate relative ranking of nations from the perspective of the economic and social liberties enjoyed by the average citizen within the state.

## 2.8 Related Work: Data Logs

### 2.8.1 Bad Packets Report (Ongoing)

The Bad Packets Report [18] scans for packets that resemble the original Mirai source code. This is possible because Mirai’s TCP sequence numbers will equal the decimal encoded value of the target’s IP address. The report records the source nation, IP and AS of Mirai-like packets and publishes them online. This information will be extrapolated along with Mirai Botnet activity to produce an indicator of the offensive malicious routing taking place inside any given nation. Unfortunately, the publicly accessible historical data only dates back to September 2018, so the comprehensiveness is somewhat limited by its short time frame however it still gives a good indication.

### 2.8.2 Mirai Botnet Activity (Data from Jan 2017 – Jan 2019)

Similarly to 2.8.1, this dataset also comes from the Bad Packets Report [18] and holds information on Mirai activity. Instead of recording instances of Mirai-like source code, it instead records all the IP interfaces observed as part of Mirai-like botnets between 2017-2019. Due to the larger time frame, this dataset is far more comprehensive than 2.8.1 and there are fewer empty areas. It will provide an insight into how many devices in any given nation are part of a Mirai-like botnet.

While this dataset will give a decent indication of the level of Mirai activity in a nation, it should be noted that a significant proportion of recorded Mirai activity is proportional to the amount of cloud services that are hosted in a given nation. This is because cloud hosting platforms are also used by criminals to host malicious software and increase botnet capability. While cloud providers in general try to remove this kind of software, due to the automated nature of the process, it is very easy for criminals to simply set up another ‘service’ under a different alias. On top of this, it would be incorrect to assume that all or even most malicious software found on cloud providers has been targeted towards that provider’s host nation. It is important to note this limitation as there is a distinction between the Mirai-like botnet’s host locations and that botnet’s various targets. This dataset records host locations only.

### 2.8.3 BGPMON (Ongoing)

BGPStream [19] records hijacks, leaks and outages in BGP. The service records the length of the outage and the AS that it occurred in. It’s intended for network engineers to work around outages, however it will also be useful for this report since the affected nations are also recorded. This information will aid in calculating the exposure to malicious routing, although it must be noted that not all BGP outages stem from malicious intent. Unfortunately, there is no practical way to determine which outages are caused by malicious actors, so it is noted that there is a degree of inaccuracy in using this dataset as a measurement for which nation’s ASes are most at risk of hijacks.

### 2.8.4 Routing Dependencies (Jan 2019)

As was explained in the Information Exposure paper [11], highly interconnected ASes have an increased chance of accidently exposing data to other ASes in different nations, even when the two end points are within the same host nation. This can be due to the service being dependent on infrastructure that is physically located in another country. For example, accessing Facebook in Mexico routes requests to Facebook’s data centre in the US, thus exposing the data to the American mass surveillance programs. Whether the data is actually compromised during flight is another matter entirely, however for this paper it will not be considered due to limited research in that area.

Data can also be exposed through the internet’s best effort approach to packet forwarding at the network level, where packets may be forwarded through routes that traverse other nations, despite the two end points being located in the same host nation. For example, data travelling from London to Manchester is frequently routed through Amsterdam because both links are of similar speed and latency. In the past, it has sometimes been favourable for intelligence agencies such as the NSA (US) to route domestic traffic across a nation’s borders. This is because the NSA were only legally allowed to intercept data packets travelling across US borders. A technique called Boomeranging was used to encourage packets to be forwarded across borders even when it was not necessary, thus allowing them to be intercepted.

Whenever data packets cross an international border, there is a chance that they’ll be intercepted. The specific nation they pass through is also important in determining the chance of interception. The Routing Dependencies Dataset gives the fraction of paths that traverse another country, as a list of all nations that data travelled through. To make this very large dataset more useful, a list of nations that perform surveillance of data in any regard will be compiled. For each nation in the Routing Dependencies Dataset, the percentage of paths that traverse nations that perform surveillance will then be compiled. This will create a more compact dataset that records the fraction of paths that traverse through nations who might intercept them. By then summing the paths together, the total exposure to surveillance programs will be calculated for each nation.

Chapter 3

# Freedom

## 3.1 Chapter Overview

This chapter will begin by looking into the problems associated with defining freedom for both a nation and for an individual, as well as the distinction between freedom and internet freedom. The gradual decline in worldwide freedom will then be explored, including a brief overview of the Chinese model of censorship as well as the difficulties this creates when trying to define malicious routing. Following this, the freedom indexes that will be used to correlate with malicious routing will be outlined and their necessity for this report will be explained.

## 3.2 Defining Freedom

While closely related in theory, and considered equal by many freedom indexes, freedom and internet freedom are distinct from one another. They can both be gauged using similar methods and both face the same difficulty in defining whether one has true freedom. Unfortunately, this remains a philosophical question that we still do not have definite answer to. It is such that two individuals in the same situation could give different answers when asked about it. Freedom is as much down to the individual’s perception of whether they are free, as it is to the physical, social and economic constrains acting upon them. When analysing each of these constraints, the distinctions between freedom and internet freedom become apparent. The internet removes many physical and economic constraints on accessibility, however it also holds different social expectations and etiquette. This has resulted in clashes between what is deemed socially acceptable online and what is socially acceptable in real life.

Furthermore, while it is possible to measure physical, social and economic constraints on freedom, it is very difficult to gather data on an individual’s perception of their own freedom. Even if such an extensive dataset were to exist, it may not even be helpful since our own standards and past experience feed into our perception of whether we are free. It is for this reason that citizens in both Pyongyang and San Francisco *might* answer “I am free” to the question of perceived freedom.

This problem translates almost exactly over to the digital world and the question of internet freedom. However, it is far beyond the scope of this report to try to provide correlations with perceived freedom, online or in the real world. Therefore, only the quantitative measures of freedom will be used. By using various individual quantitative measures of freedom such as analysing labour market regulation or the integrity of the legal system, a hypothetical perfectly free state can be envisaged from which all real nations can be compared. This report will use publicly available freedom indexes that rank nations’ freedom scores using quantitative means such as these.

## 3.3 Declining Internet Freedom & China

According to the Freedom of the Net index [15], global Internet Freedom has been in constant decline since 2010. In particular, in 2018 the report points to Facebook’s data exposure scandal involving Cambridge Analytica, security breaches affecting voters in democratic countries and the spread of false rumours and propaganda as reasons for the decline.

The consistently worst abuser of internet freedom is China, but worryingly 36 countries sent representatives to attend Chinese training and seminars on internet and information management. It has also recently been revealed that the Chinese censorship model extends beyond simple mass surveillance and censorship through bans. In her book: “*CENSORED* Distraction and Diversion inside China’s Great Firewall”, author Margaret E. Roberts [20] answers the question “Why do governments attempt to control information when these controls are easily circumvented”? Through past short term ‘crackdowns’ on circumvention methods like VPNs the Chinese government has shown that it has the capacity to make the firewall less permeable but at the current time chooses not to. Instead a more complex marriage between social engineering and computer science is being experimented with.

### 3.3.1 Social Engineering as a means to an end

Is social engineering on the internet a more effective method of control than technologically engineered solutions such as mass surveillance? Generally speaking, when a nation conducts malicious routing such as mass surveillance or censorship, their end goal is an improvement to the perceived stability of their nation. However, attempts at censorship, especially on a large scale such as the internet, are almost always imperfect and the spread of information continues regardless, albeit usually at a slower pace or only within specific clued in circles. Given time however, information (especially in the form of ideas or ideology) can resurge, only to need to be repressed again. The *figure 3.1* below outlines an issue with imperfect censorship over extended periods of time, as experienced by governments around the world:



Figure 3.1: Imperfect censorship model over time

The Chinese government fully understand this problem and rather than try to perfect their censorship technology, they’ve instead begun to explore ways to reduce the coerciveness of the censorship on the general population whilst simultaneously further expanding their control on information.

This is being done through three main methods: fear, flooding and friction [21]. The simplest of which: fear, means to threaten punishments towards those who express dissident opinions online. Recently in China, these laws have become more draconian with citizens being arrested for singing the national anthem in a distorted way on public livestreams [22]. Flooding involves not actually removing the dissident opinions, but instead attempting to drown them out by spamming pro-government or contradictory comments or links. This makes it very difficult and too time consuming for the average onlooker to find the original information that they were looking for. Friction is the most complex of the three and is defined as “increasing the cost, either in time or money of access or spread of information”. This plays off people’s impatience through techniques such as making problematic links take longer to load and slowing down communication apps during protests. The government are aware that the most politically engaged citizens will find ways around this but rely on the fact that the majority of people simply will not be bothered to take the extra steps.

The degree of success of these social engineering policies on population control is far beyond the scope of this paper. It would require an entirely separate study to gauge their impacts and whether they would even be relevant for this report. For this reason, while it is definitely important to note their significance, policies such as these will not individually be assessed in terms of their relation to malicious routing behaviour. Instead, a boolean umbrella category will record whether any citizens have been arrested for political posts to social media.

## 3.4 Freedom Indexes

### 3.4.1 Freedom of the Net

[15] The Freedom House’ Freedom of the Net Index analyses the level of digital authoritarianism across 65 nations. The term digital authoritarianism comprises of the difficulty for an individual to get online, the restrictions on digital content and the restrictions on online activity. The report found that 20% of internet users have free access, 33% have partially free access and 34% do not have free access. One of the benefits of this report is that Freedom House publishes a summary for each country individually, listing some of the principal components that made up their value judgement. Of course, it would be wrong to assume that they had perfect knowledge of all events for each specific nation, however their choice to omit the proposal and passing of articles 11 and 13 for European nations calls into question the bias of the writers. Apart from this, the index does seem to cover most of the major worldwide issues affecting freedom of the net and for this reason will be the primary index used to correlate against a nation’s malicious routing behaviour.

### 3.4.2 Freedom of the Press

[16] The Freedom House’ Freedom of the Press Index, which has been used by papers in the past as a closely related measure of internet freedom will also be used in this report to compare results with past research. As mentioned earlier in the report, while closely related in legislation, the freedom of individuals online is fundamentally different from the freedom of the press, since there are fewer social and economic constraints acting on a single internet user. Where the vast majority of the press operate in a profit-driven environment, in which straying from the overton window could damage sales, an individual internet user operating under perceived anonymity is somewhat relieved from such pressures, and therefore is more likely to share nonconforming or illegal material.

### 3.4.3 The Human Freedom Index

Fraser Institute’s [17] Human Freedom Index is a very comprehensive index recording the level of freedom for 162 nations over more than 100 features. Unlike with the previous two indexes, this index is not specific to a certain kind of freedom, however it shares generalised versions of internet freedom and freedom of the press. It will provide an interesting analysis to compare these features with the dedicated indexes to see if any differences are significant enough to impact our results. It may also be useful to compare the individual features of this freedom index with the propensity to conduct malicious routing index. This will help to determine whether any single feature has a significant correlation greater than the freedom of the net correlation.

### 3.4.4 State of World Liberty

Dr Patrick Rhamey Jr’s State of World Liberty index [30] is actually an amalgamation of many existing indices including civil liberties portions from Freedom House and economic portions from Fraser/Heritage. Both of which are already being used in this report. The intention for this index is mainly for validation against the other indices. If large variations are found between this index and the others, it would raise cause for concern about the reliability of the malicious routing dataset. It should be noted however that this dataset tends to lag behind the others as it waits for updates from all parties before publishing the most recent rankings. Due to this a small differences between this index’s correlation and the others is acceptable.

Furthermore, this index also includes a ranking of the Conservative-Progressive spectrum, detailed by conservative’s high economic liberty contrasted with progressive’s high social liberty. It may provide an interesting insight to see whether the malicious routing index correlates in any way with one side of the Conservative-Progressive spectrum over the other.

Chapter 4

# Malicious Routing

## 4.1 Chapter Overview

This chapter will focus on formally defining malicious routing and determining the important features that will create the malicious routing index. Then, the methodology to compare and correlate the indexes will be explained in depth. The correlations will then be calculated, and the results will be shown. During this chapter some important distinctions will be made, including the relationship between nations and their ASes, as well as the difference between a nations propensity to conduct malicious routing versus the level of exposure that a nation’s citizens have to malicious routing.

## 4.2 What is Malicious Routing

In the most general sense, malicious routing refers to predatory routing policies and behaviours conducted at a national level. The measurement of malicious routing will be from the point of view of the individual living in a given nation and will try to answer the question: “To what degree is an individual citizen of a given nation exposed to malicious routing?”. To answer this, the external forces acting upon the nation, its ASes and its citizens as well as the policies and actions of the nation itself will be used to develop the features of malicious routing as outlined below.

### 4.2.1 Mass Surveillance

For the citizen’s home nation, it is important to initially determine whether any mass surveillance is taking place. The data used to compile this set comes from the Freedom House [23] and partially overlaps with the Freedom of the Net Index. This will be factored in when performing the correlations. This dataset is simply a boolean metric that does not try to quantify the level of surveillance, but rather whether any surveillance is taking place. In this case it is more important to include the government’s intention to conduct mass surveillance than their level of capability. This is because as a country’s wealth increases, their government’s capability to conduct mass surveillance also increases. It shows that, with a few exceptions such as Dem. People’s Rep. of Korea, if a government decides to perform mass surveillance, their capability to conduct it is tied to their wealth and resources.

A better measurement of mass surveillance as part of malicious routing would be to gauge the level of mass surveillance in a given nation and divide by that nation’s wealth. However, this immediately runs into the issue of determining how to give a quantifiable value to mass surveillance. A problem which is made exponentially more difficult due to the secrecy around techniques used in such systems. Many of the sources that give information of that calibre are unreliable or scarce, making this feat too challenging and beyond the scope of this project. It may be an interesting area for further research though.

### 4.2.2 Arrests made over social media posts

Like with 4.2.1, this dataset has been gathered from Freedom House [23] with some updates made from current events. This dataset suffers the most from not being able to prove a direct relation between the arrests made and the techniques used to discover the perpetrator. Similarly to the mass surveillance dataset, this feature focuses on the government’s intent to control internet content within its own borders through arrests made over social media posts from its own citizens. Realistically, it would be too large of a task to quantify all arrests stemming from a social media post for each nation, so again, a boolean system is used to represent either the government’s intent or lack thereof, to control internet content.

The biggest issue with this dataset is that for a given arrest, it is very difficult to determine whether any mass surveillance tools were used to detect the content. For example, in the UK, many counter-terrorism investigations begin from content picked up from surveillance schemes on ‘private’ social media networks [24], some of which inevitably lead to arrests. At the same time however, citizens have also been arrested over out-of-context jokes made on Twitter [25] that referenced the destruction of an airport. This is not an isolated incident in the UK, as between 2010 and 2015, 2500 arrests were made in London alone over social media posts [26]. However, without examining each case individually, it would be impossible to determine which arrests were an over-reach of mass surveillance or an incident reported to the police separately.

### 4.2.3 Mirai Botnet IP Addresses

From the Bad Packets Report [18], a list of IP addresses that are confirmed to be a part of Mirai-like botnets between Jan 2017 and Jan 2019 was compiled. The *Figure 4.1* below shows the number of infected hosts that Bad Packets were able to detect between the 1st January 2019 and the 18th February 2019. This is only a sample of the full data that was collected for this feature.

Figure 4.1: The increase in Bad Packet’s detected Mirai infected hosts over time.

Each IP address was then geolocated and the total number of botnet IPs under each nation was calculated. The total number of IPs for each nation was also compiled and then the total number of botnet IPs was divided by the total number of IPs in each nation to give the proportion of IPs that were a part of a Mirai-like botnet. This data was then normalised to give each nation a value between 0 and 1 which represents that nation’s exposure to Mirai’s infectious behaviour. In more general terms, this feature gives an indication of a nation’s level of exposure to invasive malicious software.

The major limitation of this feature is the impact of Cloud Hosting providers. It is a common practice for attackers to use cloud providers to host their malware [27]. Certain countries such as The US and Netherlands are over represented in their share of the world’s cloud providers. This would mean that such nations may also be over represented in the number of infected hosts that are detected as part of the Bad Packets Report.

### 4.2.4 Mirai-like Packet Signatures

Like with the Mirai botnet IPs, this dataset was also compiled by the Bad Packets Report. The data was received as a list of IPs being targeted by packets that shared the structure and sequence number formatting with those produced by the original Mirai source code. From this, a list of nations and the number of Mirai-like packets that had been targeted at said nation between July 2018 and Dec 2018 was compiled. While the last feature represented Mirai hosting and infection, this feature represents the amount a nation is targeted by Mirai-like attacks. In a more general sense, the feature indicates the quantity of IPs in a given nation that experience malicious attacks. As with the previous feature, this feature is also divided by the number of IPs in the given nation to produce the proportion of IPs that have been targeted, then normalised to make the output more understandable.

Together, the Mirai-like packets feature and the Botnet IPs feature aim to give a representation of external offensive malicious routing affecting the citizens of any given nation.

### 4.2.5 BGP Hijacks

The data for this feature was collected by BGPmon [19]. Their dataset records victim ASes that have experienced BGP hijacks as well as the ASN (Autonomous System Number) of the attacker. In some cases, both the attacker and victim ASN are the same and BGPmon acknowledges that for many of these, the ‘hijack’ was simply a mistake on the part of that particular AS. It must be noted however, that it is not a guarantee that the ‘hijack’ was a mistake when the ASNs are equal.

The data was recorded between the 14/06/2018 to the 31/12/2018. An average of 49 possible hijacks were noted each day. In total, the attacks originated from 6524 unique ASes, but there were only 4169 unique victim ASes. The total number of recorded attacks was 9821. The *Figure 4.2* below shows the number of detected BGP hijacks per day throughout the recorded period.

Figure 4.2: Number of BGP Hijacks for each day in the period 14/06/18 to 31/12/18

A dataset mapping ASNs to nations was used to cross reference with the BGP hijacks dataset to produce a list of the number of times any AS within a given nation has experienced a BGP hijack. This gives an indication of how trustworthy a nation’s ASes are. Since ASes in a given nation operate to an extent under the instruction of that nation’s government (E.g. Boomeranging in the US), by extension this dataset also gives an indication of the degree of top down information control from a given nation’s government.

### 4.2.6 Information Exposure

As explained in 2.8.4, information exposure is calculated by taking the routing dependencies of each nation and comparing them to nations that perform mass surveillance. Due to the nature of the output being a value between 0 and 1, normalisation was not required here. While this feature gives a clear indication of the likelihood that a nation’s data will pass through a nation that performs mass surveillance, it is impossible to determine whether any given packet will be analysed by the surveillance schemes. Due to the high level of secrecy and lack of disclosure it is impossible to determine whether any given nations surveillance techniques perform better than any other nation. While in the real world this is certainly the case, for this model each surveying nation’s capability will be regarded as equal.

The BGP Hijacks feature and the Information Exposure feature together provide an indication of the degree to which a nation’s internal routing structure exposes its citizens to malicious routing.

### 4.2.7 Feature Weightings

Each dataset explained above relates to a single feature of malicious routing. To produce a single quantifiable measure of malicious routing for each nation, each feature will be given a weighting. The weightings will be between 0 and 1 and when added up will equal 1. Then, by multiplying each feature’s value with its relevant weighting and then summing the results together, a normalised measurement will be formed. From this result, the nations can then be ranked appropriately.

The feature weightings will be based upon 3 main parameters.

* Reliability

The feature’s reliability is determined by the source dataset that it originated from. Certain datasets such as Information Exposure were limited by approximations and randomness that impacted its precision and thus, its reliability. Features with low reliability will be given a lower weighting.

* Completeness

Not every dataset held a complete set of values for every nation that is being considered. Where a dataset has some missing values, its overall weighting will also need to be reduced.

* Indicator type

Different features indicate different things about a nation. In general terms, the features concerning Mirai, indicate external malicious threats towards a nation. An individual indicator such as this cannot hold a weighting too high. Therefore, features that indicate similar areas of malicious routing will not be allowed to hold a combined weight greater than 50% of the malicious routing definition.

## 4.3 Nations and Autonomous Systems

Autonomous Systems always have a home nation however can operate across borders in some cases. They are also in most cases forced to abide by, and implement changes based on the laws of the nation in which the primarily operate. A significant proportion of the datasets used in this report record data per AS, not per nation. However, for this report, it would be impractical to analyse each AS so instead ASes will be grouped by their host nation. This allows for citizens to be logically grouped by nation instead of roughly grouped by their use of certain ASes. By extension, this allows exposure to malicious routing to be calculated on a per nation basis. By indexing by nation, comparisons with legislation, policies, ideology and levels of freedom are dramatically simplified. It is important to note however, that this simplification process will come with a minor loss to precision of the exact groups being affected by malicious routing.

## 4.4 Methodology

### 4.4.1 Data Discovery and Collection

The datasets that were used were discovered mainly through background research into the area of interest. Sites such as Google Scholar [28] made the discovery process considerably easier by providing search tools for academic literature. In some cases, full access to the datasets referenced by useful academic literature was not available. In other cases, new data was collected using mirrored techniques and occasionally applications for API keys were necessary for full access. In the case of the Bad Packet’s Mirai Signatures dataset, only partial access was ever obtained. As explained earlier, this, along with the reliability of the data partially helped to form the weightings of each feature of malicious routing.

### 4.4.2 Data Reformatting

In their raw states, most of the datasets that were used did not naturally fit the malicious routing feature that they indicated.

For the Bad Packet’s Mirai Signatures, the raw data simply listed the origin IP address of the packet, the country code, the ASN and the date of discovery. An example extract from the raw data is shown in the *Table 4.1* below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Origin IP Address** | **AS** | **Country Code** | **ASN** | **Date of discovery** |
| 85.94.167.216 | ANDORRA | AD | AS6752 | 2018-10-17 01:42:46 PST |
| 91.187.79.38 | ANDORRA | AD | AS6752 | 2018-11-24 10:57:17 PST |
| 213.132.46.135 | DU-AS1 | AE | AS15802 | 2018-11-14 07:47:24 PST |
| 5.30.101.119 | DU-AS1 | AE | AS15802 | 2018-11-26 18:47:21 PST |

Table 4.1: Raw Bad Packet’s Data Sample

For the Mirai signatures feature, the number of packets originating in each nation was required. This was a simple process that just involved counting each instance of each country code and recording the final result for each nation. This was done by converting the raw data into a csv file and then running a short Python script over it.

For the Bad Packet’s Infected IPs, the raw data held the IP address of the infected device and first and last times that particular IP was detected. Only the IP address was of interest for this report. An extract from the raw data is shown in *Table 4.2* below:

|  |  |  |
| --- | --- | --- |
| **IP Address** | **Date First Seen** | **Date Last Seen** |
| 114.233.132.127 | 22/01/2019 23:25 | 22/01/2019 23:25 |
| 189.209.254.141 | 22/01/2019 23:24 | 22/01/2019 23:24 |
| 223.97.182.144 | 22/01/2019 23:23 | 22/01/2019 23:23 |
| 41.43.154.19 | 22/01/2019 23:23 | 22/01/2019 23:23 |

Table 4.2: Reformatted Bad Packet’s Dataset

In similar fashion to the other Bad Packet’s dataset, the IP addresses found for each nation of interest were counted and totalled to produce the total amount of detected infected devices in each nation. Each IP address also had to be geolocated. This was done with Maxmind’s Python geoip2 library [29].

The BGP Hijacks raw dataset had a similar structure to the last. It recorded the ASN of the victim AS and attacking AS as well as the start and end times of the hijacking. An extract from the raw data is shown below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Attacker ASN** | **Victim ASN** | **Start Time** | **End Time** |
| 265790 | 265780 | 11/11/2018 14:34 | 11/11/2018 17:07 |
| 43513 | 34434 | 11/11/2018 12:22 | 11/11/2018 13:09 |
| 204287 | 34434 | 11/11/2018 12:22 | 11/11/2018 12:34 |
| 39288 | 1299 | 10/11/2018 20:16 | 10/11/2018 20:45 |

Table 4.3: BGP Hijacks Raw Data Sample

The ASNs were first mapped to their respective nations, then the number of times a nation appeared in the dataset was recorded. This was calculated in Microsoft Excel by concatenating the original dataset with another dataset mapping ASNs to nations.

The most complex feature to reformat was the information exposure. The raw dataset gave a nation’s country code and the percentage of that nation’s data that travelled through other nations as a tab delimited string by nation.

An example record is shown below:

|  |
| --- |
| AD|ES,0.257450628366 FR,0.155834829443 GB,0.381328545781 US,0.205385996409 |

For the example shown above, ‘AD’ refers to the country Andorra. ‘ES,0.2574…’ means that roughly 25% of Andorran internet traffic routed through Spain. ‘GB,0.3813…’ means that roughly another 38% of Andorran internet traffic routed through Great Britain. This repeats for every nation that traffic was routed through for the duration of the test.

To make the dataset usable, the string format was first converted into a CSV file. Then a new csv file was created listing all the nations that performed some kind of surveillance on internet traffic. The original file was then cross referenced with the new csv file to remove nations that did not perform internet surveillance. For example, hypothetically, imagine Spain had no mass surveillance program.

For the example string from above, this would produce a new string without Spain’s routed traffic being recorded as shown below:

|  |
| --- |
| AD|FR,0.155834829443 GB,0.381328545781 US,0.205385996409 |

The remaining percentage values were then summed together to give a total value indicating exposure to nations performing internet surveillance. The operation for our example, Andorra, is shown below:

*FR 0.1558… + GB 0.3813… + US 0.2053… = 0.7424*

Therefore, for this simplified example, Andorra would have an internet exposure value of 0.7424.

It should also be noted that the raw datasets used different naming conventions for nations. For example, some datasets referred to Libya as Libya, while others used its full name, Libyan Arab Jamahiriya. For all the nations in each dataset, the full names were manually replaced to avoid any encoding and look-up issued.

### 4.4.3 Data Processing & Weighting

Where necessary, the features were normalised to make further processing simpler. Furthermore, for the Mirai-like features, the data was divided by the total number of IPs in each nation. This meant that countries with exceptionally large amounts of devices were not over-represented by this metric.

Each feature’s weighting was produced using methods explained in 4.3.8. The *Table 4.4* below shows each feature and its corresponding weighting. The sum total of the weightings equals 1.

|  |  |
| --- | --- |
| **Feature** | **Weight** |
| Mirai Botnet IPs | 0.3 |
| Mirai-like Signatures | 0.2 |
| Information Exposure | 0.2 |
| BGP Hijacks | 0.1 |
| Any Surveillance | 0.1 |
| Citizens Arrested | 0.1 |
|  |  |
| Total: | 1.0 |

Table 4.4: Weighted Values of each feature

### 4.4.4 Calculating Rankings

For each nation, each feature was multiplied by its respective weighting and the resulting values added together to get the final malicious routing indicator. An example of the feature’s values are shown in *Table 4.5* and the process is shown in *Table 4.6* below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Country** | **Mirai Botnet IPs** | **Mirai-Like Signatures** | **Information Exposure** | **BGP Hijacks** | **Any Internet/ Telecom surveillance** | **Citizens arrested over social media posts** |
| Egypt | 0.48327128 | 1 | 0.998919357 | 0.004754358 | 1 | 1 |

Table 4.5: Example feature values for the nation of Egypt.

For the example country, Egypt:

|  |  |
| --- | --- |
| ((0.48327128\*0.3) +  (1\*0.2) +  (0.998919357\*0.2) +  (0.004754358\*0.1) +  (1\*0.1) +  (1\*0.1)) = 0.745240691 | ((**Mirai Botnet IPs** \*0.3) +  (**Mirai-Like Signatures** \*0.2) +  (**Information Exposure** \*0.2) +  (**BGP Hijacks** \*0.1) +  (S**urveillance** \*0.1) +  (**Citizens arrested** \*0.1)) = … |

Table 4.6: Example malicious routing calculation

As shown in *Table 4.6,* the overall malicious routing indicator for Egypt is 0.745.

By performing this operation for every nation, it was possible to rank them all based on their malicious routing activity.

### 4.4.5 Rankings

The following *Table 4.7* shows each nation for which data was collected and its respective malicious routing indicator as well as the corresponding ranking.

|  |  |  |
| --- | --- | --- |
| **Country Name** | **Malicious Routing** | **Rank** |
| Angola | 0.185968261 | 9 |
| Argentina | 0.169791943 | 7 |
| Armenia | 0.225652715 | 19 |
| Australia | 0.401121309 | 45 |
| Azerbaijan | 0.314731064 | 32 |
| Bahrain | 0.407510384 | 47 |
| Bangladesh | 0.376556173 | 39 |
| Belarus | 0.263363177 | 23 |
| Brazil | 0.272702143 | 24 |
| Cambodia | 0.577821791 | 61 |
| Canada | 0.395792975 | 42 |
| China | 0.473346679 | 55 |
| Colombia | 0.212273654 | 15 |
| Ecuador | 0.215014948 | 17 |
| Egypt | 0.745240691 | 64 |
| Estonia | 0.090038403 | 3 |
| Ethiopia | 0.471851623 | 54 |
| France | 0.390895909 | 40 |
| Gambia | 0.546281071 | 59 |
| Georgia | 0.307816825 | 29 |
| Germany | 0.396285233 | 43 |
| Hungary | 0.193486306 | 11 |
| Iceland | 0.068743439 | 2 |
| India | 0.514315352 | 57 |
| Indonesia | 0.467152668 | 53 |
| Islamic Republic of Iran | 0.466162958 | 52 |
| Italy | 0.27398504 | 25 |
| Japan | 0.316644948 | 33 |
| Jordan | 0.530238675 | 58 |
| Kazakhstan | 0.311651078 | 31 |
| Kenya | 0.167994867 | 5 |
| Korea, Republic of | 0.428888297 | 49 |
| Kyrgyzstan | 0.196864183 | 13 |
| Lebanon | 0.309386041 | 30 |
| Libyan Arab Jamahiriya | 0.180181642 | 8 |
| Malawi | 0.050203172 | 1 |
| Malaysia | 0.32395257 | 34 |
| Mexico | 0.357246062 | 37 |
| Morocco | 0.25816448 | 22 |
| Myanmar | 0.327889242 | 35 |
| Nigeria | 0.283522179 | 26 |
| Pakistan | 0.298944596 | 28 |
| Philippines | 0.231910846 | 20 |
| Russian Federation | 0.712097116 | 63 |
| Rwanda | 0.222863317 | 18 |
| Saudi Arabia | 0.404935054 | 46 |
| Singapore | 0.398442203 | 44 |
| South Africa | 0.208803112 | 14 |
| Sri Lanka | 0.235564783 | 21 |
| Sudan | 0.212953447 | 16 |
| Syrian Arab Republic | 0.434124998 | 51 |
| Thailand | 0.48354675 | 56 |
| Tunisia | 0.429311985 | 50 |
| Turkey | 0.54925085 | 60 |
| Uganda | 0.189022844 | 10 |
| Ukraine | 0.340591358 | 36 |
| United Arab Emirates | 0.424826499 | 48 |
| United Kingdom | 0.394206337 | 41 |
| United States | 0.376459727 | 38 |
| Uzbekistan | 0.168839991 | 6 |
| Venezuela | 0.29084137 | 27 |
| Vietnam | 0.63712447 | 62 |
| Zambia | 0.153800277 | 4 |
| Zimbabwe | 0.196831159 | 12 |

Table 4.7: Malicious Routing Values and Rankings for each nation.

## 4.5 Correlation Methods

### 4.5.1 Types of Correlation

To determine whether any correlations exist, both Spearman’s rank correlation and Pearson’s correlation will be used. Spearman’s rank correlation is computed on rank and therefore depicts a monotonic relationship where the variable’s relationship changes in the same direction, but not always at the same rate. On the other hand, Pearson’s correlation is calculated on true values and instead gives information about whether linear variables increase or decrease at similar rates. To explore this dataset fully both Spearman’s and Pearson’s correlations will be calculated since the relationship between them also gives some additional information about the data.

Both correlation types were calculated using Microsoft Excel. For Spearman’s Rank, the difference between the ranks was squared and summed for each index. For Pearson’s, Excel’s inbuilt ‘CORREL’ function was used.

### 4.5.2 The FOTN Overlap Problem

As briefly mentioned when covering the freedom indexes in an earlier chapter, there is a feature of malicious routing that is shared with the Freedom of the Net index. While Freedom House do not publish their raw datasets, many of their analyses at the individual country level use a feature dubbed “Blogger or ICT user arrested, imprisoned, or in prolonged detention for political or social content”. This description is very similar to the “Citizens arrested over social media posts” feature of malicious routing and it is plausible that similar sources were used to compile both features. The exact methods that Freedom House used to gather data for this feature are unknown, so it is impossible to accurately determine the degree to which these features overlap. However, in the interest of reliability, when analysing the FOTN index, the normal malicious routing definition will be used and then compared to a recompiled malicious routing definition without the “Citizens arrested over social media posts” feature. If the difference between the two correlations equals the weighting of the feature, then it likely that original datasets were very similar if not the same. Assuming the correlation holds for both datasets, then no further action will need to be taken as the relationship has maintained consistency.

Chapter 5

# Results

## 5.1 Chapter Overview

In this chapter, the results from calculating Spearman’s ranks and Pearson’s correlations will be given and explained. Each index that is being compared will warrant its own section where the specific correlations will be detailed. Scatter plots will also be included for each index, where a single point on the plot represents a given nation. The picture of the results as a whole will not be considered here, but preliminary analysis of each individual result will be explored in preparation for in-depth analysis in the following chapter.

## 5.2 Correlations

### 5.2.1 Freedom of the Net Index (With Overlapping Features)

|  |  |
| --- | --- |
| Spearman’s Rank Correlation Coefficient | 0.412774725 |
| Pearson’s Correlation Coefficient | 0.457331 |

### 5.2.2 Freedom of the Net Index (Without Overlapping Features)

|  |  |
| --- | --- |
| Spearman’s Rank Correlation Coefficient | 0.25 |
| Pearson’s Correlation Coefficient | 0.305997 |

When the overlapping feature: ‘citizens arrested over social media posts’ is removed from the malicious routing definition, the correlation coefficient for both Spearman’s rank and Pearson’s correlations shows a decrease roughly proportional to the weighting of the removed feature. This indicates that removed feature was likely to have been very closely related to the Freedom of the Net Index. Since the drop shown between 5.2.1 and 5.2.2 is relatively small, the overall interpretation of these correlations essentially remains the same.

The correlation shows a low positive trend across both Spearman’s and Pearson’s with and without the overlapping feature. In both cases, Pearson’s correlation was slightly stronger than Spearman’s correlation. This shows that the rate at which the variables increased held a stronger correlation than their ordering (rank). For this index, it means that as a nation’s level of malicious routing increases, it is likely that the internet freedom in said nation also decreases.

### 5.2.3 Freedom of the Press Index

|  |  |
| --- | --- |
| Spearman’s Rank Correlation Coefficient | 0.267811355 |
| Pearson’s Correlation Coefficient | 0.263051 |

For both Spearman’s and Pearson’s correlations, a low positive relationship was observed. Where a low FOTP value relates to a freer press, this correlation shows that as a nation’s level of malicious routing increases, it is likely that their freedom of the press decreases.

### 5.2.4 The Human Freedom Index

|  |  |
| --- | --- |
| Spearman’s Rank Correlation Coefficient | 0.056663141 |
| Pearson’s Correlation Coefficient | -0.12923 |

Spearman’s rank showed almost no correlation between a nation’s level of malicious routing and its rank on the Human Freedom Index. For Pearson’s correlation, it must be noted that the negative value attained is due to the HFI’s real values, where high values representing high freedom levels and low values representing low freedom levels. Whereas low malicious routing values represent low levels of malicious activity and vice versa. Therefore, Pearson’s correlation coefficient of -0.12 indicates that there is a very weak relationship between increasing malicious routing in a nation and a decrease in freedom.

### 5.2.5 State of World Liberty Index

|  |  |
| --- | --- |
| Spearman’s Rank Correlation Coefficient | 0.126272081 |
| Pearson’s Correlation Coefficient | -0.15598 |

This index also showed a very weak positive correlation with malicious routing across both Spearman’s and Pearson’s correlation coefficients. Similarly, with 5.2.4, the negative value in Pearson’s correlation is due to the inverse high-low meanings for each dataset. This is because a high liberty value represents a nation with more freedom, whereas a high malicious routing value represents a nation with more malicious routing. They both, however, indicate a low likelihood that as the level of malicious routing increases in a nation, the level of liberty in said nation may also decrease.

Chapter 6

# Findings & Analysis

## 6.1 Chapter Overview

This chapter will give a generalised view of the study’s findings and the overall trends. The results will then be analysed with a focus on attempting to determine the reasons for the trends outlined in chapter 5. This will include an analysis of differences between this study’s results and any trends consistent across the freedom indexes. Any known issues with the study will also be noted and their effects on the results will be theorised.

## 6.2 Findings

### 6.2.1 The Overall Trend

Each freedom index held some level of positive correlation, however the real values varied between 0.06 and 0.45. However, when adjusted with similar features removed, the real values were between 0.06 and 0.31. After adjusting for negative values produced by Pearson’s value-based correlation, in general the correlation coefficients for each correlation method produced similar values to one another. The *Figure 6.1* below plots Spearman’s rank coefficient values for each Freedom Index where a value of 1 represents perfectly correlated data and a value of 0 represents data with no correlation.

Figure 6.1: Spearman’s Correlation Coefficients for each Index

It is clear from this that the FOTN and FOTP indexes show a weak positive correlation and the SOWL and HFI indexes show almost no correlation. The following *Figure 6.2* gives the same plot, but for Pearson’s coefficient values instead. As shown, the correlations mirror that of Spearman’s rank, with low-to-no correlation shown in SOWL and HFI, and a small positive trend for FOTN and FOTP.

Figure 6.2: Pearson’s Correlation Coefficients for each Index

### 6.2.2 Trend Implications

The general trend across all the indexes is a weak positive correlation. This means that *as a nation’s level of freedom increases, there is a small reduction in the likelihood that a citizen in that nation will be exposed to malicious routing*. However, since it is always wrong to assume that correlation equals causation, this result can potentially mean one of more of the following options:

1. That the result is pure coincidence.
2. That there is a low likelihood that freedom causes a reduction in malicious routing.
3. That there is a low likelihood that malicious routing causes a reduction in freedom.
4. That there is a low likelihood that a higher level of freedom and a lower level of malicious routing are connected by the same cause.
5. That higher levels of freedom can sometimes contribute to lower levels of malicious routing (and vice versa) in a loop.

While the results between indexes were not identical, they did all show the roughly the same general trend, albeit at various strengths of correlation. Due to this consistency the chance that these results were due to pure coincidence is low enough to rule out that option entirely. This leaves options 2 through 5 as possible explanations for the results. Since each freedom index covers a slightly different type of freedom, it is possible that each index’s correlation may be comprised of different explanations. As a basic example, due to the Freedom of the Net Index’s shared feature, it may appear that higher freedom levels contribute to lower malicious routing levels in loop. However, this is probably due to the shared overlapping definition and in reality, is not applicable. In the next section (6.3), this will be analysed in more detail for each index individually and the possible real-world explanations for the trends will be explored.

## 6.3 Analysis

This section will explore in greater detail each possible cause of the correlations as outlined in the previous chapter. Each cause will be posed as a statement and the arguments either way will be laid out, with examples provided as evidence where appropriate.

### 6.3.1 There is a low likelihood that freedom causes a reduction in malicious routing.

In full, this statement asserts that where there are higher levels of freedom, there is a greater chance for there to be lower levels of malicious routing due to said freedom. This statement is directly contradictory to the next possible cause which states: “There is a low likelihood that malicious routing causes a reduction in freedom”. Due to this clear link, these potential causes will be analysed together in this section. It must also be noted that it is possible for both potential causes to be incorrect for any given nation.

When these potential causes are compared, it begs the simple question: “Does high freedom cause low malicious routing or does low malicious routing cause high freedom?”. This initially resembles a classic ‘chicken and the egg’ problem, except that the concept of freedom far outdates the concept of malicious routing and thus suggests that high freedom would result in low malicious routing. This makes sense as by the Human Freedom Index’s standard, a socially and economically free nation that exhibits a libertarian stance on control of its citizens should not try to control the online space beyond law enforcement through the use of malicious routing. However, this model is based on the following flawed assumptions:

1. The nation existed in its current state at the conception of malicious routing.
2. The nation was free at the conception of malicious routing.
3. The nation’s government has not changed its policies over time. (e.g. becoming more authoritarian.)
4. The nation had the capability to perform malicious routing at its conception.

Essentially, these flaws can be summarised with the natural change of government policy, technology and public opinion over time. However, the rate of progression of these factors is significant in determining which potential cause is the most appropriate. Without radical means of changing these factors such as the end of war or a revolution, most nation states change very slowly over time. In democratic nations this is usually a slower process and typically mirrors ideologies popular with the general public, however in single party states or dictatorships, changes in the level of freedom or malicious routing can occur at a faster rate.

As technology has improved over time and standards have been developed, it has become increasingly easier for a nation to conduct some degree of malicious routing. This means that at the conception of many of the techniques used in malicious routing, less developed nations would not have been able to implement said techniques whether their governments wanted to or not. Furthermore, in the current day, very poorly developed nations, or nations with collapsed governments may still be unable to implement any malicious routing techniques due to technological or organisational limitations rather than desire. For example, Malawi, which in April 2018 saw the completion of a new national fibre-optic backbone from Huawei [32], was ranked 1st for least malicious routing across all weighted features. According to Freedom House [33], Malawi’s President, Peter Mutharika warned that punitive action would be taken against online hate speech, however, to date there have been no reports of citizens being arrested for online activities and there are currently no restrictions on connectivity and accessible content except from child pornography. This, alongside almost no recorded Mirai activity and very low levels of information exposure have given Malawi its respective ranking. It should be noted however that despite the new fibre-optic network, internet access still remains out of reach for the majority of Malawi’s population. The contrast between Peter Mutharika’s comments and the actual level of malicious routing, suggest that for Malawi, there is no causal relationship between freedom and malicious routing or vice versa.

However, while there may be no causal relationship for Malawi, this does not hold true across all other nations. For example, Iceland, the nation with the 2nd lowest recorded malicious routing activity does not suffer from a low level of development like Malawi. Instead Iceland ranks 6th on the Human Development Index [34] and 5th by GDP per Capita according to the World Bank [35]. Politically, Iceland has had a very stable multi-party democratic history and currently the two most popular parties generally hold liberal conservative and centrist libertarian beliefs. Icelandic citizens have experienced a consistently high level of social and economic freedom since the founding of the Republic of Iceland in 1944. This level of individual freedom has been maintained throughout the digital revolution and despite some known cooperation with the Five-Eyes Network, the overall level of malicious routing in Iceland remains remarkably low. This appears to suggest that Iceland’s principled stance on social and economic freedom has translated over to the digital world and thus would imply that freedom does cause a reduction in malicious routing. However, there are definitely other possible explanations such as alternative issues that the Icelandic government has decided to focus on instead. This would be impossible to determine at this stage and would require further detailed research which is beyond the scope of this project.

Iceland appears to be quite anomalous however, as it shares its status as a developed country with only 4 other nations in the top 20 for lowest malicious routing level. The remaining 16 nations are far more poorly developed and include countries such as Kenya, Zambia, Libya and Kyrgyzstan. Furthermore, countries at similar levels of development to Iceland, such as Great Britain, France and Germany, all exhibit the inverse relationship between freedom and malicious routing. On top of this, most of these other highly developed nations have, like Iceland, maintained stable governments over the last 60+ years that rarely reach the extremes of the political spectrum. The vast majority of them have also not seen large changes in their respective levels social and economic freedom. Despite this however, Iceland remains exceptional in this group as one of the few nations to have a high level of overall freedom across the 4 freedom indexes, be highly developed and not to also have a high level of malicious routing. Therefore, overall it would seem that in the majority of cases, freedom does **not** cause a reduction in malicious routing.

For the inverse of this potential cause to hold true, a low level of malicious routing would have to cause a high level of freedom. Due to the relatively new concepts surrounding malicious routing, the time scale in which an event such as this could’ve taken place is far shorter than for the alternative causes. Theoretically speaking however, low levels of malicious routing in a nation could improve the state of communication between citizens and thus, give a wide-reaching platform to new ideas that may improve the level of social and economic freedom in that same nation. It should also be noted that the spread of misinformation can also occur through the same process and potentially hinder or even damage the fabric of society. Comparing the benefits and downsides of a free internet’s ability to spread both good and bad ideas would provide an interesting and very challenging area of additional research. Overall however, due to the lack of real world examples, the theory that a ‘low level of malicious routing could cause a high level of freedom’ will be disregarded.

### 6.3.2 There is a low likelihood that a higher level of freedom and a lower level of malicious routing are connected by the same cause

To determine whether this cause has any applicability, the causes of freedom must be first defined. Unfortunately, as with the actual definition of freedom, the cause of freedom also remains a highly debated area with no formal definition. Furthermore, as with many historical causal relationships, it can be very difficult to determine where the absolute root cause originated from. For example, it could be stated that a certain aspect of freedom was caused by the writing of a law that guaranteed protection of that freedom. However, it could then be asked: ‘*why was the law was written to begin with?*’. Was the law written in response to a reduction in that kind of freedom? If so, then that freedom must have existed in some form before the law was written. These chains of questions can quickly become too complex to handle and many do not end with a definite answer. Furthermore, there are different definitions of freedom and separate laws to protect them for every nation individually. For some nations, such as Great Britain, an unwritten constitution defines freedom amongst many other things as being comprised of Acts of Parliament, court judgments and conventions, further complicating this seemingly impossible task.

Rather than examining individual nations for their root causes of freedom, known overall trends will be considered instead. An example of which shows that there is a positive significant relationship between economic growth and economic freedom. It was theorised that “if liberty is a normal good, an increase in wealth may spur the demand for freedom” – Jac Heckelman [36]. If an increase in wealth may subsequently increase freedom, then for this cause to hold true, then an increase in a nation’s wealth should also cause a reduction in malicious routing. To test this, the Spearman’s rank correlation coefficient between a nation’s GDP and its malicious routing was calculated. (This is intentionally a rough calculation as there are many different rankings for nations based on wealth. A separate, more in-depth study would be required to investigate the correlation between wealth and malicious routing in full.) The results of this showed a moderate negative correlation coefficient of -0.435. This means that as a nation’s GDP increased, the level of malicious routing tended to also increase. This is the direct opposite of what the cause predicted and suggests in the terms of economic freedom, that there is a causal relationship between higher GDP and higher malicious routing. This means that a higher level of freedom and a lower level of malicious routing are, in fact, unlikely to be connected by the same cause.

Although, as explained in the first paragraph, this is simply one possible cause of a single kind of freedom. There are also clear links between ideology and levels of freedom which will be left unexplored by this section but may hold their own interesting correlations if taken up for further research

### 6.3.3 higher levels of freedom can sometimes contribute to lower levels of malicious routing (and vice versa) in a loop.

While the citizens of a free nation would not normally consider the state of their own personal freedom, they do come to expect that level of freedom across all aspects of life. It is usually very noticeable to a nation’s population, when a freedom previously enjoyed is lost. If this is the case, then it should be true that citizens with higher levels of freedom should expect a low level of malicious routing and thus a high level of internet freedom. However, amongst western nations such as France, Germany and the Five-Eyes countries, this does not appear to be the case.

Following the Snowden revelations, there was a remarkably low public outcry [37] and polls showed divided public opinion 41 against, to 56 in favour (Pew research centre) on whether the surveillance was an acceptable way for the government to investigate terrorism. The reasons for this are varied but tend to focus upon a general lack of public understanding about the surveillance practices and the indirect and usually unseen impact on the lives of an average citizen. It is important to remember at this point, that an individual’s perception on whether they are free is not insignificant in determining the level of personal freedom for a given citizen (as discussed in Chapter 3).

While important, mass surveillance is only one metric that helps to determine a nation’s malicious routing. Another key component is the exposure to the effects of malicious software such as the Mirai botnet. For the most part, a nation’s government can’t do a great deal to combat threats such as these. In developed nations with defence agencies, for some public high-profile cyber-security cases the relevant agency may get involved, however an average citizen relies more upon their own preventative actions, and the mitigating steps taken from whomever they purchased the vulnerable technology from. More recent legislation such as the EU’s GDPR is encouraging private companies to better protect their customer’s personal information, however the legislators are still playing catch-up when it comes to handling issues such as poor device security as standard which was partly how botnets based on Mirai were able to spread so rapidly. Generally speaking, this kind of legislation is aimed at protecting citizens under its jurisdiction but also tends to be reactionary to recent current affairs. It may be plausible to claim that the high level of freedom experienced by the legislators creates the expectation that the same level of freedom should be at least maintained in the new laws passed. However, for this cause to hold true, the lower level of malicious routing should also contribute back to creating a higher level of freedom. In the mostly reactionary system that currently exists in many of these developed nations, a higher level of malicious routing is required before action is taken against it. Furthermore, this doesn’t even consider the actual effectiveness of the legislation and also assumes that it performs exactly as it was intended. Therefore, overall it seems that this cause is not an applicable explanation for the trends observed in the previous chapter.

It should also be noted that for this cause, many of the more poorly developed nations that have exhibited very low levels of malicious routing such as Kenya, Zambia, Libya also tend to rank middling to low across many of the freedom indexes. Due to this, there is no way that higher levels of freedom could possibly contribute to lower levels of malicious routing in these specific countries.

Chapter 7

# Conclusions

## 7.1 Chapter Overview

This chapter will review the study as a whole and attempt to draw final conclusions based on the discussions and issues raised in Chapter 6, Analysis. Finally, the limitations of the report will be reflected upon and any potential future work to take this study further will be briefly outlined.

## 7.2 Conclusions

All of this report’s initial aims and objectives have successfully been met. The resulting weak positive correlations from this study are not strong or consistent enough to reject the null hypothesis. Therefore, the null hypothesis stating that **tighter government control of the internet has no effect in reducing malicious routing,** will be accepted. Furthermore, the correlations between freedom and malicious routing do not lend themselves to a single, simple overarching trend or explanation. Every nation is unique in its own cultural values and interpretation of freedom. The overall correlation across all nations may indicate that it is slightly more likely for a free nation to exhibit lower malicious routing, but the reasons behind this are slightly different for each nation. Having said that, most of the countries that were analysed could be placed into one of the following four groupings:

1. Developing nations

Medium – Low freedom level

Poor internet connectivity

Low malicious routing

While many citizens in this group’s nations have limited or no internet access, the few citizens that do have internet access enjoy low levels of malicious routing. Government in these nations tend to have little-to-no surveillance of the online activity of their citizens are arrests based on social media posts are very rare. Due to the low level of internet infrastructure in these nations, they are also unlikely targets for cyber-attacks such those characterised by Mirai activity. Their level of internet exposure can be quite high as, is in some cases they can become reliant on services hosted in the closed developed nation. This makes their overall malicious routing level relatively low.

Example of these nations include, Malawi, Zambia, Kenya, Uzbekistan, Libya, Angola, Uganda, Zimbabwe and Sudan. All of which are in the top 15 nations by lowest malicious routing level.

1. Developing – Developed nations

Low – Very low freedom level

Medium – High internet connectivity

High malicious routing

This group of nations tends to have an oppressive government and a low level of social, economic and personal freedom. These nations can have a respectable wealth behind them and also tend to have a reasonable level of internet connectivity, especially in their cities. Alongside this, these nations also exhibit a high level of malicious routing due to high Mirai activity, high information exposure and some form of mass surveillance.

Examples include Russia, China, Jordan, Iran, Syria, UAE, Saudi Arabia, Bahrain and Egypt.

1. Developed nations

High freedom level

High internet connectivity

Medium – High malicious routing

These nations tend to be more highly developed and generally have a high level of freedom across most of the freedom indexes. They are very technologically developed and have an extremely high level of internet connectivity. (95%+ of citizens are online). However, they also tend to have comprehensive mass surveillance programs and high levels of internet exposure and Mirai activity. This gives them a medium to high level of malicious routing.

Examples include Australia, USA, Germany, UK, France, Japan, Canada and South Korea.

1. Developed nations

High freedom level

High internet connectivity

Low malicious routing

These nations have similar characteristics to group 3, however they differ in having a relatively low level of malicious routing. Their information exposure can still be quite high, however their exposure to Mirai activity is usually very low and they have few or no surveillance programs. There are far fewer nations in this group than in the others.

Examples include Iceland, Estonia, Argentina, Hungary and South Africa.

These 4 groups are not all encompassing and some nations fall outside or between groups. It would require a separate investigation into each of these nations to determine the cause of their malicious routing level. At a glance however, nations such as Venezuela stand out with medium levels of malicious routing but very low levels of freedom. Sri Lanka also stands out due to their high level of information exposure, which is likely due to its proximity and reliance on India for telecommunication services.

China ranked only 11th for most malicious routing which seems anomalous considering its well documented surveillance state and highly sophisticated content filter. On top of this, most of the freedom indexes rank it in the top 2 least free countries. There are 2 reasons why China was ranked surprisingly low on this list; its large number of internet connected devices, and the inability of this study to quantify the mass surveillance and citizens arrested features. China’s first tier cities such as Beijing, Guangzhou, Shanghai, Shenzhen and Tianjin are very technologically developed with a very high level of internet connectivity and the largest number of IP addresses in one area. It is likely that almost all of the Mirai activity detected in China was focused around these cities. It is estimated that only about 1/10th of China’s total population live in 1st tier cities. While the rest of China does have varying levels of internet connectivity, they are usually far inferior to what is available in these cities.

When calculating the Mirai activity features, the activity level values were divided by the total number of IPs in each nation respectively. One of the knock-on effects of this is that in nations with large numbers of IPs outside of the public reach (E.g. in government or corporate data centres), the impact Mirai on a given citizen is greatly diminished. Sometimes, the number of ASes in a given nation is a better indicator of the quality of the internet in that nation. Due to this, when calculating Mirai activity levels, it may have been better to instead divide by the number of ASes in the given nation, rather than the number of IPs. Since citizens of a nation are reliant on their ASes for internet connectivity, this may better represent the impact of Mirai on a given citizen. To test whether this would drastically alter the nation’s rankings, the ‘Mirai botnets detected’ and ‘Mirai signatures detected’ features were re-calculated with division by the number of ASes in a given nation. The result of this saw nations like China, who exhibited a large amount of Mirai activity across fewer ASes rise dramatically in the ranking. After the modification, China now ranked 2nd for most malicious routing exposure. However, the overall correlation coefficients did not change very much and still showed a weak positive trend across all indexes as shown in *figure 7.1* below:

Figure 7.1: Graph showing Spearman’s Rank correlation coefficients after being adjusted for division by total number of ASes.

Furthermore, while for China, there is lots of evidence detailing high level of mass surveillance and many arrests over social media posts, this kind of information is not as well documented for other nations. As well as this, much of the available information is not backed up by reliable sources and evidence. Therefore, the boolean metric was used instead of a quantifiable one. Because of this, China’s level of mass surveillance appears equivalent to any other nations that performs mass surveillance of any kind. Obviously, in the real world this is not the case and this flaw in the study has been noted.

As shown in the investigation between AS connectivity and Internet Freedom [7], AS path length and IP density were accurate (95%+) predictors of Internet Freedom (using the FOTP index). The results of this study have not come close to mirroring that correlation, as was initially predicted. While this study’s results trended in a similar positive direction, its weakness was such that it could not be used as a reliable predictor of Internet Freedom or vice versa.

Overall, while most of the nations can be characterised into one of the 4 aforementioned groups, the cause of their exposure to malicious routing differs vastly from nation to nation and does not necessarily stem from a nation’s level of freedom. Even within groups of similar characteristics the cause of malicious routing, or lack thereof, cannot necessarily be prescribed to the related level of freedom. While the overall correlation showed a weak positive trend, it seems that the cause of the correlation is different for every nation.

## 7.3 Limitations & Future Work

One of the major limitations of this project was the inability to quantify two of the features of malicious routing: mass surveillance and citizens arrested over social media posts. Both these features lacked concrete datasets to draw from due to the natural unreliability of the information being collected. It may be worthwhile in the future, trying to develop a more accurate dataset that is capable of quantifying either one of these features. For example, since the source quality varies quite widely for this kind of information, it could be useful to develop a machine learning algorithm that scans news aggregation to detect publicly recorded arrests and determine whether they were of a reliable source. Quantifying the level of mass surveillance in a given nation is far more difficult due to the high level of secrecy that surrounds such areas. Further research could be conducted to determine whether it would be possible to quantify mass surveillance based on open source and leaked information alone. However even if this were possible, it would not guarantee that a usable dataset would be produced at the end.

The Human Freedom Index comprised of many individual features, normalised and aggregated together to get a final overall value judging the freedom in each nation. If there were more time, it would be an interesting endeavour to utilise machine learning to try to determine whether a single feature in this index correlates especially highly with malicious routing. This may narrow the search for a potential cause to the level of malicious routing activity in a given nation.

Due to the boolean nature of the unquantifiable features, they were given a manual reduced weighting. If these features were to be quantifiable in the future, it would be beneficial to use a Random Forest or Decision Tree machine learning algorithm to determine the best weightings for each of the features. This was experimented with for this report, however the results were inconclusive due to the large differences between datasets with and without the unquantifiable features. Furthermore, taking multiple reading of the Mirai activity, information exposure and BGP hijacks over a longer period of time could help to verify the findings of this report and create the opportunity to measure statistical significance which was not possible with the current results from this study.

In an ideal world, I would’ve liked to have had the time to dig into the background of each nation individually and attempt to determine the root cause of their level of malicious routing. However, this task would have been far too complex and time consuming. It is feasible that could take the time it took for this project, just to fully analyse a single nation’s background in this manner. In reality, it may be better to try to find correlations with other features of each nation instead. For example, during this project the Spearman’s rank correlation coefficient between GDP and malicious routing was determined to be -0.435, which was a stronger correlation than anything found between freedom and malicious routing. With further research it could be found that a nation’s wealth is a better predictor of malicious routing than freedom.

Lastly, it was noted that a few developed, free nations did exhibit low levels of malicious routing (Iceland, Estonia, etc…). It would be interesting for further research, to determine whether there is a political, cultural or ideological link between these nations and whether other nations with similar characteristics can learn something from these unusual cases.

## 7.4 Final Thoughts

I have found this project to be incredibly interesting and I very much enjoyed the opportunity to explore real-world philosophical, political and technological problems based around the issue of Internet Freedom. I believe the project to have been a success in determining whether a correlation exists between Internet Freedom and malicious routing, but also in opening further avenues for questioning.

# References

[1] Glenn Greenwald. NSA collecting phone records of millions of Verizon customers daily | US news | The Guardian. 6th June 2013. Available at: https://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order. [Accessed 09 December 2018].

[2] Glenn Greenwald. XKeyscore: NSA tool collects 'nearly everything a user does on the internet'. The Guardian. 31st July 2013. Available at: https://www.theguardian.com/world/2013/jul/31/nsa-top-secret-program-online-data. [Accessed 09 December 2018].

[3] John Goetz and Frederik Obermaier. “Snowden reveals the names of the peering telecom companies.” Suddeutsche Zeitung, 2 Aug. 2013.  <https://www.sueddeutsche.de/digital/internet-ueberwachung-snowden-enthuellt-namen-der-spaehenden-telekomfirmen-1.1736791> [Accessed 24 Oct. 2018].

[4] Freedom House USA. “Democracy in Crisis.” Freedom in the World Index 2018. 2018 - ongoing (updating)  <https://freedomhouse.org/report/freedom-world/freedom-world-2018> [Accessed 24 Oct. 2018]

[5] CATO Institute. “Human Freedom Index.” The Human Freedom Index – 2018. <https://www.cato.org/human-freedom-index> [Accessed 10th December 2018].

[6] Damien Gayle. “Downward spiral: UK slips to 40th place in press freedom rankings.” The Guardian. 26 Apr 2017. <https://www.theguardian.com/media/2017/apr/26/uk-world-press-freedom-index-reporters-without-borders> [Accessed 20 Oct. 2018]

[7] Rachee Singh et al.  “The Politics of Routing: Investigating the Relationship Between AS Connectivity and Internet Freedom.” Stony Brook University. 2016. <https://people.cs.umass.edu/~phillipa/papers/foci16-final16.pdf> [Accessed 20 Oct. 2018]

[8] Freedom House. “Freedom of the Press Index”. 2018 – ongoing. https:// freedomhouse.org/report-types/freedom-press. [Accessed 12th December 2018].

[9] Freedom House. “Freedom of the Net Index”. 2018 – ongoing <https://freedomhouse.org/report/freedom-net/freedom-net-2018> [Accessed 12th December 2018].

[10] Anne Edmundson et al. “Nation State Hegemony in Internet Routing”. Princeton University. 2018.  <https://www.cs.princeton.edu/~jrex/papers/compass18.pdf> [Accessed 20 Oct. 2018]

[11] Peter Mell et al. “Quantifying Information exposure in Internet Routing”. National Institute of Standards and Technology. 2016. https://ws680.nist.gov/publication/get\_pdf.cfm?pub\_id=925224 .

[Accessed 17 Dec. 2018]

[12] Daniel Donni et al. “Schengen Routing: A Compliance Analysis”. University of Zurich. 2015. <https://files.ifi.uzh.ch/CSG/staff/doenni/extern/publications/Schengen_Routing_A_Compliance_Analysis_AIMS_2015.pdf>.

[Accessed 17 Dec. 2018]

[13] Jose Nazario. “Politically Motivated Denial of Service Attacks”. Arbor Networks. 2016. <http://www.senki.org/wp-content/uploads/2017/10/12_NAZARIO-Politically-Motivated-DDoS.pdf>

[Accessed 17 Dec. 2018]

[14] “Distributed Denial of Service Database”. University of Twente. Ongoing. <https://ddosdb.org/help> [Accessed 17 Dec. 2018].

[15] Freedom on the Net 2018. Freedom House. 2018. Freedom on the Net 2018. Available at: https://freedomhouse.org/report/freedom-net/freedom-net-2018. [Accessed 19 December 2018].

[16] Freedom of the Press 2017. Freedom House. 2018. Freedom of the Press 2017. Available at: https://freedomhouse.org/report/freedom-press/freedom-press-2017. [Accessed 19 December 2018].

[17] Fraser Institute. 2018. The Human Freedom Index 2018 | Fraser Institute. Available at: https://www.fraserinstitute.org/studies/human-freedom-index-2018. [Accessed 21 December 2018].

[18] Mirai-like Botnet. 2018. Bad Packets Report. Available at: https://mirai.badpackets.net/. [Accessed 21 December 2018].

[19] BGPStream. 2018. BGPStream. Available at: https://bgpstream.com/. [Accessed 21 December 2018].

[20] Roberts, M. 2018. Censored Distractions and Diversion inside China’s Great Firewall. Princeton University Press.

[21] Roberts, Margaret Earling. 2014. Fear, Friction, and Flooding: Methods of Online Information Control. Doctoral dissertation, Harvard University. Available at: <https://dash.harvard.edu/handle/1/12274299> [Accessed 27 December 2018].

[22] BBC News. 2018. Yang Kaili: China live-streamer detained for 'insulting' national anthem - BBC News. Available at: https://www.bbc.co.uk/news/world-asia-china-45859650. [Accessed 27 December 2018].

[23] Freedom on the Net Countries. Freedom House. 2019. Freedom on the Net Countries. Freedom House. Available at: https://freedomhouse.org/report/countries-net-freedom-2018. [Accessed 09 February 2019].

[24] BBC News. 2019. Security services 'prevented 13 UK terror attacks since 2013' - BBC News. Available at: https://www.bbc.co.uk/news/uk-39176110. [Accessed 09 February 2019].

[25] The Independent. 2019. Twitter joke led to Terror Act arrest and airport life ban. The Independent. Available at: https://www.independent.co.uk/news/uk/home-news/twitter-joke-led-to-terror-act-arrest-and-airport-life-ban-1870913.html. [Accessed 09 February 2019].

[26] The Independent. 2019. Arrests for offensive Facebook and Twitter posts soar in London. The Independent. Available at: https://www.independent.co.uk/news/uk/arrests-for-offensive-facebook-and-twitter-posts-soar-in-london-a7064246.html. [Accessed 09 February 2019].

[27] Jaikumar Vijayan. 2019. Attackers turning to legit cloud services firms to plant malware. Computerworld. Available at: https://www.computerworld.com/article/2484596/cybercrime-hacking/attackers-turning-to-legit-cloud-services-firms-to-plant-malware.html. [Accessed 10 February 2019].

[28] Google. 2019. Google Scholar. Available at: https://scholar.google.com/. [Accessed 15 February 2019].

[29] PyPI. 2019. geoip2. PyPI. Available at: https://pypi.org/project/geoip2/0.1.0/. [Accessed 16 February 2019].

[30] J. Patrick Rhamey Jr. 2019. 2018 State of World Liberty Index — J. Patrick Rhamey Jr. Available at: http://patrickrhamey.com/saturday-research/2018/8/6/2018-state-of-world-liberty-index. [Accessed 17 February 2019].

[31] David D. Clark. 2000. Rethinking the design of the Internet: The end to end arguments vs. the brave new world. Available at: <https://cyberlaw.stanford.edu/e2e/papers/TPRC-Clark-Blumenthal.pdf> [Accessed 22 February 2019].

[32] allAfrica.com. 2019. Malawi: Huawei Hails Government On National Fibre Backbone Project. allAfrica.com. Available at: https://allafrica.com/stories/201805010206.html. [Accessed 03 March 2019].

[33] Malawi Country Report. Freedom on the Net 2018. 2019. Malawi Country Report. Freedom on the Net 2018. Available at: https://freedomhouse.org/report/freedom-net/2018/malawi. [Accessed 03 March 2019].

[34] countryeconomy.com. 2019. Human Development Index - HDI 2017. countryeconomy.com. Available at: https://countryeconomy.com/hdi. [Accessed 05 March 2019].

[35] GDP per capita (current US$) Data. 2019. GDP per capita. Available at: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD. [Accessed 05 March 2019].

[36] Heckelman, J.C., 2000. Economic freedom and economic growth: a short-run causal investigation. Journal of applied economics. Available at: https://core.ac.uk/download/pdf/7146349.pdf [Accessed 05 March 2019].

[37] David Rieff. Why nobody cares about the surveillance state. Foreign Policy. 2013. Available at: <https://foreignpolicy.com/2013/08/22/why-nobody-cares-about-the-surveillance-state/>. [Accessed 09 March 2019].