**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](#_Toc1308064)

[1.1 Chapter Overview 7](#_Toc1308065)

[1.2 Motivation for Research 7](#_Toc1308066)

[1.3 Aims and Objectives 8](#_Toc1308067)

[1.4 Report Structure 9](#_Toc1308068)

[Background Research 10](#_Toc1308069)

[2.1 Chapter Overview 10](#_Toc1308070)

[2.2 The Politics of Routing: Investigating the Relationship between AS connectivity and Internet Freedom 10](#_Toc1308071)

[2.3 Nation-State Hegemony in Internet Routing 11](#_Toc1308072)

[2.4 Quantifying Information Exposure in Internet Routing 11](#_Toc1308073)

[2.5 Schengen Routing: A Compliance Analysis 12](#_Toc1308074)

[2.6 Politically Motivated DDoS 13](#_Toc1308075)

[2.7 Related Work: Freedom Indexes 14](#_Toc1308076)

[2.7.1 Freedom on the Net 2018 14](#_Toc1308077)

[2.7.2 Freedom of the Press 2017 14](#_Toc1308078)

[2.7.3 The Human Freedom Index (*updated for 2018*) 14](#_Toc1308079)

[2.7.4 State of World Liberty Index 2018 15](#_Toc1308080)

[2.8 Related Work: Data Logs 15](#_Toc1308081)

[2.8.1 Bad Packets Report (Ongoing) 15](#_Toc1308082)

[2.8.2 Mirai Botnet Activity (Data from Jan 2017 – Jan 2019) 15](#_Toc1308083)

[2.8.3 BGPMON (Ongoing) 15](#_Toc1308084)

[2.8.4 Routing Dependencies (Jan 2019) 16](#_Toc1308085)

[Freedom 17](#_Toc1308086)

[3.1 Chapter Overview 17](#_Toc1308087)

[3.2 Defining Freedom 17](#_Toc1308088)

[3.3 Declining Freedom & China 18](#_Toc1308089)

[3.3.1 Social Engineering as a means to an end 18](#_Toc1308090)

[3.4 Freedom Indexes 20](#_Toc1308091)

[3.4.1 Freedom on the Net 20](#_Toc1308092)

[3.4.2 Freedom of the Press 20](#_Toc1308093)

[3.4.3 The Human Freedom Index 20](#_Toc1308094)

[3.4.4 State of World Liberty 21](#_Toc1308095)

[Malicious Routing 22](#_Toc1308096)

[4.1 Chapter Overview 22](#_Toc1308097)

[4.2 What is Malicious Routing 22](#_Toc1308098)

[4.2.1 Mass Surveillance 22](#_Toc1308099)

[4.2.2 Arrests made over social media posts 23](#_Toc1308100)

[4.3.3 Mirai Botnet IP Addresses 23](#_Toc1308101)

[4.3.4 Mirai-like Packet Signatures 24](#_Toc1308102)

[4.3.5 BGP Hijacks 24](#_Toc1308103)

[4.3.6 Information Exposure 24](#_Toc1308104)

[4.3.8 Feature Weightings 25](#_Toc1308105)

[4.3 Nations and Autonomous Systems 25](#_Toc1308106)

[4.4 Methodology 26](#_Toc1308107)

[4.4.1 Data Discovery and Collection 26](#_Toc1308108)

[4.4.2 Data Reformatting 26](#_Toc1308109)

[4.4.3 Data Processing & Weighting 28](#_Toc1308110)

[4.4.4 Calculating Rankings 28](#_Toc1308111)

[4.4.5 Rankings 29](#_Toc1308112)

[4.5 Correlation Methods 31](#_Toc1308113)

[4.5.1 Types of Correlation 31](#_Toc1308114)

[4.5.2 The FOTN Overlap Problem 31](#_Toc1308115)

[Results 32](#_Toc1308116)

[5.1 Chapter Overview 32](#_Toc1308117)

[5.2 Correlations 32](#_Toc1308118)

[Analysis & Findings 33](#_Toc1308119)

[6.1 Chapter Overview 33](#_Toc1308120)

[6.2 Policy 33](#_Toc1308121)

[6.3 Case Studies 33](#_Toc1308122)

[Conclusions 34](#_Toc1308123)

[7.1 Chapter Overview 34](#_Toc1308124)

[References 35](#_Toc1308125)

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.

## Aims and Objectives

The aims of this report are therefore as follows:

* 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.

## 

## 1.4 Report Structure

This report will begin with a review of the relevant 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, the report will cover the various definitions of freedom and determine which indexes are most applicable when correlating with malicious routing. 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 will also help determine whether the best predictor of internet freedom is a nation’s propensity to conduct malicious routing activity.

The main section of the report will focus on the technical aspects of malicious routing in defining its component elements, examining the nations that partake in it and finding any correlations with freedom. This will begin by accurately defining what actions and policies 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 contrast malicious routing behaviour with freedom 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: Freedom.

Depending on the results of the correlations, there may be clear connections between certain kinds of malicious routing behaviour and freedom. This next section will examine trends seen across the data as well as stand-out and edge cases. This will be followed up with the ‘findings’ chapter that will contain a careful analysis of the results and a critical review of 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 will be 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 policieis, such as the valley-free roule. Features of this graph were then extracted and applied to various machine learning techniques that tried to predict the freedom of press index of a 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 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 my 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. I 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:



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 robustness and information exposure. An important feature when measuring a nation’s internet robustness was the number of independent routing paths between 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 1* has been extracted from the report and shows compliance levels by nation within the EU.



Figure 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, some nations will be subject to unique features that are not applicable to others. Therefore, for this report it may be necessary to assess the impact of legislation such as this for each nation individually.

## 2.6 Politically Motivated DDoS

This paper [13] documents and analyses the reasons behind many of the recent denial of service attacks directed from governments. It documents 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 the exposure of other nation’s 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.

However, predictions are frequently wrong, therefore other freedom measures listed below will also be analysed.

### 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 significantly different from 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 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 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 not any 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 make clear that 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 in use 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, this has sometimes been favourable for intelligence agencies such as the NSA (US) who were 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. Whereby, the internet removes many physical and economic constraints on an individual’s ability to access, however also holds vastly 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 paper 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], 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 choses 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 diagram below outlines an issue with imperfect censorship over extended periods of time, as experienced by governments around the world:



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 heavy-handedness 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 most simple 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 in a sea of pro-government or contradictory comments or links. This makes it very difficult or 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 by Roberts as “increasing the cost, either in time or money of access or spread of information”. This plays off the impatience of human nature through techniques such as making problematic websites take longer to load and making the messaging app WeChat lag in places where protests are taking place. 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 citizens of any given nation are exposed online social engineering policies from their own government.

## 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 will 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 margin of difference 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 closely related to that government’s amount of 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 Gross Domestic Product. 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]. 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.

### 4.3.3 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. 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.

### 4.3.4 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.

### 4.3.5 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.

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.

### 4.3.6 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.

### 4.3.8 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 will 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, allowing exposure to malicious routing can be calculated on a per nation basis.

By indexing by nation, comparisons with legislation, policies, ideology and levels of freedom dramatically simplify the process of analysis and finding correlations. 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, open source access to the datasets referenced by useful academic literature was not available. In some cases, new data was collected using mirrored techniques and in others 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 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 |

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 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 |

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 |

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 nations 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 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 |

### 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 this process is shown 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 |

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)) = … |

Therefore, 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 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 |

## 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. 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. They both 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.

Findings:

* Is there a general trend?
* What does the trend mean?

Analysis:

* Why do you think the results were as such?
  + REASONS FOR THE TREND (policies etc…)
  + Consistent differences in the results and the indexes + why.
  + Similarities ^
  + + why? (Australia High, China lower than expected…) (case studies)
  + The issue of very poorly developed countries (gdp graph)
* What were the downfalls of the study?
  + Any issues and how may have they affect the results
  + Ambiguity of “citizens arrested” and “mass surveillance”

Edge cases

Difficulties etc…

NORMALISED BY DIVIDING BY NUMBER OF IPS… HOWEVER FOR GOVERNMENT ACTIVITY, IT DOES NOT REQUIRE A LARGE NUMBER OF IPS TO PERFORM A LOT OF MALICIOUS ROUTING… MAYBE BASE VALUES WOULD BE BETTER SUITED?

* This moved china down the list because of their large number of ips, however they’re still considered very oppressed online.

## 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 bar chart 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.

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 bar chart 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.

### 6.2.2 Trend Implications

If correlation does not mean causation, then what does it mean? A correlation between variable A and variable B can mean a few things:

* A causes B
* B causes A
* Pure coincidence
* A and B are connected by the same cause
* Both A and B contribute to each other in a loop

Some users may shy away from correlation analysis because it requires some backstory, and because it is considered one of the less understood forms of analysis. For instance, homelessness and crime may be high in the same areas or low in the same areas (and thus are correlated). After having read this post, you know that to say that homelessness causes crime, or crime causes homelessness would be imprudent unless further research was conducted. The two variables could be caused by a third and unknown variable, such as education or drug addiction.

The general trend across all the indexes is a weak positive correlation.

## 6.3 Analysis

Sample text

Chapter 7

# Conclusions

## 7.1 Chapter Overview

Sample text

Conclusion?

Final thoughts?

Changes for future research?

# References

[1] Glenn Greenwald. NSA collecting phone records of millions of Verizon customers daily | US news | The Guardian. [ONLINE]. 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. [ONLINE]. 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].