A longitudinal study of DNS traffic: Understanding current DNS practice and abuse

A thesis submitted in fulfilment of the requirements for the degree of

MASTERS IN COMPUTER SCIENCE

of

RHODES UNIVERSITY

by

Ignus van Zyl

December 2015

Abstract

This thesis examines a dataset spanning 21 months, containing 3.5 billion DNS packets. Traffic on TCP and UDP port 53, was captured on a production /24 IP block. The purpose of this thesis is twofold. The first is to create an understanding of current practice and behavior within the DNS infrastructure, the second to explore current threats faced by the DNS and the various systems that implement it. This is achieved by drawing on analysis and observations from the captured data. Aspects of the operation of DNS on the greater Internet are considered in this research with reference to the observed trends in the dataset. A thorough analysis of current DNS TTL implementation is made with respect to all response traffic, as well as sections looking at observed DNS TTL values for .za domain replies and NXDOMAIN flagged replies. This thesis found that TTL values implemented are much lower than has been recommended in previous years, and that the TTL decrease is prevalent in most, but not all RR TTL implementation. With respect to the nature of DNS operations, this thesis also concerns itself with an analysis of the geolocation of authoritative servers for local (.za) domains, and offers further observations towards the latency generated by the choice of authoritative server location for a given .za domain. It was found that the majority of .za domain authoritative servers are international, which results in latency generation that is multiple times greater than observed latencies for local authoritative servers. Further analysis is done with respect to NXDOMAIN behavior captured across the dataset. These findings outlined the cost of DNS misconfiguration as well as highlighting instances of NXDOMAIN generation through malicious practice.

With respect to DNS abuses, original research with respect to long-term scanning generated as a result of amplification attack activity on the greater Internet is presented. Many instances of amplification domain scans were captured during the packet capture, and an attempt is made to correlate that activity temporally with known amplification attack reports. The final area that this thesis deals with is the relatively new field of Bitflipping and Bitsquatting, delivering results on bitflip detection and evaluation over the course of the entire dataset. The detection methodology is outlined, and the final results are compared to findings given in recent bitflip literature.

Acknowledgements

I would like to acknowledge the people and organizations that helped me realize the completion of this paper. My heartfelt thanks to Professor Barry Irwin for his guidance and support throughout the year, without which none of this would be possible. I would also like to acknowledge the staff of the Computer Science Department at Rhodes University for their support.

I thank my family for their support and love during my time at university, as it has allowed me to turn my full attention to my studies.

Many thanks to my colleagues, Kieren Hunt, Lauren Rudman, Alan Herbert and Sean Pennefather for their advice and insight given throughout the year.

Gratitude is also due to ISACA, whose scholarship allowed me to pursue my further studies.

Finally, my thanks go to Gareth Dwyer, whose love of Python is eclipsed only by his love of beer, memes and Dota 2. He has been a stalwart companion and true friend, and without him this thesis would not have been completed or proofread timeously.

This research makes use of GeoLite data created by MaxMind.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This research focuses on data gathered around the usage and implementation of different aspects of the Domain Name System (DNS), with respect to current activity on the greater Internet. It will look at both legitimate and malicious usage of the Domain Name System within the scope of the collected and analyzed data. The data was collected using an IPv4 address block used for production purposes, and as such reflects interactions between existing end-hosts and the Domain Name System as implemented on the greater Internet.

This first chapter serves as an introduction to the thesis, as well as the research herein. The problems that resulted in the instantiation of the research will be discussed, as well as the perceived significance of the research. The goals of the research will be outlined, following which the scope and limitations of the research will be considered. The last area deals with the layout of the sections that are to follow, which will comprise the main body of the thesis.

1.1 Problem Statement

The following text outlines the problems that prompted this research, as well as a motivation for the significance of this research in the current field of Computer Science.

The development of the Domain Name System came about as a result of previously existing name resolution services not being able to meet the needs of the growing network infrastructure that has developed into the modern Internet (Aitchison, 2005). DNS, as a result of its relationship with the Internet, is a dynamic system. The system experiences implementation changes and developments, as well as revisions to the aforementioned, as the requirements and functionality of the Internet and its connected end-hosts evolve. The various DNS infrastructures and capabilities of the current era are far removed in both scope and ability from the domain name system that was created to replace the Name Servers that came before them. Unfortunately, many of the principle documents outlining DNS infrastructure and implementation are archaic (Lottor, 1987;

Mockapetris, 1987a,b), yet still form the core of DNS documentation despite revisions. This would suggest that while there have been reactive actions taken to improve and utilize DNS in the modern era, there has been less consideration than necessary on how it will be affected by current Internet usage and implementation.

The distributed and fluctuating nature of the Internet and its end-hosts has enshrined DNS as an indispensable tool for network maintenance and usability (Moore and Edelman, 2010). With DNS becoming an integral part of the functioning of the Internet, it sees both legitimate and malicious usage on a large scale, in many different forms. This creates multiple opportunities for research within the field of Information Security, each of which touches on multiple disciplines within the field.

1.1.1 Significance of Research

The Domain Name System is one of the commonly used infrastructures that allow for the existence of the Internet as we know it (Agten *et al.*, 2015). Research in this area touches on multiple aspects, including but not limited to: Network configuration and usage, Network and server optimization, Network and end-host security, End-user experience, as well as different implementations of DNS and the threats generated by those implementations. The fact that research in this area contributes to so many disciplines is important with respect to the advancement of research within those disciplines, as well as Computer Science as a whole.

There is of course a plethora of research with respect to aspects of DNS, most notably in the fields of system optimization from a computational perspective, and also previous and existing threats that come about as a result of DNS implementation, configuration or usage. This research, however, also endeavors to analyze legitimate traffic and DNS configurations, of which there is a surprising lack. That is not to say that there is no work on current observable DNS traffic, merely that there is not much research that gives consideration to the normal DNS traffic that is generated through Internet usage. This research also delivers findings on the relatively new research platform of Bitflipping (Dinaburg, 2011), in the hope that it contributes to the sparse but growing collection of research on the subject.

This research will also attempt to give a South African perspective on certain aspects of DNS infrastructure and its configuration, in an attempt to make the findings of this thesis more relevant to local researchers and organizations.

Earlier findings on DNS TTL analysis using this dataset were published in the SATNAC 2015 proceedings (van Zyl *et al.*, 2015), which indicates that there is interest in this area of research.

1.2 Research Goals

There are two key aims of this research.

1.2.1 Operation

The first is to understand the ways that legitimate network entities are using the DNS infrastructure and its capabilities. This allows us to see how normal users are interacting with the infrastructure, as well as allowing us to understand how the expectations of end-users have changed with respect to DNS over the years. The two main focuses of this area will be:

- DNS TTL analysis
- DNS authoritative server geolocation and latency for .za domains

These areas will hopefully give the reader an understanding of some of the DNS implementation and configuration choices made by entities on the Internet, as well as shedding some light on current DNS practices.

1.2.2 Abuse

The second is to observe instances where DNS is being used outside of the scope of legitimate traffic, in order to better understand threats that are generated through the use and abuse of DNS and its sub-protocols. These fall into the following two categories, which were observed in the captured dataset:

- Post-attack DNS amplification scanning
- DNS NXDOMAIN analysis

The NXDOMAIN analysis, section 4.3, is interesting as it touches on both the fields of DNS operations/practice as well as possible malicious DNS use/abuse. The Post-attack scanning study, section 5.1, deals solely with DNS abuse, but makes reference to the specific infrastructures unique to DNS that make this abuse possible.

The final area of research combines the above two areas, as it has significant research value for both the legitimate and malicious spheres of DNS usage, and will be comprised of:

- Bitflipping and Bitsquatting presence in DNS

Section 5.2 explores the presence of bitflips as well as bitsquats captured in the dataset. It offers analysis on a new field of computer science, and looks specifically at examples of abuse through Bitsquatting.

1.3 Research Scope

The scope of this thesis is defined here to give a more definite context for the research presented in the following chapters. It is an important factor in not only understanding what research was possible, but also why some avenues of research were considered over others.

The dataset was made available under the condition that the source of the data was not revealed. This means that certain analysis could not be reported upon, as doing so would enable the identification of the source of the data. An example of this is NXDOMAIN analysis conducted on packets seen at the authoritative server, which were later removed from the thesis.

Malformed or mangled DNS packets were filtered out. The thesis does not concern itself with packet preservation or mangling, and as such this was considered out of scope.

A number of known misconfiguration errors were also filtered out of the dataset, as they generated millions of identical packets which did not offer opportunities for further analysis.

The scope of the research is also limited by the actual packets that were captured by the network monitor. As such there are many avenues of DNS related network activity that could not be reported on, for example amplification attacks, more specifically response packet backscatter; this was simply because there were no packets of that nature captured in the dataset.

1.3.1 Limits of Research

The first and most important limit of this research is the nature of the IPv4 block from which the data is gathered. Analysis on domains, TTLs, observed server latencies etc. will only be on domains or IPs that have interacted with the authoritative and caching servers as a result of their common usage within the IP block. As such, this thesis will not be able to deliver a holistic interpretation of current DNS activity and implementation, and can only concern itself with the traffic that made itself known to the IP block during the time of data-gathering. This is not to say that there is a lack of data from which research can be generated, but only that the research will not be able to give a representation of DNS activity for certain spheres of the Internet. For instance, since this IP block is geographically located in South Africa, it is more likely that captured traffic will be in English, and target common western and .za domains. This also means that there will be little to no captured traffic for domains specific to Malaysia, for example. It also means that the likelihood of capturing packets using other character sets (e.g. Traditional Chinese) is also very low. It is also difficult to obtain data of this nature, creating another limitation with respect to research.

Another limit of this research is the fact that, while the dataset captured scans for possible DNS amplification attacks, there was no actual DDoS attack captured on the dataset, as none of the 256 IP addresses were the target of such a DDoS attack (Rossow, 2014). As such the research relies on a third party[1], which reports DNS amplification attacks observed by an open resolver, as a validation of post-attack scanning behavior observed in the dataset.

1.4 Document Conventions

This section introduces some of the formatting and presentation conventions followed throughout the document, and seen in subsequent chapters.

Footnotes are used to indicate where tools used in this research can be accessed or downloaded.

All decimal values have been rounded to the 3rd decimal point.

All numbers split after every three digits for legibility.

All domain names italicized. All organization names in bold font.

All countries given in ISO 3166-1 alpha-2 format unless whole name is given; e.g. UK, ZA, US.

The minus symbol (-) appearing in tables indicates that there was no relevant data captured during that period.

Where figures or tables have not been referenced, they were created by the researchers themselves.

1.5 Document Structure

The document consists of six chapters, of which this is the first. Chapters two and three serve the purpose of contextualizing the findings of the thesis. Chapters four and five report on the analysis and findings of the thesis itself, while chapter six holds concluding remarks. The remainder of this document is structured as follows:

Chapter 2 gives an introduction to the technical concepts covered in the paper, discuss threats to DNS systems, as well as present a review of the relevant literature in the areas pertaining to this thesis.

Chapter 3 discusses the origin and processing of the dataset itself, as well as supplying heuristics on the data captured.

Chapter 4 focuses on DNS Operations, and delivers analysis on three areas, observed DNS TTL values; observed DNS latency and geolocation for authoritative servers of .za domains; and an analysis of NXDOMAIN traffic.

[1] *http://dnsamplificationattacks.blogspot.co.za*

Chapter 5 looks at DNS abuse, with sections on captured amplification scanning traffic; and gives an architecture for possible bitflip detection and the results of bitflipping and bitsquatting analysis.

Chapter 6 forms the conclusion of the thesis, and gives suggestions for future work in the area of DNS analysis.

Chapter 2

Background and Literature Review

This chapter is split into multiple sections, all of which are meant to familiarize the reader with the concepts present in the field of DNS analysis, as well as give the reader a broader understanding of the terms and concepts that will appear throughout the paper. Section 2.1 deals with some of the DNS specific jargon that will appear throughout the thesis. Section 2.2 discusses the various threats to DNS infrastructure and stability. Of these, two are covered extensively in the thesis. Past and current research is presented in section 2.3. Interesting concepts or analysis identified in specific sub-fields of DNS research are discussed in order to give the reader a more complete understanding of past and current research.

2.1 Technical Concepts

This section gives a simple explanation of some of the jargon that is seen throughout the thesis, and discusses how these concepts relate to the thesis itself.

2.1.1 Pcap files

Packet capture (pcap) files are datasets created by recording packet information across a connection using a passive network monitor (Williamson, 2001). The network monitor reads, or 'sniffs', packets that travel through the connection it is monitoring, but the monitor does not create or alter packets in any way (Williamson, 2001) - it merely reads and records them. The initial datasets of the thesis, before processing, were in pcap[1] format.

[1] http://www.tcpdump.org/

2.1.2 Authoritative and Caching servers

Authoritative servers are servers that only provide responses for zones for which the server is either a zone master or a zone slave, and does not allow for recursive queries (Aitchison, 2005). Apart from the zone records for which they are responsible, they do not store or communicate any other records. Caching servers are name servers that provide recursive query support to end-hosts and save responses in the local DNS cache memory (Aitchison, 2005).

2.1.3 DNS Time-to-live values

Records stored in the caching resolver memory have a 32 bit unsigned integer value called the time-to-live (TTL) value (van Zyl *et al.*, 2015). Each resource record has a TTL value set by the administrator of the DNS domain, which tells the caching resolver how long the cached record should remain in memory, in seconds (van Zyl *et al.*, 2015). Once the TTL expires, the caching server will stop replying to queries with the cached response and query the authoritative server for an updated record (Aitchison, 2005).

2.1.4 Resource records

Resource records (RR) define certain characteristics or properties contained within the domain (Aitchison, 2005). Table 2.1 describes the functions of RRs that appear throughout the thesis.

Table 2.1: Explanation of RRs (van Zyl *et al.*, 2015)

Resource record	Description
A record	Returns the IPv4 address for a host of the domain
PTR record	Returns reverse-mapped domain name of IP address
CNAME record	Returns an alias for an existing host given by an A RR
TXT record	Returns generic text associated with domain
MX record	Returns the mail servers for the domain
AAAA record	Returns forward mapping of IPv6 hosts as A does for IPv4
NS record	Returns the authoritative name servers for the domain
SOA record	Returns the key characteristics and attributes for the domain
SRV record	Allows for discovery of services provided by host

2.1.5 Network latency

Network latency forms part of the overall latency experienced by users, i.e. the amount of time between them requesting the content and content delivery, on the Internet. Various factors come into play within network latency. The first is propagation delay, which is the delay generated by the distance the packets have to travel to reach the destination (Padmanabhan and Mogul, 1996). DNS-based latency, (or name resolution latency), is the latency generated by the DNS resolution process during the overall network interaction (Jung *et al.*, 2002). The latency values given in this thesis refer to the latency generated through contacting the authoritative servers of .za domains.

2.1.6 Open resolvers

Open resolvers are public DNS resolvers that serve recursive name lookups to any client that contacts it with a DNS query (Rossow, 2014). This means that the server will respond to queries from any host on the Internet. As such, open resolvers are used as 'reflectors', allowing attackers to spoof an IP address and inundate the target IP with packets from open resolvers in different subnet blocks, effectively launching a DDoS attack (Paxson, 2001).

2.1.7 DNS Blackhole Lists

DNS blackhole lists (DNSBL), or Real-time Blackhole lists, are databases containing IP addresses and/or domain names which have been identified as spam sources, and can be queried. Queries will return that the IP address or domain is in the blackhole list, and should thus be filtered or marked, or respond that the queried value is not in the list (Miszalska *et al.*, 2007). Right-hand-side blackhole lists are a subset of DNSBLs that contain lists of spam email TLDs. The name comes from the fact that the right hand side, i.e. after the @ sign, of the email address is validated (Miszalska *et al.*, 2007).

2.1.8 Bitflipping

Bitflipping is the occurrence of random errors, as a result of software or hardware malfunction, radiation or environmental factors, that manifests as the corruption of one or more bits of the data (Dinaburg, 2011). Figure 2.1 illustrates a possible bitflip. The 'n' character is sent to be stored in memory, and has the ascii binary value 01101110. However, the one of the bits in memory becomes corrupted, resulting in the value 01101111, or 'o'.

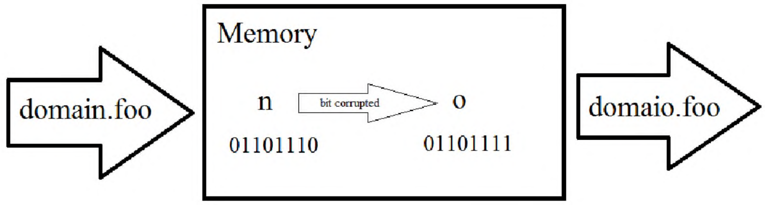


Figure 2.1: Bitflip diagram

While most bitflips do not have an impact on host activity, some bitflips create opportunities for malicious entities to gain information about or to attack end-hosts, as a result of the corrupted information being web-facing (Dinaburg, 2011).

2.2 DNS Threats

This section highlights some of the methods that malicious entities use, with respect to the DNS infrastructure, to launch or control illegitimate web activity. It covers historic DNS abuses, and also discusses some of the current threats faced by DNS implementations.

2.2.1 Historical DNS threats

A number of different methods of domain squatting have manifested themselves throughout the years.

One of the main functions of DNS is the resolution of domains to IP addresses (Aitchison, 2005), so it comes as little surprise that this functionality is targeted and abused. The first form of squatting was the aggressive registering of domains that others might want to use, and then selling these to organizations or persons that are interested in acquiring the domain, a process known as cybersquatting (Moore and Edelman, 2010). Typosquatting, the practice of registering mistyped popular domains, began as a practice in 1999 (Moore and Edelman, 2010). Typosquatting relies on the fact that users make mistakes when typing the domain (Agten *et al.*, 2015), which will then resolve to the squatter host instead of the intended host.

Soundsquatting is a variation on typosquatting, where the incorrect part of the domain will be a homophone of the correct domain (Nikiforakis *et al.*, 2014). An example of this would be textsale.ru (legitimate domain) and textsail.ru (soundsquatted domain) (Nikiforakis *et al.*, 2014), where the incorrect user input will direct users to the squatted site instead of the legitimate content server. Homograph attacks form another subset of squatting activity. Attackers will register domains that render similarly if not identically to legitimate domains (Holgers *et al.*, 2006). This form of squatting differs from others as it does not rely on the target host to mistype the domain, but rather relies on their lack of ability to distinguish between legitimate and homograph domains that are presented to them, prompting them to click on potentially malicious hyperlinks (Holgers *et al.*, 2006).

2.2.2 Cache poisoning

Cache poisoning is the act of changing or adding records to a resolver's cache, either on the client or server side, with the result of a DNS query for that domain returning the address to the attacker's domain instead of the legitimate address (Olzak, 2006). Cache poisoning attacks are carried out by querying for a legitimate domain, and then sending crafted responses, attempting to match the transaction ID of the query, in order to feed the caching resolver a malicious record (Olzak, 2006). Four large threats that face end-users are identity theft, distribution of malware, dissemination of false information and man-in-the-middle attacks, launched through the webpage that is served to the target host as a result of the poisoned record (Olzak, 2006).

2.2.3 Amplification attacks

An amplification attack is a DDoS attack that relies on the use of reflectors to generate large responses to small packets, pointed at the target host through IP spoofing (Paxson, 2001). DNS attacks, specifically, will abuse the fact that response packets can contain more data than query packets, particularly for ANY replies (Fachkha *et al.*, 2014) or EDNS0 enabled resolvers (Rossow, 2014), which generate responses sometimes orders of magnitude larger than the original query. DNS amplification attacks are commonly launched using open resolvers, as they accept and reply to queries from any source.

2.2.4 Fast-flux botnets

Service availability is a concern faced by both legitimate and malicious enterprises on the Internet (Nazario and Holz, 2008). Botmasters have been known to use dynamic DNS to ensure that bots can reach one of a number of Command and Control (C&C) hosts if the original one is taken down (Choi *et al.*, 2007). Attackers have taken this a step further by using fast-flux botnets, for which domain mappings are changed frequently to one or more of the controlled bots, which then act as a proxy for the C&C, relaying content between the botnet end-point and the malicious server (Nazario and Holz, 2008). This makes it significantly more difficult to block or request a takedown of the malicious service in question (Nazario and Holz, 2008).

2.2.5 Bitsquatting

Bitsquatting is a relatively new form of domain squatting identified by Dinaburg (2011). Bitsquatting relies on a DNS domain in memory experiencing a bitflip, which then leads to incorrect resolution of the domain through DNS (Dinaburg, 2011). Malicious entities will register domains that differ from popular domains by one bit, while still remaining a valid DNS domain, in an attempt to take advantage of the traffic routed to their servers as a result of bitflipping (Nikiforakis *et al.*, 2013).

2.3 Related Research

This section presents research that is relevant to the material seen in the thesis. The research is grouped into two main areas, *DNS Operations and Practice* and *DNS Threats and Abuse*. The former looks at the sphere of legitimate DNS use while the latter concerns itself with possibly malicious DNS activity.

2.3.1 DNS Operations and Practice

One of the first papers to deliver analysis on DNS traffic, 'An Analysis of Wide-Area Name Server Traffic' utilizes two 24-hour traffic traces to explore the performance of DNS on the network. It considers a variety of performance issues related to DNS traffic, some of which are still extremely relevant today. Three performance issues that were identified are Caching, Retransmission Algorithms and "The net effect", which is the generation of multiple queries as the result of the queried servers or their respective root servers being unreachable for a period of time (Danzig *et al.*, 1992). While this thesis does not concern itself with retransmission algorithms, both of the other issues were identified as important effectors of DNS traffic generation and bandwidth consumption. It was found that a large amount of DNS traffic was generated as a result of incorrect configuration of DNS domains and their related Resource Records, as well as misconfiguration with respect to the servers themselves (Danzig *et al.*, 1992). It was also noted that RPC timeouts and Cache leaks were a contributor to unnecessary traffic, a point that is confirmed in the TTL analysis section of this thesis. Danzig *et al.* (1992) also noted that, as of late 1991, the DNS namespace consisted of 16 000 different domains and around 1 000 000 leaf nodes that represented individual end-hosts, which serves to put into perspective how vastly different its structure is today.

'The Contribution of DNS Lookup Costs to Web Object Retrieval' looks at the lookup costs associated with DNS queries on a network. The analysis considered DNS TTLs, their values, the effect TTLs have on traffic reduction and how DNS TTLs related to overall DNS performance. The first statement was that raising TTL times would result in a higher cache hit-rate, i.e. more packets served a cached reply from the cache server (Wills and Shang, 2000). This is as a result of the records remaining live in the cache server longer and are as such able to serve more queries per cached record. The found that only 10% of records changed as their TTL record expired in the cache on a per server basis. Overall, only around 20% of records were changing between timeouts, indicating that a vast number of DNS TTL values are set too low (Wills and Shang, 2000). The researchers claimed that setting a minimum TTL value of 15 minutes would noticeably improve cache hit-rates (Wills and Shang, 2000).

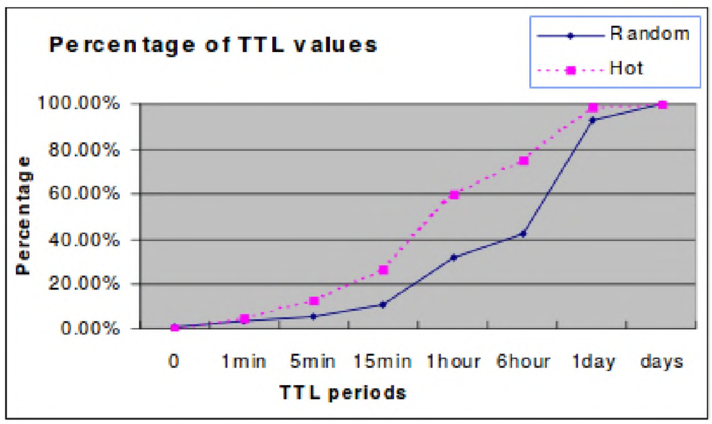


Figure 2.2: Cumulative Distribution of TTL vlues for Random and Hot Servers (Wills and Shang, 2000)

Figure 2.2 gives the cumulative distribution of TTL values seen for random and hot servers. A set of 100 popular, or hot, domains were compared to a set of 100 random domains. They found that, surprisingly, the popular domains had higher TTLs than random domains (Wills and Shang, 2000). While this analysis shows 20% of TTL values below the 10 minute mark, figure 2.3 shows a much lower TTL average thirteen years later.

An in-depth analysis on many aspects and behaviors of DNS traffic is delivered in 'An Empirical Reexamination of Global DNS Behavior', as it attempts to build on and compare results from previous analyses regarding observed DNS traffic. Comparisons with past papers yield interesting results with respect to changes in the nature of DNS traffic, including query type frequency, query success rates, DNS TTL distribution and the presence of repeated DNS queries (Gao *et al.*, 2013). Of particular interest is the comparison of TTL distribution with the previous work by Jung *et al.* (2002), mentioned above. This gives a comparative analysis of the evolution of DNS TTL practice relative to observed Resource Records, on which this thesis hopes to build. The TTL distribution in this paper is given in Figure 2.3.

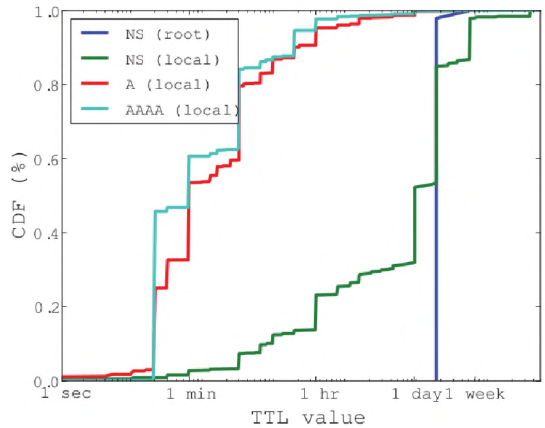


Figure 2.3: The cumulative distribution of TTLs of NS record returned by root servers, and three record types, A, AAAA and NS, returned by other servers. (Gao *et al.*, 2013)

This paper noted that there was a marked decrease in TTL values, most notably for the A and AAAA resource records observed in the dataset, when compared to the results presented in Jung *et al.* (2002). The increase in A and AAAA queries, as well as a decrease in PTR queries, was also noted when compared to previous research results. This paper provides invaluable findings with respect to TTL behavior that this thesis hopes to build on in the coming sections. Another important finding in this paper relates to the NXDOMAIN presence generated by Domain Name System Black Lists (DNSBL). This comes about as a result of how the DNSBL are configured, where queries for domains that are not on the list return NXDOMAIN response packets instead of confirmation reply packets (Gao *et al.*, 2013).

'A review of current DNS TTL practices' is a preliminary paper to the TTL analysis present in this thesis. This research looks at DNS TTL configuration over a six month period between January and June 2014. The research found that there was a strong trend towards lower TTL settings, sacrificing bandwidth and query response speed in order to decrease reaction time to downed servers or enable faster server load balancing (van Zyl *et al.*, 2015). It was noted that, while CDNs had the lowest TTL presence, other major web-based organizations such as **Google** and **Facebook** also tended to use TTL values far below those recommended (Lottor, 1987) in RFC 1033 (van Zyl *et al.*, 2015).

A short but interesting paper, 'An Exploration study into the location and routing of the most popular websites in South Africa' is on the geolocation of website hosting from a South African context. The authors geolocated web-hosters for the top 100 sites viewed from South Africa according to Alexa, and found that around 50% of the websites had locally hosted content; not including CDN content, for which the value remains unclear (Barnett and Ehlers, 2012). This study is also interesting as it notes the limitations of the MaxMind geolocation database, specifically suspected inaccuracies with respect to country bucketing(Barnett and Ehlers, 2012), which is used throughout this thesis as well. This paper explores web connectivity and international web configuration from a South African perspective, a theme that this thesis builds on.

In 'Speed Matters for Google Web Search', 'Google conducted an experiment on the effect latency had on end-users perception of the web service. They created fake latency at the server-end of some connections. Google found that slowing down search results by between 100 ms and 400 ms caused searches-per-user to decrease by between 0.2% and 0.6% (Brutlag, 2009). They found that user searches decreased further the longer they were exposed to the experiment. Users who had been exposed to a 200ms delay since the beginning of the experiment performed 0.22% fewer searches during the first three weeks, but 0.36% fewer in the following weeks. Similarly, those exposed to a 400ms delay did 0.44% fewer in the first three weeks, and 0.76% fewer thereafter (Brutlag, 2009). This seems to indicate that experienced latency has a large effect on end-user experience, and can affect user activity.

This research, 'An Empirical Study of Spam Traffic and the Use of DNS Black Lists', was motivated by a large observed increase in DNSBL DNS traffic seen on MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) between 2000 and 2004. DNSBL traffic went from 0.4% of all DNS lookups in 2000 to 14.09% of all DNS lookups in 2014 (Jung and Sit, 2004). Three reasons were identified for the observed increase. The first was the marked increase in actual mail traffic to the servers, the second was that spamming hosts were relying on open relays and compromised client machines to deliver the spam instead of sending it directly from the origin machine. The third was the increase in DNSBL services available on the web in that period of time (Jung and Sit, 2004).

2.3.2 DNS Threats and Abuse

'Winning with DNS Failures: Strategies for Faster Botnet Detection' proposes methodologies that utilize NXDOMAIN responses in order to rapidly detect the C&C for a fast-flux botnet. They found that botnets that automatically generated domains to try to reach the C&C would generate many NXDOMAIN replies in a short amount of time (Yadav and Reddy, 2012). Filtering of the data was divided into steps that generated metrics based on the source IP address. Domain entropy is then tested, where generated C&C domains should have a high entropy as they are a randomized distribution of alphanumeric characters (Yadav and Reddy, 2012). Using failure correlation was also suggested, where the entropy of failed domains is compared to successful query domains. The researchers also noted the presence of DNSBL queries, which triggered false positives when their methodology was used (Yadav and Reddy, 2012).

The paper 'An Analysis of Using Reflectors for Distributed Denial-of-Service Attacks' is one of the preliminary works on the ability of reflectors to generate Distributed Reflective Denial of Service

(DRDoS) attacks, and the subsequent threat this creates to network users. The paper discusses threats posed by TCP, UDP and ICMP services with respect to reflected attacks, but for the sake of relevance only their findings with respect to DNS will be discussed here. This paper also suggested possible defenses against reflected attacks, and an overview will also be given of those related to DNS DRDoS attacks.

DNS was identified as offering two possibilities for reflection. The first is for an attacker to spoof packets to DNS servers, which then inundate the victim with DNS replies, whose IP address is the address spoofed by the attacker (Paxson, 2001). Paxson suggests that this can be countered by filtering out packets that use port 53, the port assigned to DNS traffic (Mockapetris, 1987b), at the cost of impeding the access of the victim to DNS via external DNS services. This however can be mitigated by creating holes within the filter through which certain trusted DNS servers can be reached, restoring the victim's DNS capabilities to a certain extent (Paxson, 2001). The second reflective attack is perpetuated using DNS servers that recursively query other servers to resolve requests (Paxson, 2001). This form of reflected attack targets name servers for specific zones, which allows attackers to stream queries to other name servers for the respective zone, which then creates a bombardment of recursive queries towards the target server. This can be further supplemented by spoofing the target server as the requester, ensuring that both queries and replies are used to DoS the victim, and was identified as an early form of amplification (Paxson, 2001). This paper proves the risks generated by reflected attacks, and the need to mitigate them. It also gives insights into the ways that the methods of using reflectors to perform DNS DRDoS attacks have changed over the years, especially when comparing it to more recent works such as Rossow (2014).

A recent and very thorough work on the nature of amplification, 'Amplification Hell: Revisiting Network Protocols for DDoS Abuse' delivers research on the utilization of UDP-based network protocols in disrupting network activity and availability, through the use of DRDoS attacks. These DRDoS attacks are called reflective as the malicious entity uses a third-party infrastructure to launch the attacks against the victim, and does not directly attack the victim themselves. This paper also focuses on reflectors that allow the abuser to amplify the attack through the misuse of UDP protocols, of which DNS is one (Rossow, 2014). While this paper concerns itself with 14 seperate UDP protocols, the focus of this review will be on the results gathered with respect to DNS abuse, as these results are most relevant to the thesis. The data gathered for this paper comes from 130 real-world DRDoS attacks as well as scans captured across two darknets (Rossow, 2014). This paper noted that DNS is an interesting case with respect to amplification protocols, as the number of available amplifiers is known, unlike other protocol amplifiers. This comes about as a result of dedicated projects in existence that track the number of open resolvers that could be used to launch DNS DRDoS attacks. One of these is the Open Resolver Project[2], which has identified over 20 million unique-IP open resolvers currently active on the Internet, of which over 15 million

[2]http://openresolverproject.org/

respond to all queries, indicating that they pose a significant amplification threat (Mauch, 2013). Two ways of evaluating the amplification factor of attacks were suggested in this paper, and are given below.

$BAF = \frac{len(UDP\,payload)\,amplifier\,to\,victim}{len(UDP\,payload)\,attacker\,to\,amplifier}$ $\qquad PAF = \frac{number\,of\,packets\,amplifier\,to\,victim}{number\,of\,packets\,attacker\,to\,amplifier}$

BAF represents the bandwidth amplification factor of the attack while PAF represents the packet amplification factor of the attack (Rossow, 2014). Results on observed amplification factors are further broken down into three levels in the paper. These were the average observed amplification factors of the whole dataset, worst 50%, and worst 10% respectively. The results for DNS amplification are further broken down into ANY lookups at authoritative name servers (NS) and ANY lookups at open resolvers (OR), and are given in Table 2.2. As is seen, the DNS bandwidth amplification factor achieved by abusing name servers was 54.6 times the attacking packet on average, while the open resolver abuse resulted in an average bandwidth amplification of 28.7 times the attacking packet. The packet amplification was 2.08 times of the attacking packet for name servers and 1.32 times for open resolvers.

Table 2.2: Observed average amplification factors

	B	A	F	P A F
Protocol	All	50%	10%	All
DNS (NS)	54.6	76.7	98.3	2.08
DNS (OR)	28.7	41.2	64.1	1.32

This paper noted that most, if not all, of the observed queries attempting DRDoS attacks were ANY queries, which allow attackers to enforce high amplification rates , as resolving ANY queries for domains will result in large responses (Rossow, 2014). This research is relevant to the Post-attack-amplification scanning research seen later in the thesis, as it not only gives a BAF and PAF baseline with which to compare observed results, but also notes various behavioral characteristics of amplification packets that will aid in their identification within the dataset.

'Fingerprinting Internet DNS Amplification DDoS Activities' is a study on using darknet packet captures to infer DDoS activity on the Internet. Traffic captured by a darknet is filtered for possible amplification packet traffic generated by attacks. Of this traffic, queries with the ANY RR set were found to make up the majority of possible amplification packets. It was stated that the increase in ANY traffic seen in darknet space over recent years could be as a result of an increase in amplification attack popularity (Fachkha *et al.*, 2014). Of the domains captured, the most popular was Root, as attackers attempted to request a large amount of zone information to maximize packet amplification (Fachkha *et al.*, 2014). Their analysis showed that DNS amplification attacks would sometimes vary and slow the attack rate to make the attack less detectable (Fachkha *et al.*, 2014). This paper is one of the few examples of amplification attack inference using passive packet collection

The first published paper on the Bitsquatting, 'Bitsquatting: DNS Hijacking without Exploitation' offers comprehensive analysis on the reasons for bitflips, analysis on captured bitflips and recommendations with respect to mitigating the threat presented by bitflips. The three main causes of observed bit-errors were identified as manufacturing defects and contamination; operating outside environmental tolerances and radiation (Dinaburg, 2011). Dinaburg also raises the issue that many manufacturers do not use error checking and correction (ECC) schemes in their hardware, including high-grade mobile devices. The occurrence of flipped bits in the RAM pose a serious security threat when the flipped bit occurs in the domain string. A flipped domain bit will lead to a connection being established with the possibly bitsquat domain, instead of the intended domain, allowing the domain owner to send phishing pages, browser exploits or executable scripts, or make other attempts at compromising the security of the end-host (Dinaburg, 2011). The bitflip analysis found that a high occurrence of queries for a single bitflip showed bit-errors at the responding server, while less frequent and more varied bitflips were usually indicative of end-hosts. The paper also found that certain operating systems were more prone to flipped bits than others. Dinaburg (2011) noted a smaller bitflip presence for Apple OS HTTP User-Agents while a larger number of Other OS HTTP User-Agent bitflips were recored; when compared to average OS User-Agent frequency for visits to Wikipedia. Other OS in this case refers to gaming console and mobile operating systems, as well as less common computer operating systems. The suggestions for bitsquatting mitigation were two-fold. The first was to register all possible bitflips of the domain intended for use, while the second was the adoption of integrity checks, such as Cyclic Redundancy Checks, and ECC Memory to decrease the chances of a bitflip error remaining undetected(Dinaburg, 2011).

The paper 'Bitsquatting: Exploiting Bit-flips for Fun, or Profit?' was written in an attempt to discover if malicious entities were attempting to take advantage of the bitflip behavior reported in Dinaburg (2011). The researchers generated bitflipped domains for the Alexa top 500 domains[3]. They then performed varied analysis from the rate of bitflip site registration to the content served by bitsquatted sites. It was found that 40% of the bitflipped domains were owned by legitimate entities (Nikiforakis *et al.*, 2013). For the most part, the domains were found to be owned by the same organization that owned the top domain investigated. Another 15% of the squatted domains were parked websites, i.e. domains run by domain-parking agencies which serve advertisements relevant to the domain name in order to encourage misdirected users to click on them for revenue (Nikiforakis *et al.*, 2013). Other interesting domain behavior was also noted, 15% of the registered domains redirected to unrelated websites or the websites of competitors, a further 10% of domains listed as 'for sale' on the domain itself, and 6.8% showed advertisements but were not affiliated with a domain-parking agency. Of the investigated domains, 3.2% were serving malware, either

[3] http://www.alexa.com/topsites

through the direct inclusion of malicious scripts, or indirectly through the advertising network present on the site (Nikiforakis *et al.*, 2013). Four defenses against bitsquatting were suggested in this paper. Addressing the problem at the hardware level, through the implementation of CRC and ECC, is the suggested approach. Mitigating errors at the software level is also suggested through the use of the DNSSEC, TLS and/or SSL protocols (Nikiforakis *et al.*, 2013). Another suggested approach is to remove the incentive of Bitsquatting, by implementing legal frameworks that restrict the activity of obviously squatted domains (Nikiforakis *et al.*, 2013). The final option is for the owner of the legitimate domain to register all of the possible bitflips of the domain. It should be taken into consideration, however, that the most frequently resolved domains are more at threat than minor domains, as an increase in queries and responses for any given domain increases the likelihood of a domain being flipped (Nikiforakis *et al.*, 2013). This in effect means that the only domain owners that should consider this defence are those that see enough traffic to mitigate the cost of flipped-domain registration.

2.4 Chapter Summary

This chapter has three main aims. The first was to introduce the reader to certain technical concepts that appear frequently in the thesis, and which need to be understood in order to gain meaning from the findings delivered in the rest of the paper. The most important of these are the following: Authoritative servers exist to serve records for which they are responsible, and will not serve other records. Caching servers will recursively query domains queried through them, unless the domain exists in the cache and has a live TTL, in which case the cached record will form the response packet. DNS TTL values determine the length of time for which a record can remain in the local caching resolver's memory before a new record has to be fetched. Bitflipping is the process by which one or more bits becomes corrupted in memory.

The second aim was to introduce the reader to some of the threats of abuse with respect to DNS. The two most relevant are amplification attacks, whereby attackers will spoof the IP of the target address and flood it with DNS responses larger than the query packets sent to reflectors, and bitsquatting, the targeted domain squatting of domains that differ from popular domains by one bit.

The third aim was to introduce the reader to some of the relevant literature in the field of DNS traffic and DNS threat analysis. The articles cover aspects of the areas discussed later in the thesis, including DNS TTL characteristics, the effects of latency, amplification attack behavior, and the prevalence of bitsquatting on the Internet.

Chapter 3

Data and Data Processing

This chapter discusses the source of the data on which this research is based, as well as an overview of the datasets on which analysis was performed. Data collection is covered in section 3.1. The tools used for data evaluation, analysis and visualization will be discussed, where they were not produced by the researchers themselves, in section 3.2. The preprocessing of the data is handled in section 3.3, followed by a high-level overview of the entire dataset in section 3.4. Section 3.5 discusses some cases of anomalous packet activity seen in the dataset. Sections 3.6 to 3.9 describe the subsets of data that have been isolated from the overall dataset.

3.1 Data Collection

The dataset was collected from a production /24 IPv4 network block which is part of the 196/8 IPv4 network block. The data was gathered from the 1st of October 2013 to the 31st of August 2015. Between this period of time there is one gap within the dataset between April and June 2014. The data was collected by capturing traffic observed, either from or destined to all IP addresses within the /24 block, both TCP and UDP packets, across port 53. The IP block included two authoritative DNS servers and two caching DNS servers within the live end-hosts. Figure 3.1 illustrates the topology of the /24 IPv4 network from where data was gathered. The packet sniffer was connected to the Internet-facing port of the firewall, which was also connected to the /24 IP block. Two authoritative servers as well as two caching servers make up four of the end-hosts within the IP block. The captured packets were subsequently stored as pcap files in a database.

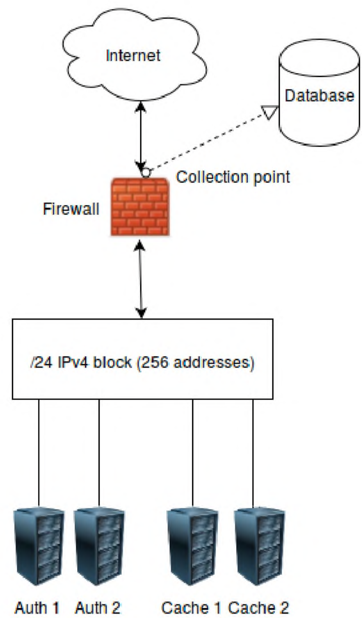


Figure 3.1: Configuration of /24 IP block from which data was collected

3.2 Description of Tools

This section describes the various tools used both in the processing and analysis of the data that were taken from external sources.

3.2.1 Libtins

The C++ libtins[1] (Fontanini, 2015) library, a multiplatform network packet sniffing and crafting library, was used for packet processing of the pcap files. Libtins was selected because of its efficient performance, an important consideration when dealing with large pcap files. Benchmark testing showed that libtins had a faster parsing speed than other well known packet libraries, including dpkt and scapy (Fontanini, 2015). The documentation available with respect to the libtins library also made it a more reasonable choice than libraries such as dpkt.

3.2.2 Python

Python was selected for three reasons as the core programming language. The first is that the language integrates extremely well with unix-based operating systems, on which the majority of this research was done. Python also offers an elegant and simple syntax that promotes code readability and user-friendliness (Sanner, 1999). Third, Python delivers excellent performance with regards

[1] http://libtins.github.io/download/

to parsing, string manipulation and dictionary searches (Prechelt, 2000), which enable faster data processing and analysis during the research process.

3.2.3 Maxmind Geolocate Database

The Maxmind GeoLite databases maintained by Maxmind allow for the mapping of IPv4 addresses to the geographic positions of their end-hosts. It was used in this research in conjunction with the pygeoip[2] library, which is based on Maxmind's GeoIP C API (Ennis, 2015). The GeoLite City database[3] was used in this research.

3.2.4 IPv4 heatmap

The ipv4-heatmap package[4] was created by the Measurement Factory. This allows the mapping of the one-dimensional IPv4 address space onto a two-dimensional image represented using a 12th order Hilbert curve (Irwin and Pilkington, 2008). Each pixel of the generated 4096x4096 image represents a single /24 network containing 256 hosts (The Measurement Factory, 2015).

3.2.5 fping

fping[5] is a tool used for conducting ping sweeps to search for live hosts (Teo, 2000). fping was selected for its functionality, which allowed users to give fping a list of IP addresses instead of pinging each IP seperately. fping also allowed variables to be set regarding the number of pings that would be sent for each IP address given, which allows for a more comprehensive and accurate look at latency averages observed with respect to these IP addresses.

3.2.6 Wireshark

`The editcap`[6] tool is a program designed to read some or all packets from a pcap file, optionally converting or filtering them in various ways before writing the remaining packets to another pcap file. `The editcap tool` was used here to separate the captured pcap files into monthly blocks. The `mergecap`[7] tool is a program designed to merge multiple pcap files into one, and was used to merge pcap files containing different halves of a single month, after they had been filtered using `editcap`.

[2]https://pypi.python.org/pypi/pygeoip/
[3]http://dev.maxmind.com/geoip/legacy/geolite/
[4]http://maps.measurement-factory.com/software/index.html
[5]http://fping.org/
[6]https://www.wireshark.org/docs/man-pages/editcap.html
[7]https://www.wireshark.org/docs/man-pages/mergecap.html

The Wireshark[8] packet inspection tool was also used to inspect packets to allow for more concrete analysis of the findings in this thesis.

3.3 Preprocessing

First `editcap` was used on the available pcap files to separate the datasets into months. `mergecap` was used on datasets separated by `editcap` that had a month split between them in order to concatenate the dataset into monthly pcap files; which were used for subsequent processing.

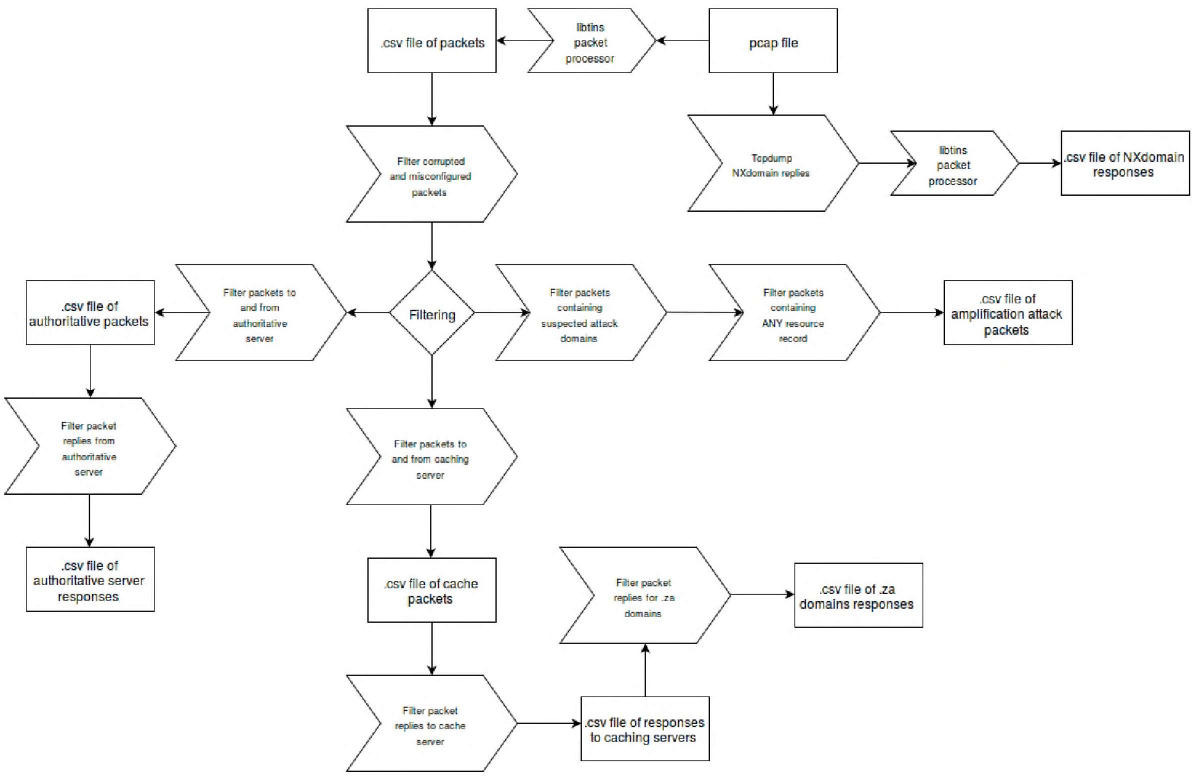


Figure 3.2: Preprocessing method to create datasets for analysis

Figure 3.2 gives a representation of the preprocessing carried out in order to create datasets for analysis. The monthly pcaps were processed using libtins (Fontanini, 2015) to create comma separated value (CSV) data files of the relevant pcaps.

After this certain packets were filtered out. These packets were corrupted as a result of server misconfiguration or corruption during the routing process, which resulted in them being illegible and/or parsing incorrectly. Figure 3.3 shows example output generated by corrupt packets when parsed.

[8]https://www.wireshark.org/

Figure 3.3: Corrupted packet output

There was also a known misconfiguration error for IP 196.x.x.130, within the monitored range, which generated PTR queries directed at two IANA black hole[9] servers, which were also filtered from the dataset. IP fragments were also filtered from the datasets.

From here, the CSV files were filtered using various characteristics to create data subsets. Filtering by source and destination IP address for the four servers resulted in the creation of authoritative and cache server datasets. These datasets were further filtered to create additional datasets comprised only of authoritative server responses, and responses received by cache servers. The responses to the cache servers were further filtered by domain to create a .za response dataset.

Filtering by identified amplification generated a dataset of known or suspected amplification attacks. This was followed by filtering for packets with the ANY RR flag set to remove false positives from the dataset. The subsequent research was performed using these monthly CSV data files and their subsets.

The NXDOMAIN dataset was filtered at the pcap level using `the command below:`

```
tcpdump -n -r <input.cap> -w <output.cap> "udp[11] & 0xf = 3
```

[9]The IANA black hole servers exist to respond to reverse-lookup queries for IP addresses reserved by RFC 1918

where 0xf points to the part of the header that contains the error number space, and 3 is the RCODE for Non-Existent Domain (NXDOMAIN) failures (Eastlake, 2013). The resulting output was then processed into a CSV file via libtins .

The various filtering processes, unless otherwise specified, were performed using tools developed in Python.

3.4 Overview of Dataset

The dataset spans twenty two months between the period of October 2013 and August 2015. Table 3.1 gives information relating to the overall composition of the dataset. It should be noted that the dataset size (given in bytes) is a representation of the size of the pcap files and includes packets that were filtered out as a result of misconfiguration, as mentioned in section 3.3. The number of packets however represents packets in the dataset **post-filtering**, and counts only the packets on which analysis was performed. The total number of packets on which analysis was performed is just under 3.5 billion, comprising some 578 GB of initial data.

Table 3.1: High level view of processed data

Month	# of days	% of hours	# of packets	% of total packets	# of unique IPs	Size (bytes)	% of total bytes
October 2013	31	100	137 792 142	3.940	136 461	23 808 353 832	4.116
November 2013	30	100	133 145 106	3.807	134 638	20 958 712 584	3.624
December 2013	31	100	175 661 225	5.022	116 174	23 101 638 356	3.994
January 2014	31	100	236 963 425	6.775	127 704	31 622 040 972	5.468
February 2014	28	100	155 029 695	4.432	160 289	31 351 807 424	5.421
March 2014	31	100	408 824 999	11.689	164 629	54 340 269 380	9.396
April 2014*	12	39.444	242 632 653	6.937	107 204	31 482 598 264	5.443
May 2014*	3	6.586	2 392 107	0.068	31 243	355 478 064	0.061
June 2014*	25	80.972	111 205 783	3.179	129 837	18 151 581 540	3.139
July 2014	31	100	133 495 938	3.817	137 296	23 787 923 592	4.113
August 2014	31	100	94 691 272	2.707	128 793	15 292 726 124	2.644
September 2014	30	100	155 549 492	4.447	136 745	24 801 018 752	4.288
October 2014	31	100	171 123 957	4.893	162 515	29 751 124 984	5.144
November 2014	30	100	184 681 747	5.280	130 114	31 528 195 160	5.452
December 2014	31	100	80 872 961	2.312	100 525	12 621 737 824	2.182
January 2015	31	100	137 860 035	3.941	126 387	22 717 030 796	3.928
February 2015	28	100	156 387 164	4.471	128 858	26 176 830 264	4.526
March 2015	31	100	178 941 264	5.116	132 041	31 037 853 360	5.367
April 2015	30	100	84 355 017	2.412	109 043	13 610 216 616	2.353
May 2015	31	100	183 408 170	5.244	126 693	32 552 629 472	5.629
June 2015	30	100	173 990 025	4.974	123 125	28 944 171 700	5.005
July 2015	31	100	164 889 418	4.714	129 222	28 368 275 908	4.905
August 2015	31	100	127 261 293	3.638	112 586	21 954 632 708	3.769
Total	625	14936	3 497 665 267	100	722 394	578 316 847 676	100

* datasets do not represent a complete monthly capture.

Most of the months in Table 3.1 are complete from 00:00:00 on the 1st to 23:59:59 on the last day of the month, as is indicated by the number of hours represented in each dataset. Exceptions to this are April, May and June of 2014, due to a gap in the available data. As a result of this, April and May 2014 have been excluded from all analysis in order to not let incomplete datasets skew the results. After consideration, June 2014 was included as it was able to record days 6 through 30 in the dataset, and is considered mostly complete.

Figure 3.4 gives a Hilbert Curve representation of the IPv4 IP space (Irwin and Pilkington, 2008) observed across the entire dataset. Here it can be seen that there is a distributed IP presence within the dataset itself, representing communication between many different subnets with the /24 IPv4 subnet from which the data is drawn.

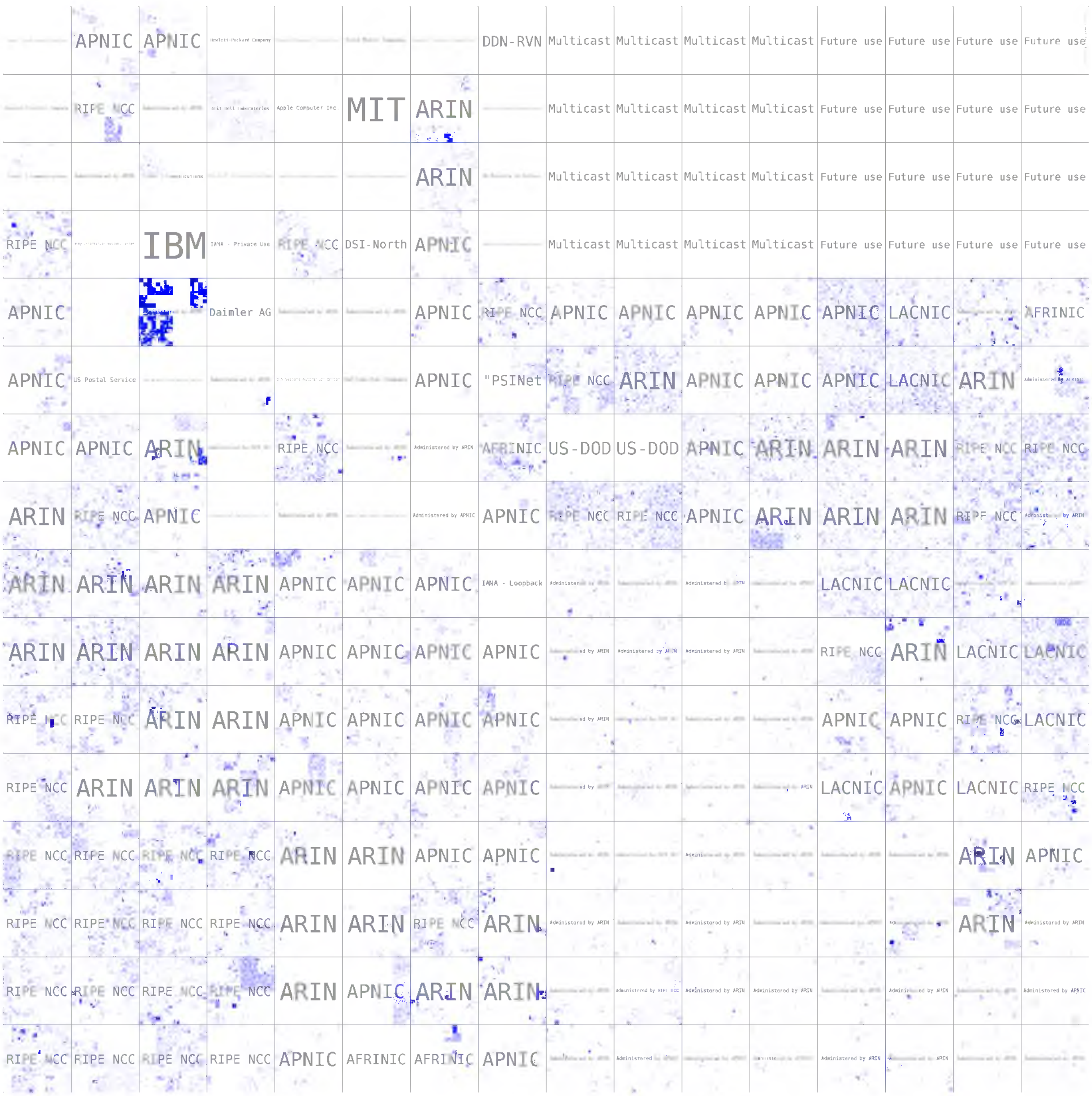


Figure 3.4: IPv4 Hilbert Curve of IP addresses in dataset

The meaning behind this visualization of IP addresses,as well as the software used to produce it, is discussed in section 3.2.4. This heatmap uses the IANA IPv4 Space Registry overlay, showing the registries of the various IP blocks. This allows for easy identification of which registries are communicating with the observed IP block.

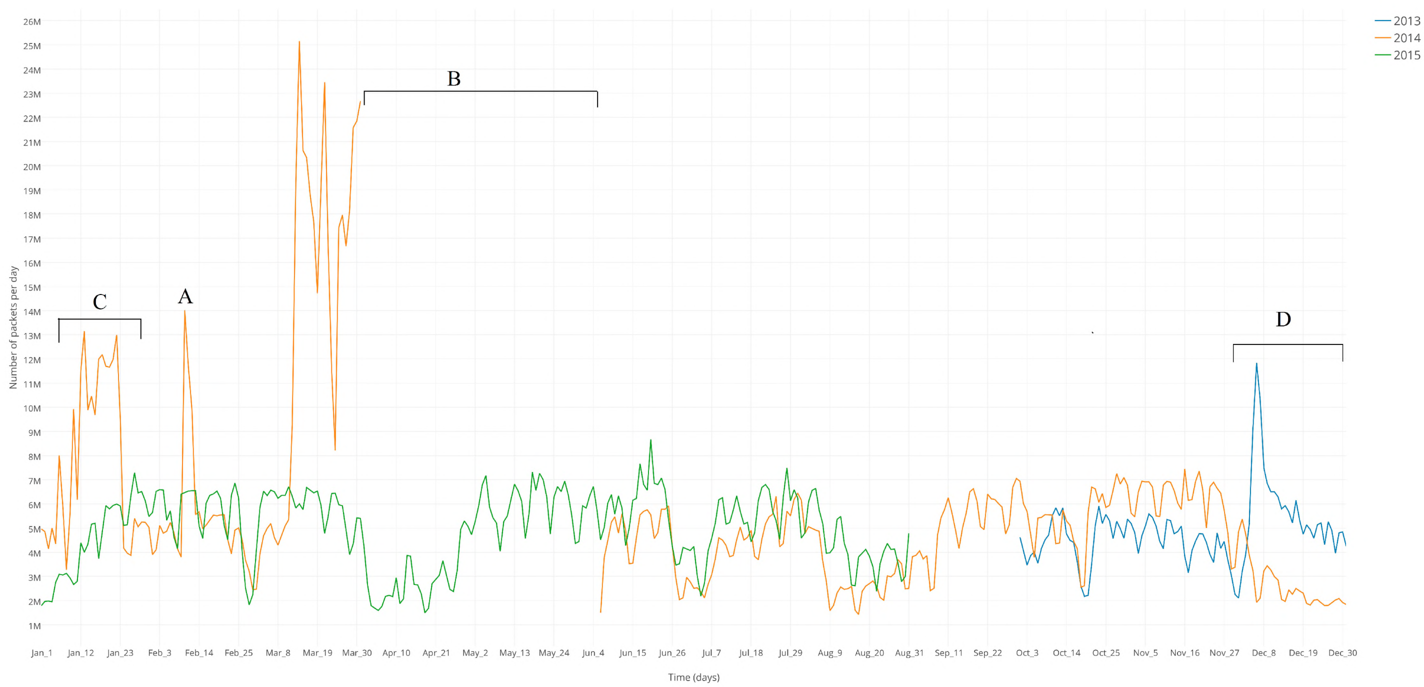


Figure 3.5: Time series of packets from 1 October 2013 to 31 August 2015

Figure 3.5 is a time series of packet traffic across the entire dataset. While the first 12 days of April and first 3 days of May 2014 were recorded, they were omitted from the time series as they will not be included in the discussions on the dataset and analysis to follow. This then leaves a gap in the time series from 1 April to 6 June 2014, marked as B in figure 3.5. The packet presence is noticeably larger for March 2014, which is expected as it comprises the largest percentage of the total dataset. There is a trailing spike in traffic, marked D, observed in December 2013, an increase in packet frequency across the month of January 2014, marked C, and a large singular spike in traffic on the 10th of February 2014, labeled A. The latter will be further looked at in section 3.5.

3.5 Observations of packet behavior across dataset

This section deals with some anomalous packet behavior observed across the dataset. While this is not completely relevant to the analysis that will follow this chapter, the captured activity is presented so that the reader may better understand why some observed values in the datasets are counter-intuitive.

3.5.1 Traffic spike observed 10 February 2014

Figure 3.6 shows a packet time series for the month of February. The month itself only holds 155 million packets so the fact that one day's worth of traffic would comprise nearly 10 percent of the total traffic, at over 14 million packets, stands out.

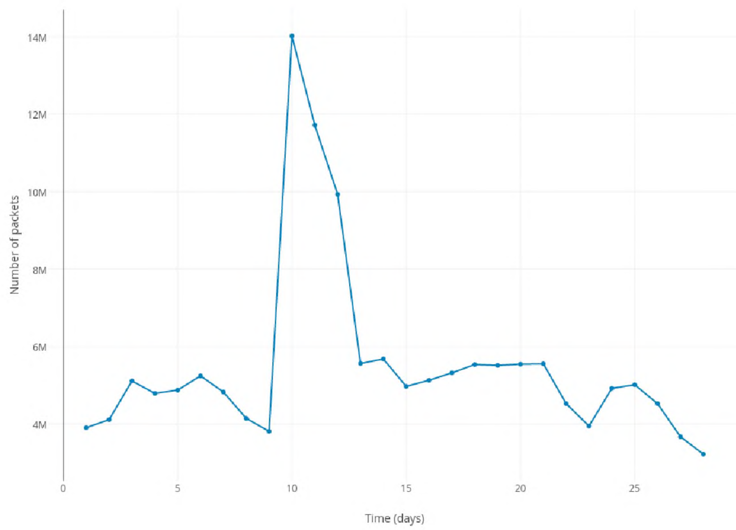


Figure 3.6: Timeseries of packets captured for February 2014

Analysis of the day in question led to the following findings. A significant amount of traffic was caused by a server misconfiguration of one of the end-hosts in the IP block, which resulted in malformed packets constantly being sent to 108.61.239.225. This accounted for just over 9 million packets during the course of the day. No replies were received from any of the IP addresses within the 108.61.239/24 IP block.

3.5.2 Shift in authoritative and caching packet presence for May - July 2015

There is a notable increase in the overall presence of authoritative traffic coupled with a decrease in the presence of caching traffic between May and July of 2015.

Table 3.2: Top 10 source IP blocks seen for April 2015 authoritative and caching datasets

		/16				/24		
Rank	Authoritative	IPs in block	Cache	IPs in block	Authoritative	IPs in block	Cache	IPs in block
1	192.221	3597	205.251	2041	8.0.6	255	205.251.199	254
2	8.0	2941	192.185	728	192.221.163	255	205.251.198	254
3	66.249	1466	156.154	361	192.221.162	255	205.251.197	254
4	61.220	1208	173.245	359	192.221.151	255	205.251.196	254
5	74.125	281	192.254	240	216.40.44	253	205.251.192	254
6	216.40	253	216.21	124	8.0.18	251	205.251.195	253
7	52.12	217	193.108	111	8.0.16	251	205.251.193	251
8	173.252	194	208.76	103	66.249.76	251	205.251.194	251
9	151.164	194	217.160	101	8.0.23	246	193.108.91	192
10	12.121	170	50.87	100	8.0.10	240	173.245.58	179

Table 3.2 gives a /24 and /16 IP block breakdown of IP addresses communicating with the two servers. This table is a strong representation of trends seen in most of the other months. The 192.221/24 and 8.0/24 IP blocks usually contribute the most unique IP addresses to the authoritative dataset, while the 205.251/16 IP block dominates IP presence in the caching dataset. The IP addresses communicating with the authoritative servers are queries for domains while the IPs communicating with the caching servers are query responses.

Table 3.3: Top 10 source IP blocks seen for May 2015 authoritative and caching datasets

		/16				/24		
Rank	Authoritative	IPs in block	Cache	IPs in block	Authoritative	IPs in block	Cache	IPs in block
1	192.221	3231	205.251	2043	205.251.197	254	205.251.199	254
2	8.0	2656	192.185	918	205.251.195	254	205.251.198	254
3	205.251	2045	156.154	404	205.251.194	254	205.251.197	254
4	66.249	1310	173.245	399	205.251.193	254	205.251.196	254
5	61.220	1050	192.254	319	205.251.199	253	205.251.195	254
6	192.185	717	216.21	268	205.251.198	253	205.251.194	254
7	173.245	362	184.154	232	205.251.196	253	205.251.193	254
8	156.154	352	50.87	222	192.221.162	253	205.251.192	254
9	192.254	273	193.108	219	205.251.192	252	173.245.58	200
10	173.252	254	208.76	204	8.0.18	249	173.245.59	195

Only one month later, the 205.251/16 IP block has an equally strong presence in both the authoritative and caching IP contributors, and ranks third overall for unique IPs seen for authoritative servers, despite not being a top IP contributor in any of the previous datasets. The 205.251/16 IP block held seven out of ten positions for /24 IP blocks in June, as well as retaining its third rank for /16 IPv4 IPs contributed. July saw the IP block holding eight of the top ten /24 positions while ranking first for overall IP contribution by a /16 IPv4 block. This is not continued by the August 2015 dataset however, which retains characteristics similar to the IP contributors of previous months.

Packet flows indicated that during this time one of the authoritative servers, 196.x.x.75, was acting as a caching server by sending queries to and receiving responses from the IP blocks in question.

It is suspected that this was configured as such because of one of the caching servers, 196.x.x.77, being offline. This was most likely done in order to balance the cache server load, which is more strenuous than the load on the authoritative servers.

3.6 Overview of Authoritative Dataset

The authoritative dataset was filtered from the total dataset as it forms part of the TTL dataset seen in section 4.1. This was done by filtering the dataset for packets that had a source or destination IP address that belonged to one of the authoritative servers in the observed IP block. Table 3.4 gives a summary of the authoritative datasets filtered from the CSV files. It should be noted here that the size (in bytes) is of these files and not the .cap files, and the accompanying percentages are also calculated against the CSV file size for that month. This holds true for all subsequent data size comparisons. Authoritative servers hold domain records, and are queried by other end-hosts for information on those domains. The authoritative dataset was created by filtering for packets destined to or sent from the IP addresses of the two known authoritative servers.

Table 3.4: High level view of processed Authoritative data

Month	# of packets	% of total monthly packets	# of unique IPs	Size (bytes)	% of total monthly bytes
October 2013	2 993 563	2.137	42 891	329 016 262	2.442
November 2013	4 050 830	3.042	46 571	446 009 496	3.393
December 2013	3 661 817	2.085	45 287	402 344 104	2.453
January 2014	3 971 267	1.676	46 216	436 145 491	1.939
February 2014	6 133 090	3.956	51 652	653 935 585	4.286
March 2014	5 392 802	1.319	56 606	589 739 291	1.521
June 2014*	4 285 626	3.854	50 799	466 891 814	4.105
July 2014	5 022 776	3.762	54 063	548 088 455	3.895
August 2014	5 307 639	5.605	53 068	578 293 706	5.749
September 2014	5 881 709	3.781	55 223	641 799 056	3.859
October 2014	6 341 608	3.706	59 276	690 493 931	3.646
November 2014	5 478 611	2.967	52 681	596 528 112	2.944
December 2014	4 741 805	5.863	48 603	515 800 706	6.101
January 2015	5 896 603	4.277	53 040	646 006 657	4.257
February 2015	7 779 344	4.974	54 703	855 775 895	4.795
March 2015	7 228 079	4.039	55 186	794 857 236	4.136
April 2015	7 043 041	8.349	51 137	775 126 280	8.690
May 2015	34 266 603	18.683	92 935	3 851 282 313	20.050
June 2015	27 022 755	15.531	81 153	2 997 293 863	16.624
July 2015	16 091 494	9.759	76 526	1 776 308 825	10.078
August 2015	6 261 108	4.919	41 184	688 459 139	5.045
Total	174 852 170	4.999	344 445	19 280 196 217	5.715

* datasets do not represent a complete monthly capture.

While there is usually a large unique IP presence in the authoritative datasets, there is only a small packet percentage presence for most of the months. This does not hold true, however, for the months of May, June, and - to a lesser extent - July of 2015, where recorded unique IP addresses as well as packet percentage representation are far above other observed values.

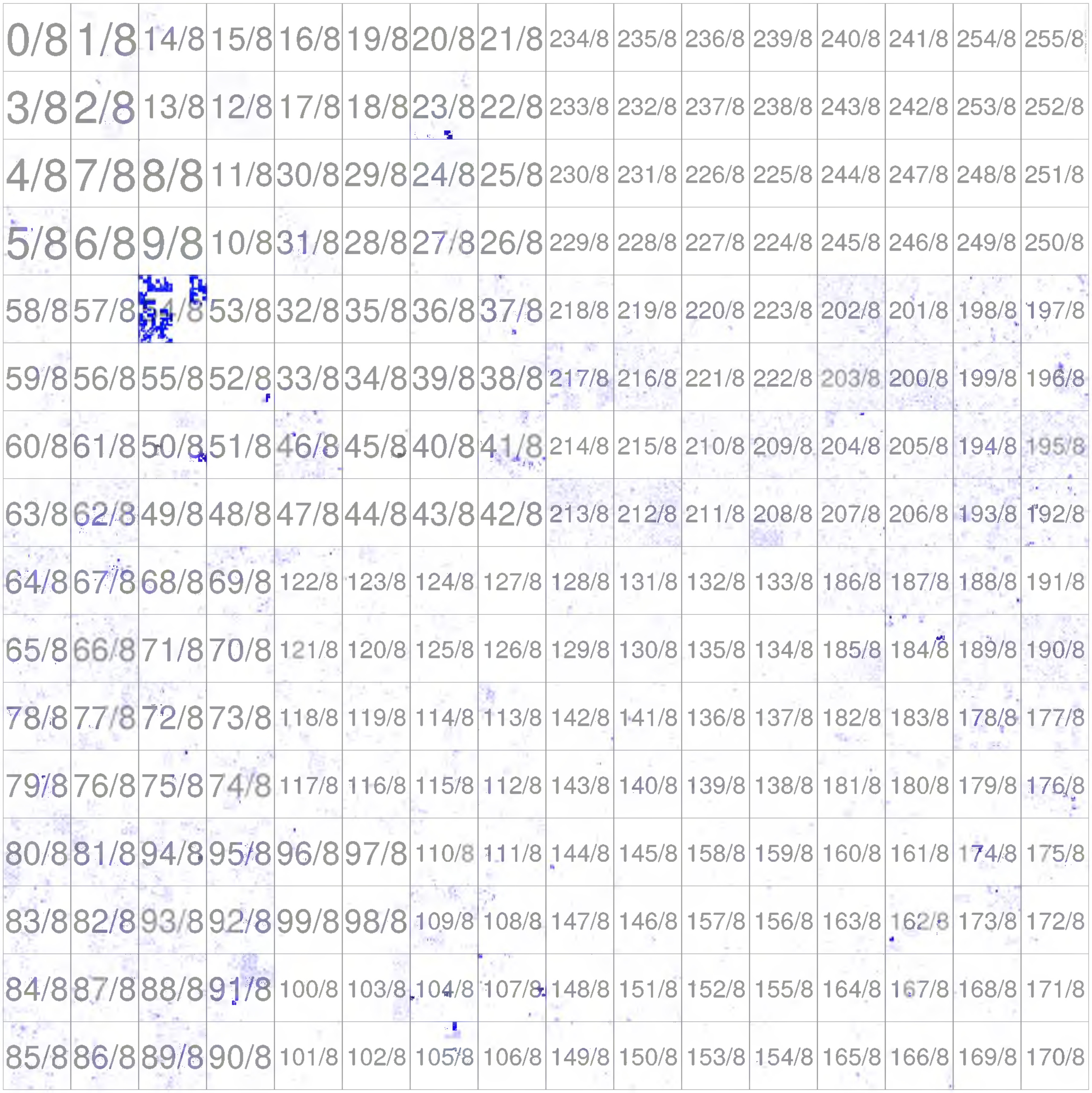


Figure 3.7: IPv4 Hilbert Curve of client IP addresses in dataset

The reasons for this are discussed in section 3.5. The authoritative dataset overall holds a mere 175 million packets, much smaller than the captured cache dataset. This is to be expected as authoritative servers will generally see less traffic than their caching resolver counterparts unless they serve very low TTLs coupled with the fact that they are authoritative for very popular domains. It is this that makes the large authoritative presence between May and July so anomalous, as the packet percentage presence of the two servers actually approach one another.

The IPv4 Hilbert curve of observed addresses in figure 3.7 shows a spread across most of the /8 IPv4 blocks. The 54/8 presence is much more pronounced than other blocks, and indicates that the 54/8 presence in the overall Hilbert curve can be attributed to traffic to and from the authoritative servers. Interestingly, none of the /24 IP blocks in the 58/8 IP block contribute a significant number of IP addresses to the authoritative cache. The /16 presence for the 58/8 IP block is more notable, and some of the /16 IPv4 blocks will occasionally rank in the top ten contributors of IP addresses to the authoritative dataset.

Table 3.5 gives a breakdown of observed /24 and /16 IP blocks within the authoritative dataset for July 2014. While there is no single large /24 block IP contributor presence from the 58/8 IP block, there are a number of IP addresses from different /24 IP blocks that create a larger /16 IP presence, with respect to unique IP addresses.

Table 3.5: Top 10 source IP blocks seen for authoritative datasets July 2014

IP block size	/16		/24	
Rank	Authoritative	IPs in block	Authoritative	IPs in block
1	192.221	4356	192.221.150	253
2	8.0	3447	66.249.74	252
3	66.249	1314	8.0.15	249
4	61.220	1305	192.221.151	249
5	74.125	371	192.221.143	249
6	54.90	219	192.221.139	246
7	54.203	187	192.221.138	246
8	54.91	185	192.221.134	245
9	54.89	183	192.221.167	241
10	54.74	183	66.249.66	240

Amazon technologies does have a number of /10, /11, /12 and /13 subnets in the 54/8 IPv4 block, while the block itself is administered by ARIN. This traffic is most likely generated by **Amazon Web Services** ,through third parties using their cloud hosting software, which are trying to reach domains for which the server has authoritative records.

3.7 Overview of Cache Dataset

The cache dataset was filtered from the total dataset as sections 4.1 and 4.2 are made up of this dataset. Section 4.2 more so, as authoritative replies to the two caching servers for .za domains were filtered out in order to gather the necessary data for authoritative server geolocation. An overview of the caching dataset is given in Table 3.6. The caching dataset represents a much greater packet percentage when compared to other datasets. There are two reasons for this, the first is that the filtering criteria is broader than the NXDOMAIN and Amplification filters, as it targets any packet to and from the caching resolvers, the second is that the caching servers saw more traffic on average than the authoritative servers, as they were not authoritative for extremely popular domains, while popular domains were constantly queried through the caching resolvers. This dataset also contains the largest number of unique IP responses as the caching servers receive response packets from authoritative servers across the Internet.

Table 3.6: High level view of collected data

Month	# of packets	% of total monthly packets	# of unique IPs	Size (bytes)	% of total monthly bytes
October 2013	51 121 777	37.101	81 053	6 148 029 736	45.627
November 2013	53 454 301	40.147	76 908	6 398 171 422	48.674
December 2013	82 793 019	47.132	57 572	9 597 971 528	58.520
January 2014	109 809 929	46.340	72 906	12 781 938 116	56.816
February 2014	56 686 532	36.565	100 948	6 781 035 956	44.448
March 2014	188 760 346	46.171	98 549	21 941 444 555	56.579
June 2014*	43 951 343	39.523	75 569	5 283 291 552	46.455
July 2014	42 028 142	31.483	70 846	5 078 619 295	36.089
August 2014	35 342 637	37.324	68 431	4 255 975 551	42.011
September 2014	64 346 538	41.367	73 708	7 754 985 986	46.627
October 2014	59 127 006	34.552	93 967	7 183 009 480	37.924
November 2014	65 954 042	35.712	71 748	7 963 868 754	39.301
December 2014	30 867 320	38.167	47 534	3 661 454 337	43.310
January 2015	50 346 994	36.520	69 706	6 070 039 859	40.001
February 2015	52 128 094	33.332	71 509	6 359 485 172	35.632
March 2015	60 244 649	33.667	72 768	7 357 672 570	38.288
April 2015	27 035 565	32.050	51 087	3 285 589 272	36.834
May 2015	40 508 020	22.086	60 877	4 902 267 514	25.521
June 2015	47 264 942	27.165	64 449	5 863 556 885	32.521
July 2015	49 546 971	30.048	70 570	6 125 539 381	34.755
August 2015	34 040 395	26.748	59 828	4 144 218 993	30.369
Total	1 245 358 562	35.605	346 316	148 938 165 914	41.729

* datasets do not represent a complete monthly capture.

There is a notable dip in percentage packet representation for May, June, and - to a lesser extent - July 2015. This is linked to the larger authoritative presence noted over that period and will be discussed further in 3.5. Overall the caching dataset holds just under 1.25 billion packets, and represents around 35% of all packets captured in the dataset.

As is seen in figure 3.8, there is a smaller but noteworthy 54/8 IPv4 presence in the caching dataset. This is as a result of **Amazon** cloud authoritative servers existing in this IP block, of which they control certain sub-blocks as mentioned previously.

Figure 3.8: IPv4 Hilbert Curve of IP addresses in dataset

This Hilbert curve plot indicates a good spread of communication across IPv4 address space at the /8 level captured in these datasets, as almost all of the /8 IP blocks, not including those reserved for future use, are populated to some extent. When compared to the authoritative Hilbert curve in figure 3.7, there seems to be fewer densely populated clusters, like the 54/8 IP block, but overall traffic from many of the blocks seen in the authoritative set. Figure 3.8 also shows an increase in smaller concentrated clumps of IPs when compared to figure 3.7, particularly in the 173/8 to 193/8 IP range.

This heatmap and the Authoritative heatmap in figure 3.7 use the /8 IP block overlay instead of the IANA registry overlay seen in figure 3.4. This is done to show the /8 IP block distribution of the Hilbert curve, as well as to make the figures more meaningful to the reader.

3.8 Overview of Amplification Dataset

The amplification dataset is the smallest of the filtered whole datasets to appear in this thesis. It was filtered from the dataset using reported attack domains[10] as an identifier. The packets were then further filtered, accepting only ANY RR packets, in order to remove false positives. These false positives were generated as some of the attack domains are legitimate domains and as such generate non-amplification query traffic. Table 3.7 describes the amplification presence in the monthly datasets. The largest amplification presence amounts to merely 0.222% of a monthly dataset. One of the key reasons for the dataset being so small is that there are no open resolvers present in the 196.x.x.x/24 IP block for these scans to take advantage of, and as such no response packets have been recorded. The dataset will be discussed further in section 5.1.

Table 3.7: High level view of collected data

Month	# of packets	% of total monthly packets	# of unique IPs	Size (bytes)	% of total monthly bytes
October 2013	306 364	0.222	85	31 030 187	0.230
November 2013	102 010	0.077	22	10 405 133	0.079
December 2013	37 298	0.021	13	3 882 634	0.024
January 2014	21 818	0.009	11	2 215 590	0.010
February 2014	52 700	0.034	16	5 262 240	0.034
March 2014	44 505	0.010	15	4 669 692	0.012
June 2014*	6 009	0.005	15	629 147	0.006
July 2014	9 952	0.007	21	1 013 588	0.007
August 2014	13 666	0.014	25	1 390 746	0.014
September 2014	10 902	0.007	20	1 107 062	0.007
October 2014	7 900	0.004	16	822 639	0.004
November 2014	6 148	0.003	16	632 665	0.003
December 2014	6 593	0.008	25	669 385	0.008
January 2015	4 338	0.003	16	440 621	0.003
February 2015	4 199	0.003	14	429 931	0.002
March 2015	4 974	0.003	15	503 684	0.003
April 2015	2 583	0.003	7	264 280	0.003
May 2015	3 044	0.002	7	312 795	0.002
June 2015	4 095	0.002	10	418 065	0.002
July 2015	2 793	0.002	5	284 768	0.002
August 2015	978	0.001	5	99 119	0.001
Total	652 869	0.019	325	66 483 971	0.011

* datasets do not represent a complete monthly capture.

[10] http://dnsamplificationattacks.blogspot.co.za/

3.9 Overview of NXdomain dataset

The NXDOMAIN dataset contains all responses with the NXDOMAIN error flag (Andrews, 1998) set, indicating that the queried domain does not exist. Overall, this dataset accounts for just under 143 million packets. This dataset was filtered using the `tcpdump` command mentioned in section 3.3. NXDOMAIN responses were filtered into a seperate dataset as these responses are usually indicators of anomalous activity (Yadav and Reddy, 2012). It must be taken into account that only the replies are captured here; queries have not been included in the filtered dataset. This dataset was filtered using tcpdump, as mentioned in section3.3. Further analysis on NXDOMAIN traffic is done in section 4.3.

Table 3.8: High level view of collected data

Month	# of packets	% of total monthly packets	# of unique IPs	Size (bytes)	% of total monthly bytes
October 2013	4 800 190	3.484	17 442	769 302 413	3.231
November 2013	4 341 797	3.261	16 927	692 394 489	3.304
December 2013	4 298 557	2.447	19 616	696 180 506	3.014
January 2014	4 698 313	2.017	17 853	761 163 116	2.407
February 2014	4 780 242	3.083	20 722	768 147 593	2.450
March 2014	5 803 551	1.420	19 837	944 724 915	1.739
June 2014*	3 199 786	2.877	21 026	503 612 858	2.774
July 2014	4 429 901	3.318	23 619	815 629 109	3.429
August 2014	4 040 682	4.267	20 643	702 114 083	4.591
September 2014	4 768 943	3.066	23 472	807 022 101	3.254
October 2014	6 431 531	3.758	24 706	1 106 018 089	3.718
November 2014	6 530 777	3.536	18 552	1 185 156 713	3.759
December 2014	6 034 674	7.462	14 774	1 030 469 783	8.164
January 2015	7 490 611	5.433	17 978	1 341 201 081	5.904
February 2015	8 210 950	5.250	19 022	1 445 870 257	5.523
March 2015	9 155 147	5.116	18 718	1 531 014 630	4.933
April 2015	7 488 729	8.877	14 978	1 244 785 242	9.146
May 2015	8 779 328	4.787	19 389	1 527 449 716	4.692
June 2015	13 503 773	7.761	20 057	2 500 045 944	8.637
July 2015	14 066 828	8.531	22 158	2 537 062 858	8.943
August 2015	9 868 415	7.754	18 911	1 707 882 116	7.780
Total	142 722 725	4.081	121 379	24 617 247 612	4.257

The * notes datasets that do not represent a complete monthly capture.

3.10 Chapter Summary

This chapter discusses the dataset, the filtering and preprocessing of the data, and gives overviews of the dataset as well as the subsets created for more focused analysis.

A number of tools were used in the processing and filtering stage of the research. The most notable of these are libtins, a C++ packet parsing library with which the pcap reader was written;

Wireshark, a packet sniffer and analysis UI, as well as its derivative tools `editcap` and `mergecap`, which aided in the processing of the pcap files; fping, a ping sweep tool which was used to determine authoritative server latency; the MaxMind Geolocate Database which was used in order to correlate IP addresses to the countries in which they are based.

The total dataset spans 21 months and holds close to 3.5 billion packets. From this dataset, four separate subsets were formed that are used in three of the analysis sections. The authoritative and caching subsets are analyzed for TTL implementation and behavior. Analysis on the NXDOMAIN reply dataset forms its own section, as does analysis on the amplification dataset on post-attack amplification scanning. The section on bitflipping and bitsquatting utilizes the entire dataset.

Chapter 4

DNS Operations

This chapter deals with three areas related to the practice of DNS implementation and usage captured by the dataset. Section 4.1 gives a breakdown of observed DNS TTL values across the dataset. Section 4.2 looks at the geolocation of authoritative servers for queried .za domains, as well as the latency generated for .za queries as a result of the location of the authoritative servers; from a South African context. Section 4.3 looks at NXDOMAIN responses for queries captured over the various months.

4.1 Breakdown of DNS response TTLs

DNS time-to-live values are implemented to instruct servers caching the responses of DNS resolvers as to how long the record should remain viable within the cache memory. Once this value times out, the cache is instructed to query the authoritative server of the domain instead of replying to queries with the cached data. This section gives a breakdown of observed TTL values across the dataset.

4.1.1 Observed TTL frequency

Table A.1 gives a ranked breakdown of the frequency of DNS TTL values observed throughout the dataset, and is found in the appendix. Figure 4.1 illustrates the ranked position of the TTL values throughout the dataset. The results here are slightly skewed in favor of lower TTL values. This is because lower TTL values result in more queries to the respective authoritative server of the domain, as the cached records become stale faster. As a result, more replies with low DNS TTL values present. While the presence of lower TTL values is then expected, it would also be expected that the lowest TTL value would rank the highest given an equal distribution of TTL value configurations, which is not seen in this case.

September 2014 stands out as it is the only month to have a TTL value of 1 in the top 10 ranking. This is even more curious as it would be expected that if the TTL was related to commonly queried domains that it would appear in more, or maybe even all, months given the nature of its low TTL. Further analysis revealed that most of the 1 value TTL values were linked to query responses for sc.xx.rules.mailshell.net, where xx is a placeholder of two numeral characters and not part of the domain itself. These queries were linked to botnet activity from an unidentified botnet by Kwon *et al.* (2014). Six subdomains of *.rules.mailshell.net*, sc21, sc18, sc19, sc17, sc1 and sc2, contributed almost all of the 1 TTL packets observed in the dataset. While these domains are present in other months, they are not nearly as large as the presence recorded in September 2014. This botnet was used to launch DDoS attacks towards the end of September 2014, as is seen in figure 4.2.

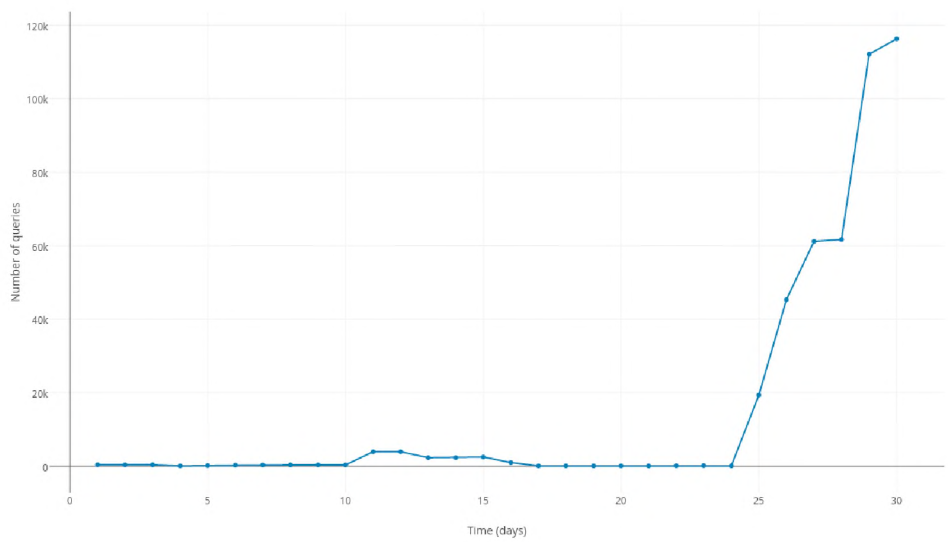


Figure 4.2: Queries for sc.xx.rules.mailshell.net domains during September 2014

Figure 4.2 is a timeseries of queries for the aforementioned domains from one host in the dataset, 196.x.x.162, a known proxy for a local network, during September 2014. This IP, among others, targeted five servers with more than 10 000 queries and a number of others with queries totaling less than 1000 of these. One of the target end-hosts, IP 155.232.135.5 was the most affected victim, receiving over 300 000 queries from 196.x.x.162 alone. It should be noted here that the mailshell.net TLD has been identified in other research as part of the DNSBL infrastructure (Metcalf and Spring, 2014)

4.1.2 Normalised TTL frequency

Table 4.1 gives the frequency of DNS TTL values after the data has been normalised in order to remove duplicated responses. This is done in order to counter the frequency skewing that comes about as a result of low TTL values generating more response queries than cached records with higher TTL values, given the same query frequency on the network (Jung *et al.*, 2002).

Table 4.1: Normalised TTL frequency

Rank	1		2		3		4		5		6		7		8		9		10	
Month	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs
October 2013	86400	21.973	300	17.529	3600	13.406	7200	9.153	900	8.824	14400	5.290	600	3.009	43200	2.724	1800	2.370	21600	1.957
November 2013	300	21.566	86400	20.279	3600	13.680	7200	8.798	900	6.486	14400	5.023	600	3.081	1800	2.555	43200	2.459	28800	1.860
December 2013	86400	34.433	3600	13.392	300	11.956	7200	7.687	900	7.189	43200	4.614	172800	2.604	14400	1.954	28800	1.944	0	1.667
January 2014	300	20.911	86400	19.836	3600	12.254	900	8.948	7200	7.342	0	5.907	14400	4.588	43200	2.890	600	2.270	1800	1.951
February 2014	86400	25.698	300	13.842	3600	13.261	900	10.579	14400	6.898	7200	5.933	43200	3.936	0	3.481	172800	2.457	600	1.871
March 2014	86400	18.568	300	18.195	900	13.295	3600	11.518	14400	8.471	0	5.198	3200	3.406	43200	2.978	7200	2.425	600	2.148
June 2014	86400	18.025	300	14.145	3600	14.063	900	11.490	14400	6.904	28800	6.486	0	4.667	600	3.097	7200	2.893	1800	2.814
July 2014	86400	17.545	300	17.424	3600	16.133	900	9.733	14400	5.877	28800	4.388	600	3.522	7200	3.106	1800	3.010	0	2.935
August 2014	300	22.596	86400	16.861	3600	14.004	900	10.775	14400	6.668	600	3.150	1800	3.065	0	2.713	7200	2.659	21600	2.276
September 2014	300	24.476	86400	15.412	3600	14.133	900	8.188	0	6.623	14400	5.584	600	3.597	1800	3.040	7200	2.741	60	2.264
October 2014	300	20.111	3600	16.558	86400	16.486	900	8.975	14400	8.080	600	4.019	7200	3.478	1800	3.206	60	3.128	43200	2.016
November 2014	300	25.508	86400	16.346	3600	14.805	900	9.760	14400	6.042	600	3.799	60	3.514	7200	2.759	1800	2.742	43200	1.913
December 2014	86400	183813	300	18.514	900	15.185	3600	14.228	14400	5.090	60	3.947	600	3.449	28800	2.786	7200	2.563	1800	2.519
January 2015	300	20.420	86400	17.445	3600	14.423	900	13.541	14400	6.722	60	3.655	600	3.253	43200	2.534	1800	2.440	7200	2.436
February 2015	300	25.710	86400	16.179	3600	14.497	900	9.796	14400	6.413	60	3.768	600	3.226	1800	2.663	7200	2.596	43200	2.219
March 2015	300	26.166	86400	15.842	3600	15.152	900	9.891	14400	6.639	60	3.773	600	3.335	7200	2.510	1800	2.497	28800	2.117
April 2015	300	22.285	86400	16.633	3600	15.108	900	11.951	14400	5.128	60	4.773	600	3.575	7200	2.683	1800	2.547	28800	2.483
May 2015	300	25.199	3600	15.830	86400	15.418	900	9.691	14400	5.686	60	4.294	600	3.822	21600	3.406	7200	2.619	1800	2.604
June 2015	300	21.290	86400	19.459	3600	15.412	900	10.389	14400	4.978	60	3.385	600	3.569	21600	2.985	7200	2.736	1800	2.265
July 2015	86400	24.266	300	16.443	3600	16.054	900	9.956	14400	4.551	600	3.426	60	3.340	7200	2.959	21600	2.700	43200	2.528
August 2015	86400	22.889	3600	14.821	900	13.082	300	12.629	14400	4.411	60	4.228	600	3.549	21600	3.245	7200	2.826	43200	2.316

It is interesting to note that apart from the 86400 (one day) TTL value, none of the TTL values that rank in the top 4 exceed two hours, and from January 2014 none of them exceed one hour. This is a clear indicator that domain administrators are favoring lower TTL values, as noted by Gao *et al.* (2013). It seems clear when looking at table 4.1 that the 300, 86400, 3600 and 900 TTL values are most favored as domain TTL values.

One of the largest contributors to the 300 TTL values seen in the dataset are domains related to Google. The *gstatic.com, google.com, googlevideo.com* and *googlehosted.com* top-level domains all contributed a number of subdomain responses with TTL values of 300. The Akamai CDN family also responded with 300 TTLs for subdomains for three TLDs, *akadns.net, akamaihd.net* and *edgesuite.net*. Other notable corporate contributors of the 300 TTL are *yahoo.com, yahoodns.net, skype.net* and *photobucket.com*. This seems to indicate that Internet-based organizations are making use of lower TTLs in an attempt to better control the distribution of connections between servers, much like a CDN, while also enabling quick configuration changes to ensure minimal downtime due to server failure.

The largest contributor of 86400 TTL values are responses for PTR queries. This value is recommended in RFC 1035 (Mockapetris, 1987b) and further defined as the default value for PTR and other RR types not subject to constant change in RFC 1912 (Barr, 1996). There is also a notable presence of *apple.com* and *icloud.com* subdomains that respond with 86400 TTLs. Some *google.com* subdomains also responded with the one day TTL, but less so than those that responded with 300 TTLs.

Microsoft domains, most notably the *windows.com, hotmail.com* and *live.com* TLDs, were some of the largest contributors of 3600 TTL values. The *mcafee.com* subdomains were also a large contributor of this TTL.

The largest contributor of 900 TTL responses was *spamhaus.org*, a Domain Name System Black List (DNSBL) (Jung and Sit, 2004). A DNSBL allows mail recipients to query sending hosts against the list, filtering out known spam hosts (Jung and Sit, 2004). The queries were either A or TXT RR queries, taking the form of {IPv4address}.*zen.spamhaus.org*. For the month of January 2015 alone, *spamhaus.org* subdomains were responsible for 88.719% of unique 900 TTL responses captured on the dataset.

There were no significant contributors for 600 TTL domains, which showed a less concentrated spread of domains than other TTL values. Four main contributors of this TTL were *xvideos.com* CNAME queries, as well as various PTR queries that make up around 5% of the 600 TTL responses, while *softonic.com* and *avast.com* subdomains each make up 2.5% of the 600 TTL replies respectively.

The 60 TTL presence was relatively interesting, as most of the unique subdomains that responded with this TTL fell under the *mailspike.net* and *spamhaus.org* domains. There was also a noticeable presence from akadns.net CDN domains, as well as subdomains present for *googlevideo.com* and *amazonaws.com*, indicating that these domains are mirroring CDN configuration and behavior. The Facebook CDN - *fbcdn.com* - also responded with a number of 60 TTL replies. CDN TTL values are typically low to allow the CDN to change domain mapping quickly in order to facilitate server load balancing (Krishnamurthy *et al.*, 2001).

Gao *et al.* (2013) stated in their paper that observed TTL values have decreased when compared to past research, and papers cited by them backed their findings. This research shows close results to that paper, showing that TTL values at or below one hour are far more prevalent than TTL values above two hours. This of course does not include the 86400, or 1 day, TTL the prevalence of which is most likely due to the fact that it is the recommended TTL time(Lottor, 1987), and in many cases the default TTL assigned to records, not including MX . This becomes even more obvious when RFC 1912 states "Popular documentation like [RFC 1033] recommended a day for the minimum TTL, which is now considered too low except for zones with data that vary regularly." (Barr, 1996).

Of particular interest are is the 0 TTL presence that appears between December 2013 and September 2014, which will be discussed in 4.1.4.

4.1.3 Normalised TTL frequency for resource records

TTL frequency for nine separate resource records was also investigated. These nine were chosen as they consistently appear throughout the dataset, which then allows for a broader comparison of

TTL behavior. Table 4.2 gives a breakdown of the most popular TTL setting for various resource records observed across the dataset.

Table 4.2: Normalised TTL frequency by resource record by month

RR	A		PTR		CNAME		TXT		MX		AAAA		NS		SOA		SRV	
Month	TTL	% of RR	TTL	% of RR	TTL	% of RR	TTL	% of RR	TTL	% of RR	TTL	% of RR	TTL	% of RR	TTL	% of RR	TTL	% of RR
October 2013	300	23.742	86400	56.342	3600	19.370	900	73.042	3600	22.339	300	75.573	86400	36.792	86400	30.667	300	50.000
November 2013	300	29.323	86400	52.533	3600	22.076	900	59.260	3600	20.379	300	82.764	86400	32.223	86400	63.636	300	50.000
December 2013	300	29.830	86400	51.723	3600	22.949	900	76.091	3600	18.790	300	53.333	3600	43.548	86400	62.500	86400	40.000
January 2014	300	31.025	86400	47.865	3600	22.535	900	77.641	14400	36.071	300	79.297	3600	46.535	86400	70.000	300	50.000
February 2014	300	27.275	86400	50.968	3600	22.592	900	89.852	14400	59.295	300	82.425	86400	42.466	86400	41.667	86400	28.571
March 2014	300	27.488	86400	49.029	3600	22.655	900	90.446	14400	58.009	300	78.786	86400	36.047	86400	45.455	300	60.000
June 2014	300	17.670	86400	47.852	3600	21.967	900	77.105	14400	43.436	300	70.424	86400	38.806	86400	56.25	300	57.142
July 2014	300	22.688	86400	49.161	3600	21.710	900	68.615	3600	24.783	300	70.634	3600	39.726	86400	39.130	86400	37.500
August 2014	300	31.432	86400	51.554	3600	21.804	900	72.255	14400	34.819	300	84.189	86400	41.176	86400	46.154	300	50.000
September 2014	300	31.178	86400	51.160	3600	21.612	900	64.361	3600	23.292	300	83.344	3600	36.066	86400	27.027	300	57.143
October 2014	300	24.481	86400	50.594	3600	21.649	900	64.341	3600	28.144	300	88.845	86400	38.356	86400	43.077	300	57.143
November 2014	300	32.981	86400	50.211	3600	20.466	900	67.852	3600	25.831	300	89.350	3600	49.206	86400	44.681	300	33.333
December 2014	300	26.876	86400	49.807	3600	21.787	900	72.055	3600	25.062	300	77.253	3600	44.595	7200	40.741	86400	40.000
January 2015	300	28.564	86400	51.148	3600	21.082	900	78.415	14400	33.272	300	87.008	3600	45.205	7200	26.829	300	38.462
February 2015	300	34.066	86400	48.536	3600	21.657	900	71.124	14400	27.994	300	89.827	3600	49.383	86400	32.653	3600	28.571
March 2015	300	34.394	86400	48.819	3600	21.008	900	68.493	14400	27.653	300	89.377	3600	40.000	7200	32.500	3600	30.000
April 2015	300	29.541	86400	48.138	3600	20.462	900	65.729	3600	24.914	300	79.940	3600	35.556	86400	39.130	300	37.5
May 2015	300	33.035	86400	46.974	3600	22.456	900	61.371	86400	23.752	300	47.224	3600	38.462	86400	45.833	3600	46.154
June 2015	300	30.338	86400	44.936	3600	22.306	900	69.287	3600	23.548	300	40.556	3600	31.429	86400	44.000	3600	50.000
July 2015	300	24.287	86400	46.859	3600	22.990	900	72.936	3600	23.713	300	34.641	3600	34.545	86400	30.769	3600	41.667
August 2015	300	16.379	86400	46.344	3600	23.042	900	74.033	3600	22.530	300	41.168	86400	40.000	86400	42.857	3600	36.364

There seems to be a clear tend towards lower TTL configurations for many of the resource records present in table 4.2. The 5 minute TTL is the most popular A record TTL across all months, which suggests that domain administrators value the ability of promptly redirecting domain traffic to different servers more than the increased bandwidth cost created by smaller TTL values. The lower 300 TTL presence for A record traffic in June of 2014 comes about as a result of a larger presence of 0 TTL and 28800 TTL A queries than in other months. The increased 28800 TTL presence in this month comes about as a result of responses for subdomains of rhs.mailpolice.com, which is a Right Hand Side Black List (RHSBL). RHSBLs contain domain names belonging to TLDs, and derive their name by storing and filtering emails based on the domains given at the right hand side of the @ in the address (Miszalska *et al.*, 2007). While these are similar to DNSBL services, they filter using the TLD and not the IP address of the sender (Jung and Sit, 2004). This indicates that the IP block was the target of email spamming during this month. While August 2015 shows a similar reduction in the prevalence of 300 TTL values, this is because of a decrease in ratio between this and other common A record TTL values, and is not as a result of anomalous traffic. This decrease can also be partially, but not wholly, attributed to an increase in 900 TTL A record traffic as a result of a larger spamhaus.org, a DNSBL, presence in the dataset.

The one day TTL is the most popular PTR TTL choice, and its prevalence in the dataset remains fairly consistent throughout the captured months. The 3600 and 300 TTL values rank second and third respectively for PTR records for most of the months in the dataset, but their presence is much lower than the one day TTL. 43200 and 28800 TTLs also rank in the top 5 TTLs seen for most months. This seems to indicate that there is less need to have frequently updated PTR records as opposed to other record types. This difference in TTL selection could also be explained by the targets of different resource records. PTR records may not necessarily target the same domains as A records, which would create a discrepancy in TTL frequency.

The one hour TTL remains the most frequent CNAME TTL throughout the dataset. This is in part due to CNAME requests for services hosted through the Akamai CDN domains, whose CNAME TTLs are set to 3600. While the Akamai edgesuite.net and *akadns.net* domains are the two largest contributors of 3600 CNAME TTLs, they are closely followed by subdomains for *apple.com*, *amazonaws.com*, *windowsupdate.com*, *skype.net* and *icloud.com*, indicating that domain administrators from various spheres and organizations are favoring the one hour TTL for this resource record.

An overwhelming number of TXT responses yielded a 15 minute, or 900 TTL. This presence is largely due to the configuration of TXT responses for *spamhaus.org* domains. 86400 was the second most used TTL configuration for TXT records, and would have ranked first if it was not for the large 900 presence created by the *spamhaus.org* domains. Interestingly, the 10 TTL had the third largest presence, which was generated by *sophosxl.net* subdomains. These are generated by anti-virus software from **SophosLabs**, which runs its SXL protocol using DNS queries as part of their threat detection infrastructure[1].

The distribution of 3600 and 14400 TTL values remains similar throughout most of the months, indicating that they are both equally popular TTL choices for MX domains. The 14400 TTL is prevalent as it is the default TTL value for MX records. The large 3600 presence is seen for a number of unique MX domains, and indicates a shift in industry towards lower MX TTL values, which corresponds to findings for other TTL values across resource records.

The 3600 and 86400 TTLs are the most popular choices for NS records. The 3600 TTL suggests that some domains require more flexible name server configurations, but it can also be as a result of administrators leaving flexibility in case a name server for their domain fails, and may not point to the name servers for those domains being less reliable than for other domains.

The 300 TTL is by far the most popular configuration for AAAA records, a similar trend to the A records captured in the dataset. An increase in 86400, 3600 and 1800 TTL values for AAAA records led to the decrease in 300 TTL presence in May 2015. This change in TTL configuration remains the same in the months following May, barring August 2015 where the 1800 TTL was favored over the 3600 TTL, but otherwise similar to previous months. The noticeable drop in 300 TTL frequency for December 2013 comes about as a result of a large 200 TTL presence for subdomain responses to the exodus.desync.com domain; examples being *EXOduS.DEsYNc.Com, exOduS.DEsYnC.coM, EXODUs.dEsYNc.COm* among others. This phenomenon could be due to 0x20 bit manipulation (Dagon *et al.*, 2008), and is discussed further in section 5.3.2.

SOA records show a strong preference for the one day TTL. RFC 1035 stated that SOA TTL

[1] https://www.sophos.com/en-us/support/knowledgebase/117936.aspx

values should be set to 0 to prevent caching (Mockapetris, 1987b), but this was amended in RFC 2181, which notes the comment and refutes it, suggesting that SOA records can utilize other TTL values (Elz and Bush, 1997).

Interestingly, the 300 and 3600 TTL presence for SRV records is frequent throughout the dataset. RFC 2782 states that SRV weight should only be used statically, and dynamic server selection would require lower TTL values that would clutter network caches and increase bandwidth use (Gulbrandsen *et al.*, 2000). It would seem that the increase in network speed, bandwidth and cache memory have allowed administrators to take greater advantage of SRV records with respect to the estimation and selection of services related to their domain.

RFC 1033 recommends TTL values of between one day (86400 seconds) and 1 week (604800 seconds), and suggests only lowering the TTL values if changes are expected (Lottor, 1987). This RFC was written in 1987, and its recommendations are far removed from the current TTL distribution observed in the dataset, in which the majority of the top ranking TTL values fall under one day.

4.1.4 0 TTL presence analysis

Below is an explanation to the large 0 TTL presence seen between December 2013 and September 2014. The presence of disposable domains will be discussed and a review non-disposable 0 TTL domain activity given.

4.1.4.1 Disposable Domains:

Approximately 99% of the unique 0 TTL packets in each dataset were hosted by *mailshell.net*, a domain owned by `mailshell`[2], an Internet security firm that offers email-, web- and dns-filtering as well as anti-phishing solutions. This is as a result of their employment of disposable domains in their service (Chen *et al.*, 2014). The 0 TTL is set, in this instance, to ensure that DNS cache servers are not overloaded by creating cached records for multiple thousands of one-use domains, which would severely affect the performance and memory of the DNS caching sever in question. While both *spamhaus.org* and *mailshell.net* are mentioned in Chen *et al.* (2014), **spamhaus** domains did not result in a noticeable influx of 0 TTL packets .

4.1.4.2 Non-disposable 0 TTL presence

Not included in table 4.3 are in-addr.arpa responses for TeamViewer servers. TeamViewer is a remote access and online collaboration service[3], and will as a result generate one-use records so

[2]http://www.mailshell.com/ns/
[3]https://www.teamviewer.com/en/index.aspx

that end hosts can connect to the created server on the end-host that is hosting the session. These PTR responses also had a 0 TTL set, but these responses are similar to disposable domains in the sense that they are one-time use and so have not been included in the following analysis. Table 4.3 concerns itself with only the months that had the 0 TTL present in the top ten TTL ranking.

Table 4.3: Top 10 normalized frequent 0 TTL domains

Rank	Jan 14	Feb 14	Mar 14	Jun 14	Jul 14	Aug 14	Sep 14
1	outlook.com	outlook.com	outlook.com	outlook.com	outlook.com	outlook.com	outlook.com
2	espier.mobi	hichina.com	domobile.com	spotify.com	dstv.com	site2unblock.com	dstv.com
3	dstv.com	dstv.com	sharesdk.cn	dstv.com	spotify.com	ctnsnet.com	spotify.com
4	nbpush.com	live.com	dstv.com	supersport.com	supersport.com	dpliveupdate.com	supersport.com
5	live.com	lyrics007.com	hichina.com	live.net	oldmutual.co.za	dstv.com	playtime.bg
6	sinkdns.org	topnewinfo.cn	dressthat.com	tedro2.fr	live.net	hidebux.com	tedro5.fr
7	supersport.com	supersport.com	sales200.com	oldmutual.co.za	amnetsal.com	ns37.net	oldmutual.co.za
8	live.net	live.net	live.com	live.com	vitalteknoloji.com	supersport.com	greentreeapps.ro
9	greentreeapps.ro	greentreeapps.ro	joyogame.com	wwiionline.com	greentreeapps.ro	narutoget.com	dsintic.net
10	domobile.com	export-supply.com	goodphone.mobi	vitalteknoloji.com	veeam.com	ani1.net	vitalteknoloji.com

The *outlook.com* domain is almost always the leading contributor of 0 TTL responses, not including disposable domains. This is as a result of Microsoft configuring their *outlook.com* replies to have a TTL of 0, most likely to prevent an overconsumption of DNS memory on caching resolvers. The domains *live.com* and *live.net* also fall under the Outlook DNS infrastructure. Almost all of these queries are A queries. The three most interesting results here are dstv.com, supersport.com and oldmutual.co.za, not only because of the South African context, but also because all three of them (Old Mutual to a lesser extent) are among the top 10 contributors to the 0 TTL response traffic. A breakdown of these three domains will be given below.

dstv.com : DSTV subdomain responses have TTLs of either 600 or 0. All of the 0 TTL responses are for A queries, and have 18 individual subdomains responding with a 0 TTL. Responses for the *dstv.com* domain also return 0 TTLs. There is evidence that the resolved IP changes between queries for this domain, which suggests either active server load-balancing or the mimicking of CDN-like behavior from the domain.

supersport.com: The supersport responses are also all A queries. While there is a positive TTL presence for supersport CNAME queries, all A queries return a 0 TTL for seven subdomains seen in all ten months and two subdomains seen in May and June. Responses for *supersport.com* and *supersport.mobi* also return 0 value DNS TLLs, which suggests that the domain administrators set the TTLs to prevent record caching.

oldmutual.co.za: Old Mutual returns 0 TTLs for six subdomains present in each month, including responses for A queries for *www.oldmutual.co.za*. As with the previous two, all of the 0 TTL queries are A queries.

It was suggested in Larson and Barber (2006) that a 0 TTL presence indicates that the owner of the domain is planning to change the way their domains are configured, and wants to ensure that the expiring configuration is not cached. This does not seem to be the case with the three domains in question, as they sustain their TTL values throughout the ten month period. One reason that this TTL value is set to 0 would be that it gives the managing entity of the authoritative server the ability to instantaneously reroute traffic to different servers for each query. While this has the benefit of allowing for maximum data distribution management with respect to servers, it creates a much larger consumption of network bandwidth at the authoritative server, as it is queried every time a query is processed for the relevant domain. Setting a TTL of 0 is detrimental to both bandwidth consumption and load experienced by the authoritative server of the domain (Larson and Barber, 2006), as the domain query is forwarded to the authoritative server every time the query is made by an end host, instead of being served by a local cache server.

4.2 Analysis on authoritative servers for .za domains

Authoritative servers for domain records are not necessarily topologically or geographically near the servers that query for those domains, nor are they necessarily in close proximity to the content servers for which they hold the domain record. This could occur for a number of reasons, including but not limited to: off-shore hosting being cheaper than local alternatives; international web-hosts offering a more secure or complete service than local counterparts; or multinational corporations that manage their DNS infrastructure from the original country. This chapter looks at the geographical distribution of authoritative servers responsible for .za domain replies, in order to determine the geographical authoritative server presence for local (in this case .za) domains. Further analysis is done on the DNS-based latency experienced when querying for these domains. This is intended to give the reader an idea of the effect that server location has on end-user latency experience, as well as allowing for a comparison of experienced latency times for international servers.

4.2.1 Geolocation of .za authoritative servers

Table 4.4 shows the top ten countries that have a .za domain authoritative server presence. These countries are ranked by the percentage of unique IP addresses that respond with replies to .za queries. The country names have been abbreviated using ISO 3166-1 alpha-2 codes [4].

[4] http://data.okfn.org/data/core/country-list

Table 4.4: Distribution of unique IP responses for .za domains

Rank	1		2		3		4		5	
Month	Country	%ofTotal	Country	%ofTotal	Country	%ofTotal	Country	%ofTotal	Country	%ofTotal
October 2013	US	44.012	ZA	36.267	UK	7.028	DE	5.050	NL	1.296
November 2013	US	44.080	ZA	36.605	UK	6.532	DE	5.344	NL	1.432
December 2013	ZA	43.003	US	37.947	UK	6.741	DE	5.618	CA	1.379
January 2014	US	43.627	ZA	36.542	UK	6.623	DE	5.814	NL	1.386
February 2014	US	43.526	ZA	36.573	UK	6.705	DE	5.605	NL	1.419
March 2014	US	42.479	ZA	37.191	UK	6.422	DE	5.563	NL	1.545
June 2014	US	42.119	ZA	38.763	UK	6.109	DE	5.505	NL	1.395
July 2014	US	43.053	ZA	37.601	UK	5.804	DE	5.276	CA	1.477
August 2014	US	40.726	ZA	39.516	UK	6.268	DE	5.389	CA	1.430
September 2014	US	44.322	ZA	36.349	UK	6.281	DE	5.174	NL	1.432
October 2014	US	44.541	ZA	36.166	UK	6.576	DE	4.932	NL	1.582
November 2014	US	45.008	ZA	36.116	UK	5.650	DE	4.623	NL	1.477
December 2014	US	41.434	ZA	40.361	UK	5.496	DE	4.895	NL	1.417
January 2015	US	44.425	ZA	37.456	UK	5.575	DE	5.192	NL	1.498
February 2015	US	45.586	ZA	34.836	UK	6.337	DE	4.666	NL	1.608
March 2015	US	45.372	SA	35.579	UK	6.203	DE	4.691	NL	1.480
April 2015	US	44.813	ZA	37.871	UK	5.206	DE	4.126	CA	1.350
May 2015	US	42.717	ZA	38.261	UK	5.652	DE	5.000	NL	1.449
June 2015	US	43.711	ZA	37.256	UK	5.865	DE	5.275	NL	1.475
July 2015	US	43.586	ZA	36.917	UK	5.649	DE	4.555	NL	1.931
August 2015	US	42.964	ZA	38.808	UK	5.331	DE	4.665	NL	1.411

Rank	6		7		8		9		10	
Month	Country	%ofTotal	Country	%ofTotal	Country	%ofTotal	Country	%ofTotal	Country	%ofTotal
October 2013	CA	1.296	FR	0.819	AU	0.478	MU	0.307	SG	0.273
November 2013	CA	1.048	FR	0.838	AU	0.454	SG	0.349	MU	0.314
December 2013	NL	0.919	FR	0.817	AU	0.511	MU	0.460	PL	0.255
January 2014	CA	1.155	FR	0.770	AU	0.616	MU	0.308	SG	0.270
February 2014	CA	1.100	FR	0.816	AU	0.745	SG	0.319	MU	0.319
March 2014	CA	1.408	FR	1.133	AU	0.618	IE	0.378	SG	0.309
June 2014	CA	1.244	FR	1.207	AU	0.641	MU	0.377	IE	0.264
July 2014	FR	1.196	NL	1.161	AU	0.739	SG	0.352	MU	0.352
August 2014	NL	1.393	FR	1.173	AU	0.550	SG	0.513	MU	0.330
September 2014	CA	1.334	FR	0.911	AU	0.683	SG	0.358	MU	0.325
October 2014	CA	1.179	FR	1.055	AU	0.713	SG	0.310	IE	0.310
November 2014	CA	1.252	FR	1.220	AU	0.610	CH	0.353	IE	0.353
December 2014	CA	1.374	FR	1.073	AU	0.472	SG	0.429	CH	0.386
January 2015	CA	1.289	FR	1.080	AU	0.592	MU	0.314	CH	0.244
February 2015	CA	1.513	FR	1.230	AU	0.820	SG	0.410	MU	0.284
March 2015	CA	1.322	FR	1.165	AU	0.756	SG	0.378	CH	0.252
April 2015	NL	1.311	FR	1.080	AU	0.810	MU	0.347	IE	0.347
May 2015	FR	1.268	CA	1.268	AU	0.580	SG	0.362	MU	0.326
June 2015	FR	1.475	CA	1.180	AU	0.627	SG	0.295	MU	0.295
July 2015	FR	1.567	CA	1.239	AU	0.656	SG	0.364	MU	0.328
August 2015	FR	1.294	CA	1.254	AU	0.784	MU	0.314	CH	0.274

The United States of America (US) is almost always the largest responder with respect to .za domains. South Africa (ZA), for which the country code top-level domain (ccTLD) is reserved, holds the second largest presence of unique IP responders. These two countries represent around 80% of the total replies seen for .za domains across the dataset. The United Kingdom (UK) and

Germany (DE) rank third and fourth respectively, showing a smaller authoritative server presence than the US or ZA presence, but consistently 3.5-4% higher than the responder presence seen from other countries. Canada (CA), France (FR) and the Netherlands (NL) are each responsible for between 1-2% of the responding servers, while countries that appear towards the end of the ranking include Australia (AU), Singapore (SG), Mauritius (MU) and Switzerland (CH).

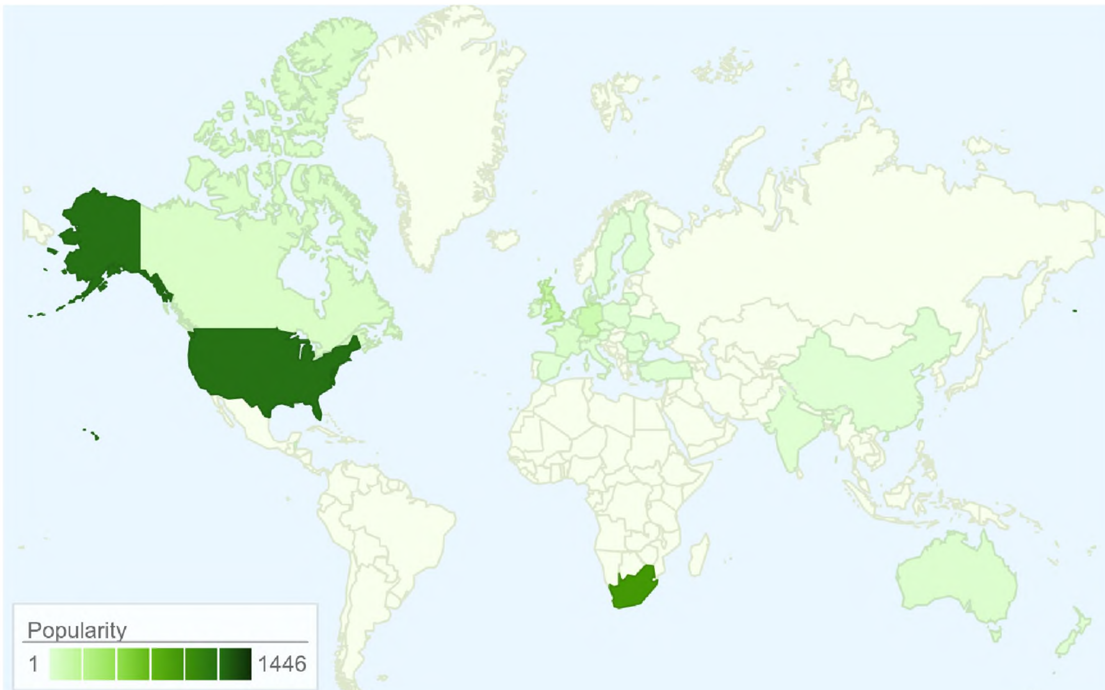


Figure 4.3: Geographic heatmap of authoritative server presence for February 2015

Figure 4.3 shows a server distribution heatmap for February 2015. The United States of America and South Africa are the most densely populated. A lot of the servers are spread across Europe, with the West being favored slightly over the East. The Far East and Australia also have a notable presence, but there is no contact from most of Africa, South America and the Middle East (not including Turkey).

4.2.2 Topology of .za domains

Table 4.5 describes the IP address and domain distribution with respect to /16 IP blocks observed in the dataset. The 196.0/16 and 197.0/16 IP blocks are local IP blocks, while the other /16 IP blocks represent international authoritative servers.

Table 4.5: Top 5 /16 network topology seen in .za dataset

Rank		1			2			3			4			5	
Month	/16	% of unique IPs	# of domains	/16	% of unique IPs	# of domains	/16	% of unique IPs	# of domains	/16	% of unique IPs	# of domains	/16	% of unique IPs	# of domains
October 2013	205.251	2.661	82	164.151	2.013	111	174.120	1.774	103	196.41	1.672	992	196.25	1.535	594
November 2013	192.185	3.912	141	205.251	3.353	90	164.151	1.886	95	208.76	1.851	60	196.41	1.711	965
December 2013	192.185	3.779	86	205.251	2.298	40	197.242	2.145	629	196.25	2.043	362	196.15	1.788	171
January 2014	192.185	5.468	173	205.251	3.966	108	164.151	1.848	94	197.221	1.810	1 262	208.76	1.617	60
February 2014	192.185	5.676	218	205.251	4.292	115	197.221	2.0575	1 545	164.151	1.987	103	197.242	1.667	1 434
March 2014	192.185	4.945	206	205.251	3.915	103	208.76	2.679	60	164.151	1.889	105	197.221	1.751	1 583
June 2014	205.251	4.713	88	192.185	4.374	163	208.76	2.413	47	164.151	2.112	102	197.242	1.735	3 384
July 2014	192.185	5.522	193	205.251	4.151	87	208.76	2.075	48	197.242	2.075	1 351	164.151	1.970	102
August 2014	192.185	4.985	174	205.251	3.959	71	164.151	2.383	94	208.76	2.089	44	197.221	2.016	1 520
September 2014	192.185	6.606	258	205.251	4.491	97	164.151	2.278	117	208.76	2.180	66	197.242	2.115	2 060
October 2014	192.185	7.041	326	205.251	4.311	112	197.242	2.295	1 596	208.76	1.985	56	164.151	1.892	115
November 2014	192.185	6.613	267	205.251	5.104	120	197.242	2.119	1 501	208.76	1.958	45	164.151	1.894	96
December 2014	205.251	4.637	82	192.185	4.594	120	197.242	2.362	870	208.76	1.803	28	164.151	1.760	61
January 2015	192.185	6.237	223	205.251	4.913	111	197.242	2.195	1 311	208.76	2.056	39	197.221	1.777	1 457
February 2015	192.185	7.030	285	205.251	5.643	137	197.242	1.955	1 494	208.76	1.892	41	164.151	1.702	77
March 2015	192.185	6.958	292	205.251	5.006	135	197.242	2.141	1 513	208.76	1.732	39	197.221	1.700	1 750
April 2015	205.251	5.708	114	192.185	5.438	176	197.242	2.352	1 131	173.245	1.928	96	197.221	1.735	1 180
May 2015	205.251	5.290	114	192.185	5.145	177	197.242	2.029	1 301	173.245	1.812	88	164.151	1.812	70
June 2015	205.251	6.192	133	192.185	5.040	180	173.245	2.088	106	197.242	1.836	1 305	197.221	1.584	1 471
July 2015	205.251	5.696	116	192.185	4.650	166	197.242	2.235	1 240	173.245	2.127	100	164.151	1.766	70
August 2015	205.251	5.802	109	192.185	4.861	150	197.242	2.313	1 152	173.245	2.195	98	164.151	1.803	68

International IP blocks show a greater unique authoritative server IP population with respect to the overall dataset, when compared to local IP blocks. This is expected as the overall dataset shows a greater international IP presence overall when compared to the local authoritative IP presence, as seen in table 4.4. What is interesting is that the local /16 IP blocks, while showing a lower concentration of unique authoritative IPs, resolve a larger number of unique domains when compared to international /16 IP blocks. This would suggest that, while there are less authoritative servers for .za domains in local IP blocks than in those abroad, the local authoritative servers are each responsible for a greater number of unique domains when compared to their international counterparts. This means that while there are many fewer individual authoritative servers resolving for .za domains in ZA IP space, the authoritative servers have a higher concentration of unique domains for which they are responsible, while international authoritative servers will usually not be responsible for more than a few domains.

4.2.3 Latency seen for .za domains

Table 4.6 gives a breakdown of average latencies observed when pinging authoritative servers for IP addresses in countries that showed the largest unique authoritative presence for .za domains. The latency average was calculated using five ping times for each IP address gathered using fping, mentioned in section 3.2.5.

Table 4.6: Average latency observed for top geographical responders (ms)

Month	US	ZA	UK	DE	CA	FR	NL	AU	MU	SG	CH	average
October 2013	276.286	34.618	207.554	211.347	302.967	210.893	199.195	463.123	48.587	424.825	469.554	183.012
November 2013	279.754	35.199	208.689	215.718	290.801	209.699	199.618	490.719	48.047	430.861	495.64	185.890
December 2013	272.860	35.129	208.314	209.746	282.925	212.030	198.139	472.424	48.102	424.287	205.27	170.524
January 2014	277.805	34.482	207.785	211.277	306.240	208.524	200.576	484.551	49.845	414.000	358.864	189.830
February 2014	278.857	37.287	208.382	211.259	301.336	211.363	200.669	455.243	51.378	432.027	337.486	190.986
March 2014	281.572	34.663	206.385	212.941	305.733	209.892	200.666	472.953	49.849	431.149	449.01	190.241
June 2014	275.636	35.303	204.663	213.830	307.309	210.019	199.286	472.308	48.748	404.518	460.546	183.843
July 2014	277.602	34.908	204.865	212.954	297.361	210.130	199.647	419.709	48.835	421.563	469.72	197.677
August 2014	275.084	35.398	204.850	212.215	295.412	210.669	199.638	426.577	48.248	405.615	392.098	181.817
September 2014	275.447	35.714	206.487	213.049	295.243	210.843	199.208	471.075	46.435	405.493	372.460	190.042
October 2014	276.721	35.995	207.785	213.422	300.604	214.637	199.433	407.581	52.717	421.670	233.153	190.171
November 2014	272.153	34.200	206.062	212.113	292.630	215.673	201.617	439.158	48.026	416.347	276.569	188.244
December 2014	266.606	35.351	206.050	210.893	295.532	212.514	198.222	474.102	50.521	425.329	268.484	175.467
January 2015	272.137	34.988	205.739	212.516	297.721	214.112	205.208	415.962	50.592	392.406	281.334	184.883
February 2015	271.639	35.014	206.615	211.855	293.922	210.437	207.940	406.656	48.813	430.411	238.043	190.670
March 2015	272.920	34.823	206.891	211.327	294.616	215.111	202.221	428.444	53.523	403.356	263.058	189.835
April 2015	265.890	36.384	208.618	210.568	297.100	217.660	200.310	417.436	51.074	412.997	231.114	182.553
May 2015	263.138	35.241	204.626	212.392	299.306	211.390	207.056	419.618	48.033	423.698	269.632	178.598
June 2015	262.072	34.721	207.098	212.169	297.443	210.875	199.401	395.310	49.879	423.375	233.248	179.586
July 2015	263.363	36.387	208.168	211.364	297.779	211.598	205.483	399.366	52.033	411.820	216.381	180.947
August 2015	261.170	35.878	207.233	215.389	299.126	217.183	201.157	418.732	52.653	405.993	266.483	176.032

Despite the US having the largest authoritative server presence for .za domains, the average latency observed for these servers is higher than half of the top ten countries, being lower than only Canada, Australia, Singapore and Switzerland on occasion. The average US ping is roughly nine times that of its ZA counterparts, indicating that there is a significant DNS latency introduced by over half of the authoritative servers queried, when other locations are taken into consideration as well. Surprisingly, although Mauritius shows less than 1% of the overall authoritative server presence, it shows much lower latencies than more favored authoritative servers. Mauritius is also the only non-local area to offer latency rates below the observed average for the eleven most prevalent geographic authoritative server clusters.

One aspect of latency generation noted during research is that there is much greater fluctuation in latency generation in larger countries than smaller ones. This is believed to be the result of the distribution of authoritative servers across the landmass. For example a server in California would have a different distance and routing path from the pinging host than a server in New York. This fluctuation is more noticeable in the Australia dataset than the US and CA counterparts despite all three of them representing large land masses. This is as a result of there being less authoritative servers present from the AU region than the others, which leads to the data being less able to create a stable average than counterparts with a higher presence. The Switzerland latencies show a marked decrease from the beginning to the end of the dataset. There are two possible reasons for this. The first is that the number of unique IP addresses from the CH region is small, resulting in non-normalized fluctuation of latencies. The second is that pings to those IP addresses are instead rerouted to different servers for a response, or the reported geographic area is incorrect with respect to the IP address. The latter is observed in other cases below.

Three IP addresses in the US dataset, 196.220.43.240, 196.220.42.13 and 196.220.42.14 respond with pings as low as 14ms throughout the dataset. While the IP addresses are registered for use in the United States, and placed in Atlanta according to RobTex[5], they consistently return pings lower than the local average, which is physically impossible. Electrons simply cannot travel between the US and ZA in 14ms. This means that the pings are either being rerouted to a local server, or the end-host location of the IP address is not in the country for which it is registered. The IP 176.124.112.100 showed a similar ping time for the UK dataset. All four of these IP addresses are linked to name servers for .za domains. These observed latencies are expected to be the result of the use of Anycast in the DNS implementation of these IP addresses. End-hosts that query these domains will be routed to the closest replica of the Anycast group (Sarat *et al.*, 2006). Anycast is used not only to increase the availability and reliability of DNS records, but also to reduce experienced latency times (Sarat *et al.*, 2006).

It is clear, when considering the results in subsection 4.2.1, that many queries to authoritative servers for .za domains experience much greater DNS-based latency than .za domains with local authoritative servers. The question, then, is why and to what extent the experienced latency is important.

In a 2012 paper discussing the reduction of network latency via redundancy, it was noted that increased latency times would decrease user visits to and interaction with the web medium which was affected by the latency (Vulimiri *et al.*, 2012). The paper went on to cite two separate latency studies performed by Google (Brutlag, 2009) and Bing (Souders, 2009). Bing noted that a 500ms delay resulted in a 1.2% revenue drop, while a 2s delay resulted in a loss of 4.3% (Vulimiri *et al.*, 2012). Google stated that a latency increase of between 100 and 400ms resulted in a reduction in user searches between 0.2% and 0.6%, with search frequency decreasing further as the time users were exposed to latency increased. They also found that search frequency would take time to recover even after the latency was removed (Brutlag, 2009). Amazon also stated that every 100ms latency penalty implies a 1% decrease in overall sales (Singla *et al.*, 2014).

Older studies on user perception of injected latency found that users, when asked to rate web-pages, would rate the pages lower as loading time increased. Interestingly, when asked to rate how interesting the web-page was, users would identify the faster loading web-pages as more interesting than the latency injected ones (Ramsay *et al.*, 1998). This has large implications for international web-hosting, as increased local latency times could negatively affect the perception of the hosted web content, which is in some cases also the target userbase for the hosted content.

These papers suggest that even experienced latency at the level of a few hundred milliseconds could have a large impact on end-user experience, and that web-hosts using non-local authoritative

[5] https://www.robtex.org

servers are negatively impacting their website delivery and user experience, when compared to their local peers.

4.2.4 Domain breakdown

This section takes a deeper look at .za domain characteristics by splitting the datasets into those for various sub-TLDs for the .za ccTLD. An overview of authoritative server geolocation and latency experienced is given. The TTL values and RR frequency for the different TLDs will also be considered and discussed with reference to the previous section. All of the TTL and RR values are taken from normalized datasets, which were processed in the same way as the normalized TTL dataset to allow for accurate comparisons. The geolocation figures as well as the TTL and RR tables are excerpts from tables A.2 to A.11 found in the appendix, which contain a top five breakdown for the former, a top ten breakdown of the latter, as well as relevant percentages.

.co.za Figure 4.4 illustrates the distribution of servers for the top five geographic responders. Servers from the US make up the largest authoritative contribution, followed closely by servers situated in ZA. There is a constant presence from UK and DE servers, which each represent around 5% of the total server presence. The fifth ranked country is for the most part the Netherlands, but is displaced by Canada in April 2014 and France in June 2015. These ratios indicate that a large number of DNS responses to .co.za queries experience higher latencies than the dataset average, particularly the large US presence that generates more latency than any other ranked country barring Canada.

Table 4.7: Top 5 observed TTL and RRs for .co.za

Rank	1	2	3	4	5
Month	TTL	TTL	TTL	TTL	TTL
October 2013	7200	14400	86400	3600	600
November 2013	7200	14400	86400	3600	600
December 2013	7200	86400	3600	14400	600
January 2014	7200	3600	14400	86400	600
February 2014	7200	86400	14400	3600	600
March 2014	7200	86400	3600	14400	600
June 2014	86400	7200	3600	14400	600
July 2014	7200	86400	3600	14400	600
August 2014	7200	86400	3600	600	14400
September 2014	7200	86400	3600	14400	600
October 2014	7200	3600	86400	14400	600
November 2014	7200	3600	14400	600	86400
December 2014	7200	3600	600	14400	86400
January 2015	7200	3600	600	14400	86400
February 2015	7200	3600	14400	600	86400
March 2015	7200	14400	3600	600	86400
April 2015	7200	3600	600	14400	86400
May 2015	7200	3600	600	14400	86400
June 2015	7200	600	3600	14400	86400
July 2015	7200	600	3600	14400	86400
August 2015	7200	600	3600	14400	86400

Rank	1	2	3	4	5
Month	RR	RR	RR	RR	RR
October 2013	A	MX	CNAME	TXT	PTR
November 2013	A	MX	CNAME	TXT	AAAA
December 2013	A	MX	TXT	CNAME	SOA
January 2014	A	MX	CNAME	TXT	AAAA
February 2014	A	MX	CNAME	TXT	AAAA
March 2014	A	MX	CNAME	TXT	AAAA
June 2014	A	MX	CNAME	TXT	SOA
July 2014	A	MX	CNAME	TXT	AAAA
August 2014	A	MX	CNAME	TXT	AAAA
September 2014	A	MX	CNAME	TXT	AAAA
October 2014	A	CNAME	MX	TXT	AAAA
November 2014	A	MX	CNAME	TXT	AAAA
December 2014	A	MX	CNAME	TXT	AAAA
January 2015	A	MX	CNAME	TXT	AAAA
February 2015	A	MX	CNAME	TXT	AAAA
March 2015	A	MX	CNAME	TXT	AAAA
April 2015	A	MX	CNAME	TXT	AAAA
May 2015	A	MX	CNAME	TXT	AAAA
June 2015	A	MX	CNAME	TXT	NS
July 2015	A	MX	CNAME	TXT	PTR
August 2015	A	MX	CNAME	TXT	PTR

.org.za The .org.za domains show a similar server distribution to .co.za domains. Figure 4.5 shows some differences when compared to figure 4.4 . The two most notable differences are the larger ZA server presence as well as the more prominent CA presence across the board. The .org.za responses were most frequently seen from ZA servers, accounting for around 50% of unique server responses. US servers were ranked second overall for server contribution. Interestingly, Canada ranked consistently fifth throughout the dataset. Canada records some of the worst latency averages in the dataset, which will affect between 2% and 4% of replying .org.za authoritative server packets.

A summary of the .org.za TTL and RR frequency is given in table 4.8. The TTL frequency is similar to that observed by the .co.za subset, but the 7200 TTL is more favored, accounting for between 29% and 51% of the observed TTLs. A records make up a more substantial part of this subset when compared to the .co.za subset. Between 76% and86% of RRs were A records. This suggests that .org.za queries also experience significant DNS based latency, but less so than the previous subset, as there is a higher ZA presence of authoritative servers coupled with the increase in the 7200 TTL prevalence.

Table 4.9: .gov.za domains using non-local authoritative servers

Domain (.gov.za)	Server	Local IPs	International IPs	Domain hosted locally
aarto	NZ	1	1	Yes
breedevallei	CA	0	1	Yes
camdeboo	US	3	1	Yes
chrishanidm	UK	0	2	Yes
ecdoe	US	0	2	Yes
ecdoh	US	0	2	NXDOMAIN
ecdoeresearch	US	0	2	No - US
engcobolm	US	1	1	Yes
george	US	0	1	Yes
gis.bcmm	DE	0	1	Yes
hessequa	US	0	1	No - DE
johannesburg	DE	0	1	No -DE
kouga	US	1	1	Yes
kznunemployedgrads	US	0	2	Yes
lesedilm	US	0	1	Yes
rustenburg	US	0	1	No - US
stellenbosch	US	2	1	No - DE
srvm	US	2	1	Yes
bvm	UK	0	1	Yes

The .gov.za subset shows a large 600 TTL presence, accounting for between 26% and 34% of TTLs. This is also the only subset that has the 300 TTL ranked in the topfive5. This, coupled with the large A RR query presence (75%-86%) means that cumulative DNS based latency would be far greater than other subsets if a similar number of non-local authoritative servers were responsible for these domains. Fortunately, almost all of the responding authoritative servers are locally based, and while DNS based latency is still experienced, to individual end users it would represent a DNS latency cost of around 30ms, which is almost instantaneous when compared to other observed latency figures, such as the average latencies recorded in table 4.6. The fact that low TTLs are predominantly favored increases the risk of offshore authoritative servers. If the server crashes, locally cached records will time out relatively quickly, which will then prevent end-hosts from reaching the site, as the authoritative server is no longer responding.

Table 4.10: Top 5 observed TTL and RRs for .gov.za

Rank	1	2	3	4	5
Month	TTL	TTL	TTL	TTL	TTL
Oct13	600	3600	86400	7200	300
Nov13	600	86400	3600	7200	14400
Dec13	600	3600	86400	7200	300
Jan14	600	3600	86400	7200	10800
Feb14	600	3600	86400	10800	300
Mar14	600	3600	86400	300	7200
Jun14	600	3600	86400	7200	300
Jul14	600	3600	86400	300	10800
Aug14	600	3600	86400	7200	10800
Sep14	600	3600	86400	300	7200
Oct14	600	3600	86400	300	10800
Nov14	600	3600	86400	7200	300
Dec14	600	86400	3600	7200	300
Jan15	600	3600	86400	7200	300
Feb15	600	3600	86400	7200	300
Mar15	600	3600	86400	7200	300
Apr15	600	3600	86400	7200	43200
May15	600	3600	86400	10800	14400
Jun15	600	3600	86400	300	43200
Jul15	600	3600	86400	300	14400
Aug15	600	3600	86400	300	43200

Rank	1	2	3	4	5
Month	RR	RR	RR	RR	RR
Oct13	A	CNAME	MX	TXT	NS
Nov13	A	CNAME	MX	TXT	NS
Dec13	A	CNAME	TXT	MX	N/A
Jan14	A	CNAME	TXT	MX	N/A
Feb14	A	CNAME	TXT	MX	N/A
Mar14	A	CNAME	TXT	MX	N/A
Jun14	A	CNAME	TXT	MX	N/A
Jul14	A	MX	CNAME	TXT	N/A
Aug14	A	CNAME	MX	TXT	N/A
Sep14	A	CNAME	TXT	MX	N/A
Oct14	A	CNAME	TXT	MX	N/A
Nov14	A	CNAME	MX	TXT	N/A
Dec14	A	CNAME	MX	TXT	N/A
Jan15	A	CNAME	TXT	MX	N/A
Feb15	A	CNAME	TXT	MX	N/A
Mar15	A	CNAME	TXT	MX	N/A
Apr15	A	CNAME	TXT	MX	N/A
May15	A	CNAME	TXT	MX	N/A
Jun15	A	TXT	CNAME	MX	N/A
Jul15	A	MX	TXT	CNAME	N/A
Aug15	A	TXT	MX	CNAME	N/A

.ac.za The geographic distribution of authoritative servers for .ac.za domains is illustrated in figure 4.7. The .ac.za dataset shows a strong ZA presence, but is ranked third for both ZA frequency and US frequency when compared to other .za subsets. This subset also shows a favoring of DE authoritative servers over UK servers when compared to the overall results, as well as more notable AU presence. The large ZA presence is also expected here, as most academic institutions manage their domains internally. The .ac.za domains will overall generate less latency than their .co.za and .org.za counterparts, but more than .gov.za and other .za domains as a result of the larger US and AU presence.

The ac.za dataset shows an 86400 TTL presence much higher than the dataset average, with between 41% and 53% of all TTL values for this subset, and retains the first rank throughout all months. While the A record presence is lower on average than other .za subsets, it shows greater variation, with between 56% and 80% of RRs being A records. While this subset has a strong ZA authoritative server presence, it also has a non-negligible international server presence, including a higher Australian authoritative server percentage than other subsets, which results in high latency values. Nonetheless, the 86400 TTL presence as well as the 10800 TTL presence in the top five ranks, works to greatly mitigate the effects of DNS-based latency on queries, as they are more likely to have live entries in the local cache servers than other subsets, based on TTL alone.

which offers average latency only slightly above local latency averages.

Table 4.12: Top 5 observed TTL and RRs for other .za domains

Rank	1	2	3	4	5
Month	TTL	TTL	TTL	TTL	TTL
Oct13	3600	84600	86400	7200	600
Nov13	84600	86400	3600	7200	600
Dec13	84600	7200	86400	3600	600
Jan14	84600	86400	3600	600	7200
Feb14	84600	3600	86400	7200	600
Mar14	84600	3600	86400	7200	600
Jun14	600	84600	86400	3600	7200
Jul14	84600	3600	600	86400	7200
Aug14	84600	3600	86400	600	300
Sep14	84600	86400	600	3600	7200
Oct14	84600	86400	600	7200	3600
Nov14	84600	86400	3600	600	300
Dec14	84600	3600	600	86400	7200
Jan15	84600	600	3600	86400	7200
Feb15	84600	3600	86400	600	7200
Mar15	84600	3600	86400	600	7200
Apr15	86400	600	84600	3600	7200
May15	86400	84600	600	3600	7200
Jun15	86400	3600	84600	600	7200
Jul15	86400	84600	600	3600	7200
Aug15	86400	600	84600	3600	7200

Rank	1	2	3	4	5
Month	RR	RR	RR	RR	RR
Oct13	A	MX	CNAME	SOA	N/A
Nov13	A	MX	CNAME	SRV	SOA
Dec13	A	MX	CNAME	N/A	N/A
Jan14	A	MX	CNAME	TXT	N/A
Feb14	A	MX	CNAME	TXT	SOA
Mar14	A	MX	CNAME	TXT	SOA
Jun14	A	MX	SOA	CNAME	SRV
Jul14	A	MX	SOA	CNAME	AAAA
Aug14	A	MX	CNAME	SOA	N/A
Sep14	A	CNAME	MX	SOA	TXT
Oct14	A	CNAME	SOA	MX	N/A
Nov14	A	MX	CNAME	SOA	TXT
Dec14	A	MX	SOA	SRV	CNAME
Jan15	A	MX	SOA	CNAME	NS
Feb15	A	MX	SOA	CNAME	TXT
Mar15	A	MX	CNAME	SOA	TXT
Apr15	A	MX	CNAME	TXT	SOA
May15	A	MX	CNAME	TXT	N/A
Jun15	A	MX	TXT	CNAME	SOA
Jul15	A	MX	TXT	CNAME	N/A
Aug15	A	MX	CNAME	TXT	SOA

4.3 NX domain analysis

Authoritative NXDOMAIN status codes are returned with responses when a domain queried at an authoritative server does not exist. Some NXDOMAIN traffic can be indicative of malicious network behavior; an example of this being continued queries for domains that respond with correct NXDOMAIN responses that show positive TTLs (Oberheide *et al.*, 2007). However, more often than not, NXDOMAIN status code generation is the result of host misconfiguration (Kumar *et al.*, 1993) or as a result of Internet spam-filtering services utilizing the DNS protocol in their service (Jung and Sit, 2004), examples of which can be seen in table 4.15.

4.3.1 Observed NX TTLs and RRs

Tables 4.13 and 4.14 describe the top ranked TTL and RR values observed for normalised NXDOMAIN traffic across the dataset. The 86400 TTL is for the most part the top ranked TTL, with the exception of June and July 2014.

Table 4.13: Top 5 observed TTL and RRs for other .za domains

Rank	1		2		3		4		5	
Month	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs
October 2013	86400	31.240	3600	8.311	900	6.779	10800	2.068	7200	2.016
November 2013	86400	43.271	900	12.297	3600	11.122	7200	2.680	1800	2.276
December 2013	86400	30.868	3600	14.983	900	8.552	7200	3.903	600	2.976
January 2014	86400	39.490	3600	12.674	900	9.325	7200	3.129	1800	2.731
February 2014	86400	49.997	3600	10.428	900	8.319	7200	2.385	300	2.047
March 2014	86400	48.802	3600	10.819	900	10.416	7200	2.526	1799	2.419
June 2014	900	34.501	86400	27.912	3600	9.909	1799	3.044	7200	2.552
July 2014	900	35.680	86400	24.218	3600	7.385	7200	1.905	600	1.749
August 2014	86400	37.359	900	25.000	3600	6.195	1799	4.560	7200	1.694
September 2014	86400	35.655	900	20.132	3600	7.146	1799	2.902	7200	1.647
October 2014	86400	23.018	900	9.952	3600	3.798	1799	1.363	10800	2.116
November 2014	86400	12.031	900	3.012	3600	2.518	1799	1.011	10800	0.926
December 2014	86400	13.369	3600	5.004	900	4.928	600	0.910	10800	0.848
January 2015	86400	7.397	900	3.213	3600	2.778	1799	0.718	600	0.551
February 2015	86400	25.401	3600	1.897	900	1.857	1799	0.749	600	0.381
March 2015	86400	18.275	900	2.646	3600	2.497	1799	1.002	600	0.495
April 2015	86400	18.300	900	2.902	3600	2.293	1799	0.881	600	0.546
May 2015	86400	26.532	900	4.077	3600	2.950	1799	1.893	600	0.709
June 2015	86400	16.499	3600	5.318	900	3.102	1799	1.847	21599	0.771
July 2015	86400	20.874	3600	6.776	900	4.577	1799	2.238	600	0.770
August 2015	86400	21.630	900	1.904	3600	1.543	1799	1.290	600	0.393

Two interesting TTL values, the 1799 and 21599 TTLs, stand out as they are not standard TTL values, which are almost always multiples of 60. Surprisingly, both of these TTL values are response TTLs from 8.8.4.4 and 8.8.8.8, the two Google DNS servers. Clients in the observed IP block generating NXDOMAIN queries by sending misconfigured packets directly to the Google DNS servers instead of routing them through the local cache servers will often see these TTL values in response.

The increase in the 900 TTL frequency is directly related to the increase in MX RR frequency seen for the same months. These months saw the cache servers send MX queries with between four and six seemingly random letters, and is expected to be the result of a brute force mail server search from an affected spam bot, or a compromised host being used as an open mail relay (Schonewille and van Helmond, 2006).

Table 4.14: Top 5 observed TTL and RRs for other .za domains

Rank	1		2		3		4		5	
Month	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs
October 2013	A	72.580	TXT	12.737	PTR	5.874	MX	3.500	AAAA	2.108
November 2013	A	67.910	TXT	17.072	PTR	5.129	MX	3.613	AAAA	2.955
December 2013	A	42.509	TXT	25.945	PTR	15.480	MX	7.499	SOA	3.420
January 2014	A	56.586	TXT	21.369	PTR	7.447	MX	5.983	AAAA	3.167
February 2014	A	64.250	TXT	14.608	PTR	8.908	MX	5.397	AAAA	3.064
March 2014	A	66.166	TXT	14.573	MX	6.792	PTR	5.922	AAAA	2.897
June 2014	A	39.743	MX	31.369	TXT	15.923	PTR	4.243	AAAA	4.182
July 2014	A	34.893	MX	33.308	TXT	12.800	PTR	5.600	SRV	5.014
August 2014	A	46.427	MX	23.659	TXT	10.988	AAAA	7.385	PTR	4.328
September 2014	A	52.574	MX	17.590	TXT	10.515	AAAA	8.461	PTR	4.790
October 2014	A	53.176	AAAA	12.731	SOA	10.780	MX	8.961	PTR	8.808
November 2014	SOA	40.942	A	40.185	AAAA	7.169	PTR	5.366	TXT	2.952
December 2014	A	49.396	SOA	31.331	MX	5.319	AAAA	5.174	PTR	4.255
January 2015	SOA	53.659	A	29.980	AAAA	4.686	PTR	3.549	TXT	3.382
February 2015	SOA	41.084	A	32.423	AAAA	16.222	PTR	4.863	TXT	2.688
March 2015	SOA	33.997	A	33.229	AAAA	13.882	PTR	12.331	TXT	3.723
April 2015	A	58.532	AAAA	15.694	PTR	12.396	SOA	7.297	MX	2.704
May 2015	A	56.061	AAAA	19.855	PTR	7.920	SOA	6.710	TXT	5.326
June 2015	A	63.967	AAAA	16.170	PTR	6.397	SOA	5.624	TXT	3.799
July 2015	A	58.649	AAAA	14.324	SOA	9.936	TXT	7.458	PTR	4.91
August 2015	A	78.924	AAAA	11.475	SOA	4.204	PTR	1.680	TXT	1.674

These two months show the largest MX presence, and exhibit a much higher percentage of overall RR traffic than other research suggests should be seen (Zdrnja *et al.*, 2007). The large SOA presence in November 2014, as well as between January and March 2015, are as a result of misconfigurations seen for 196.x.x.162, and will be mentioned in subsection 4.3.3.1.

4.3.2 Top Cache NX traffic

Most of the observed NXDOMAIN responses for the caching servers were from DNSBLs (Jung and Sit, 2004). Many DNSBL services receive queries for IP addresses attached to a domain registered for that DNSBL, and will respond with an NXDOMAIN response if the identified IP address is not found on the list itself (Yadav and Reddy, 2012). Table 4.15 is a list of the DNSBL services for which NXDOMAIN responses appeared. These responses were filtered out as they are not true NX responses, but form part of the DNSBL framework (Yadav and Reddy, 2012). As such, they are not actually representative of queries for domains that do not exist.

Table 4.15: Domain extensions seen in DNSBL NXDOMAIN responses

Domain TLD
spamhaus.org
uribl.com
multi.surbl.org
spamcop.net
score.senderscore.com
sa-accredit.habeas.com
sa-trusted.bondedsender.org
dnsbl.sorbs
sibl.support-intelligence.net
list.dnswl.org
iadb.isipp.com
mailspike.net
mailpolice.com

4.3.2.1 Responses from 146.231.128.1

The authoritative server hosted at 146.231.128.1 is responsible for the largest number of NXDOMAIN responses within the caching dataset. Table 4.16 describes the observed packet traffic as well as some of the commonly seen domain queries that led to NXDOMAIN responses.

Table 4.16: Packet breakdown of 146.231.128.1 NXDOMAIN traffic

Month	# of packets	% of NX packets	# of domains	Top domain	# of responses
Oct13	33 946	0.007	16 439	::1.async.org.za	860
Nov13	30 248	0.007	16 488	::1.async.org.za	954
Dec13	21 715	0.005	8 736	::1.async.org.za	755
Jan14	24 282	0.005	10 251	::1.async.org.za	943
Feb14	34 240	0.007	18 126	kwc-ntfs007.kc.ecape.school.za	1 082
Mar14	36 687	0.006	19 346	ns3.24ohone2014.co.za.async.org.za	961
Jun14	19 092	0.006	7 897	::1.async.org.za	816
Jul14	21 586	0.005	7 834	::1.async.org.za	1 475
Aug14	24 343	0.006	10 059	::1.async.org.za	1 368
Sep14	20 360	0.004	7 166	::1.async.org.za	1 049
Oct14	10 670	0.002	5 080	hn.kd.ny.adsl.async.org.za	290
Nov14	9 806	0.002	4 896	d011.albene.info.async.org.za	633
Dec14	8 895	0.001	5 613	hn.kd.ny.adsl.async.org.za	151
Jan15	14 146	0.002	8 209	tiffin2.voipphoneonline.com.async.org.za	219
Feb15	15 385	0.002	7 094	superpositions.enjoydaring.com.async.org.za	424
Mar15	12 432	0.001	6 071	gc.gc.ecape.school.za	144
Apr15	10 230	0.001	4 058	mta1.imx14.info.async.org.za	470
May15	17 559	0.002	4 183	mta1.imx14.info.async.org.za	1 958
Jun15	15 110	0.001	5 036	dealzzy.net.async.org.za	1 366
Jul15	10 652	0.001	5 552	ne73-nat.renet.ru.async.org.za	222
Aug15	11 397	0.001	5 870	193.189.116.67.host.e-ring.pl.async.org.za	410

Queries for the *::1.async.org.za* domain are believed to be an IPv6 configuration error which creates malformed IPv6 packets. Other domains of this type that appear throughout the dataset include *::1.org.za, fe80::1.async.org.za, ::1* and *fe80::181f:20e5:e533:d17.org.za* amongst others. Table 4.16 also highlights instances of misconfiguration that leads to TLD domains being appended to the queried domain, for example *dealzzy.net.async.org.za.* The most common appended TLDs are *async.org.za, org.za, school.za* and *ecape.school.za.* These two misconfigurations generated most of the observed caching NXDOMAIN responses throughout the entire dataset. Section 4.3.2.3 will look at some of the unexpected NXDOMAIN traffic that remains after these misconfigurations have been filtered out.

4.3.2.2 Response cluster

There was a cluster of NXDOMAIN responses captured in February 2015, as seen in table 4.17. This is the only such cluster of responses captured throughout the caching dataset. All of the captured queries are for domains within the *.local* or *.vp.local* domain space. This is the result of a server misconfiguration that led to local network addresses being queried through global DNS.

Table 4.17: Packet clustering seen in February 2015

Source IP	# of packets	# of domains	Top domain	# of responses
192.58.128.30	11 892	4 548	local	216
192.5.5.241	11 874	4 544	local	193
202.12.27.33	11 816	4 509	local	178
199.7.83.42	11 804	4 459	local	176
198.41.0.4	11 804	4 490	local	165
193.0.14.129	11 784	4 507	local	197
192.203.230.10	11 769	4 421	local	183
192.33.4.12	11 729	4 469	local	200
192.228.79.201	11 725	4 465	local	177
192.36.148.17	11 724	4 481	local	199
192.112.36.4	11 689	4 454	local	209
128.63.2.53	11 652	4 497	local	171

The IP addresses seen in the Source IP column all correspond to root DNS servers, which responded to these packets as the local TLD is not a registered or recognized TLD in DNS infrastructure. It has been suggested that the TLD be prohibited in an Internet-Draft (Chapin and McFadden, 2011) that expands on RFC 2606 (Eastlake and Panitz, 1999), the RFC that specifies reserved DNS TLDs. It was suggested that this be done as a result of the use of the .local TLD on private networks, where the presence of a global DNS .local TLD may cause differences in resolution behavior for the TLD on different local networks, which poses a security threat (Chapin and McFadden, 2011).

4.3.2.3 Unexpected NXDOMAIN traffic seen at caching servers

The .su ccTLD is the ccTLD used by the old Soviet Union, which has persisted in the ccTLD registry after the dissolution of the state itself (Von Arx and Hagen, 2002). Continued legitimate use of the ccTLD is non-existent, but its use has been noted even in current years with respect to malicious traffic (Ling *et al.*, 2014). **Snort**[6], an intrusion detection system, flags .su domains as possible malware activity (Hermanowski, 2015). A 2012 study that tracked DDoS activity listed the .su TLD as the sixth most frequent TLD observed for victim URLs present in that dataset (Büscher and Holz, 2012). Table 4.18 describes the captured .su ccTLD activity across the caching dataset.

Table 4.18: Packet breakdown of .su ccTLD cache traffic

Month	# of packets	# of domains	Top domain	# of responses	# of source IPs
Oct13	116	40	ns.neic.nsk.su	39	25
Nov13	81	33	ns.neic.nsk.su	29	26
Dec13	72	18	ns2.transfer.su	42	18
Jan14	32	12	finley.su	11	16
Feb14	90	21	ns2.transfer.su	44	24
Mar14	34	20	www.su	10	18
Jun14	27	17	www.su	5	20
Jul14	29	23	ns.neic.nsk.su	4	18
Aug14	20	15	redsun.lvk.cs.msu.su	2	14
Sep14	34	21	host-176-107-248-11.it-net.su	5	16
Oct14	53	24	sunnyweek.su	7	20
Nov14	92	29	nitmurmansk.su	30	27
Dec14	25	22	46-161-129-86-nts.su	3	16
Jan15	236	20	bnswhat.su	105	15
Feb15	644	13	bnswhat.su	382	9
Mar15	169	59	bnswhat.su	61	18
Apr15	130	15	host-94.198.132.vernet.su	103	16
May15	145	68	luposer.su	19	15
Jun15	30	19	ns2.airlink.su	6	16
Jul15	12	8	tracker.irc.su	2	8
Aug15	20	15	host-77.91.195.112.vernet.su	3	15

The amount of captured .su traffic for the caching dataset is small. Some of the captured domains are most likely the result of DNS misconfiguration, for example the *www.su* domains captured in March and June 2014, as they are lacking an actual domain to resolve. Two months that exhibit strange domain activity are March and May 2015, where a number of pseudo-random, single-query domains with the .su ccTLD were captured. Examples of the captured domains are *jjheuuxwbvidq.su* and *flrsdtduo.su*, among others. This traffic behavior seems to point to the

[6] https://www.snort.org/

presence of a fast-flux botnet (Yadav *et al.*, 2012), using the .su ccTLD to generate domains on which its C&C is hosted (Stalmans and Irwin, 2011).

4.3.3 Other NX traffic

This section looks at anomalous NXDOMAIN traffic captured between hosts on the local IP block and non-local DNS servers. Section 4.3.3.1 describes the packet behavior of a proxy server of a local network, while section 4.3.3.2 discusses captured .su ccTLD domains seen in the same dataset.

4.3.3.1 196.x.x.162

The system with IP 196.x.x.162 acts as a proxy address, and is a network address translation (NAT) server for a local network within the monitored IP block. As is seen in table 4.19, the misconfiguration presence began in July 2014, and has persisted throughout the rest of the dataset.

The large SOA presence seen for the months of November 2014, and January through March 2015, seen in table 4.13 are generated by this IP address. The SOA queries are the result of misconfiguration at the end-host, where unqualified domains are queried through the global DNS instead of of on the local network. These queries will query localhost SSID names as domain names, for example Emmas-iPad, android-2d73b2b20838c999 and John-PC. XxxXxxxxx-HP (X/x has been used in place of the name and surname that identified the PC) is another such example, and was the most queried domain for the months of July and August 2014. The query misconfiguration seems to be related to the presence of end-hosts on the wireless network, and their subsequent interaction with the NAT server.

Each month also shows a large Web Proxy Auto Discovery (WPAD) query domain presence. Browsers utilize Proxy Auto Configuration (PAC) to eliminate the need for manual proxy configuration, and WPAD, which is automatically enabled in most browsers, requests the URL of the PAC script from the DHCP and DNS servers (Smith, 2010). Attackers have been known to create malicious PAC servers in an attempt to compromise target hosts through WPAD, where the PAC server will deliver malicious code to the end-hosts (Pashalidis, 2003). Attackers are able to identify WPAD addresses, and subsequently respond with URLs that lead to malicious PAC scripts, by sniffing network queries (Smith, 2010). At least in the case of 196.x.x.162 traffic, this could lead to multiple WPAD addresses being compromised. One known attack of this kind had an infected computer masquerade as a WPAD proxy, which then identified a compromised server as a Microsoft Update server, which led to uninfected hosts downloading malware that they believed to be Windows Updates (Sullivan, 2015).

Table 4.19: Packet breakdown of 196.x.x.162 NXDOMAIN traffic

Month	# of packets	% of month NX packets	# of domains	Top domain	# of responses	# of source IPs
October 2013	47	0.000	1	example.fake	47	1
November 2013	-	-	-	-	-	-
December 2013	-	-	-	-	-	-
January 2014	-	-	-	-	-	-
February 2014	-	-	-	-	-	-
March 2014	-	-	-	-	-	-
June 2014	682	0.000	152	ns4.stileproject.com	57	126
July 2014	384 617	0.086	25 701	XxxXxxxxx-HP	4839	3 134
August 2014	204 984	0.051	18 482	XxxXxxxxx-HP	2583	1 944
September 2014	242 187	0.051	25 866	10.in-addr.arpa	6245	2 591
October 2014	443 409	0.069	39 807	local	11 929	2 682
November 2014	537 288	0.082	34 902	local	13 723	2 479
December 2014	203 914	0.034	16 186	local	6 693	1 377
January 2015	568 487	0.076	34 491	local	19 964	2 399
February 2015	595 689	0.073	26 889	local	29 322	1 421
March 2015	453 410	0.050	16 639	local	35 166	98
April 2015	207 477	0.028	17 214	mail	17 325	1208
May 2015	266 140	0.030	17 991	local	37 444	66
June 2015	243 009	0.018	24 085	local	32 473	78
July 2015	283 559	0.020	18 865	local	41 035	112
August 2015	270 744	0.028	17 118	mail	80 259	1 168

Another misconfiguration that is seen at this address is the presence of PTR queries for addresses defined in RFC 1918, in this case the 172.16.0.0/16 and 192.186.0.0/8 subnets (Rekhter *et al.*, 1996). This gives potential attackers insights into the IP address space used behind the NAT, and creates a security threat as it better enables them to target end-hosts behind the NAT itself. The presence of RFC 1918 address misconfigurations has been highlighted in other papers, most notably Zdrnja (2006).

The presence of the 47 *example.fake* NXDOMAIN queries in October 2013 stands out, as at that time none of the many server misconfigurations seen in later months were present. These queries seem indicative of malware that points to a static DNS domain hosted on a compromised end-host. All 47 queries occurred within 2 ms, and were sent to an IP located in Vietnam. It is also possible, however, that the query presence is as a result of local queries for a fake DNS domain leaking to the global DNS during Network File System (NFSv4) testing by setting up a fake nameserver[7].

4.3.3.2 .su ccTLD traffic

The .su ccTLD presence in the non-authoritative and non-caching NXDOMAIN dataset is sporadic, but contains a greater number of packets than those captured in the caching dataset. The *nitmurmansk.su* presence stands out as anomalous. Not only did it generate large volumes of packet traffic when compared to the other .su domains, it was also the most queried .su domain in November 2014 of the caching dataset, as seen in table 4.18.

[7] http://wiki.linux-nfs.org/wiki/index.php/Fake_DNS_Realm

Table 4.20: Packet breakdown of .su ccTLD traffic from other servers

Month	# of packets	# of domains	Top source IP	# of source IPs	Top destination IP	# of responses	Top domain	# of responses
October 2013	-	-	-	-	-	-	-	-
November 2013	-	-	-	-	-	-	-	-
December 2013	-	-	-	-	-	-	-	-
January 2014	13	1	8.8.4.4	1	196.x.x.210	13	finley.su	13
February 2014	-	-	-	-	-	-	-	-
March 2014	1	1	8.8.8.8	1	196.x.x.227	1	www.su	1
June 2014	-	-	-	-	-	-	-	-
July 2014	5	2	195.58.27.158	4	196.x.x.162	5	www.su	3
August 2014	2	1	195.58.1.145	1	196.x.x.162	2	ns.e-burg.su	2
September 2014	-	-	-	-	-	-	-	-
October 2014	61	5	155.232.135.5	2	196.x.x.162	61	lowbalance.su	20
November 2014	1 404	5	8.8.8.8	3	196.x.x.162	1 404	nitmurmansk.su	912
December 2014	1 170	4	155.232.135.5	3	196.x.x.162	1 170	nitmurmansk.su	896
January 2015	1 509	6	155.232.135.5	5	196.x.x.162	1 507	nitmurmansk.su	1 278
February 2015	735	5	155.232.135.5	3	196.x.x.162	735	nitmurmansk.su	603
March 2015	1 338	5	155.232.135.5	3	169.x.x.162	1 338	nitmurmansk.su	1 109
April 2015	51	5	155.232.135.5	3	196.x.x.162	51	nitmurmansk.su	26
May 2015	1	1	8.8.8.8	1	196.x.x.162	1	www.su	1
June 2015	24	6	41.0.1.1	21	196.x.x.162	6	invisible.msk.su	16
July 2015	135	3	193.232.156.17	7	196.x.x.80	134	crazyerror.su	118
August 2015	407	3	193.232.156.17	136	196.x.x.80	407	crazyerror.su	393

The *nitmurmansk.su* domain is interesting as it is the top domain in the Other dataset from November 2014, as seen in table 4.20, and is the top domain for the caching dataset for November 2014 as well. This domain is also responsible for more packets each month than many other months combined. The large increase in packet frequency suggests a malware infection trying to reach a server for commands and updates. This is supported by the fact that the packets are usually generated by three IP addresses, but in the case of 196.x.x.162, there could be multiple infected hosts behind the NAT.

4.4 Chapter Summary

This chapter builds on past work in the DNS operations sphere. Section 4.1 describes the current TTL implementations and practices seen on the Internet. The research shows that organizations are favoring lower TTL values, most likely because of the infrastructure flexibility that it provides, despite increasing network traffic and bandwidth cost. Many of the TTL values seen fall below the 15 minute value recommended by Wills and Shang (2000). Observed CDN TTL values are typically lower than other observed TTLs, ranging from 20 seconds to 10 minutes. The author believes that, overall, TTL values will decrease further as network performance increases in the future. The presence of large amounts of 0 TTL disposable domains is of interest, as this is created by DNSBL interaction. Jung and Sit (2004) noted an increase in DNSBL related traffic, and this data suggests that its presence has increased further, not only in quantity but also in variation of use.

The geolocation, and subsequent latency generation of authoritative servers for .za domains is discussed in section 3.6. The findings showed that the United States held the most unique authoritative servers, followed by South Africa, for the entire dataset. There were however large differences in authoritative server distribution for subsets of the .za domain dataset. The .org.za and .co.za domains were more likely to have international authoritative servers, while the .ac.za, .gov.za and (other).za domains were more likely to have local authoritative servers. The comparison of server location to generated latency showed that the large number of servers present at sites in the United States, Canada, and to a lesser extent Australia, were generating DNS-based latencies above the internationally observed average, and orders of magnitude higher than locally observed latencies.

Section 3.9 describes the NXDOMAIN response activity captured in the dataset. A large amount of this traffic was found to be generated by DNSBL services, which would use NXDOMAIN responses in their infrastructure to send confirmation that the tested address was not in the blacklist. After that had been filtered, it was found that the largest contributor to NXDOMAIN responses were server misconfigurations. Varying misconfigurations were captured, the most dangerous of which were WPAD queries, which give attackers information about WPAD server IDs, and PTR queries for addresses on the local network, which give attackers and monitors insight into the address block used by local networks. Filtering the discontinued .su ccTLD also revealed interesting packet activity, much of which could be considered an indicator of malware activity on the network.

This chapter built on the knowledge of the fields relating to DNS TTL values and NXDOMAIN response analysis, while also introducing new research from a South African context in the form of server geolocation for authoritative servers of .za domains. This is especially important, considering the evidence that latency times in the order of hundreds milliseconds, affect user Internet experience and site loyalty. Considering this, .za sites that target a local audience should consider shifting their authoritative server to a locally based alternative, in order to not suffer the cost of DNS-based latency on their userbase.

Chapter 5

DNS Abuse

This chapter focuses on the malicious use and abuse of DNS infrastructure. Section 5.1 deals with DNS amplification attack scans captured in the dataset, and will look at packet throughput as well as the temporal relationship between amplification scans captured and reported amplification attacks. Section 5.2 describes the methodology used for identifying bitflips, as well as the identification and subsequent filtering of false positives. Section 5.3 discusses the analysis of the identified bitflips, bitflip identifiers, and possible bitsquats detected in the dataset.

5.1 DNS post-attack amplification scanning

Denial-of-Service (DoS) attacks are a type of attack used by malicious entities in an attempt to limit or discontinue legitimate services connected to a network. DoS attacks fall under two main categories. Crafting packets with the intent to exploit vulnerabilities in the implemented software of the victim host is the first category, while the second category focuses on the consumption of critical system resources, e.g. network bandwidth (Kambourakis *et al.*, 2008), in an attempt to incapacitate the target host. Distributed Reflective Denial-of-Service (DRDoS) attacks are DoS attacks that use reflectors (Paxson, 2001), public servers that utilize UDP-based network protocols and respond to packet requests without the need for validation (Rossow, 2014), as part of the attack framework. The attacker will spoof the source IP of packets before sending them to the reflector, which will in turn forward the reply to the target host. These attacks are considered category two attacks as the aim is the consumption of resources and bandwidth. DNS amplification attacks are a class of DRDoS attack that exploits the fact that DNS protocols allow reply packets to be much larger than query packets. These attacks also take advantage of the fact that UDP packets are easily spoofed. This results in what is known as the amplification factor, which is the ratio of the size of the response to the request (Anagnostopoulos *et al.*, 2013). DNS amplification attacks are conducted using open resolvers, which are public resolvers that process queries from any client

(Rossow, 2014). The Open Resolver project (Mauch, 2013) has identified over 19 million servers that reply to DNS packets, just over 14 million open resolvers returning the correct query response, which is considered to pose a significant network security threat (as of 3 November 2015). One of the reasons an open resolver attack is so effective is as a result of caching, which enables the resolvers to send the attack packet from its cache rather than repeatedly querying the attack domain, significantly increasing the speed and throughput of the attack.

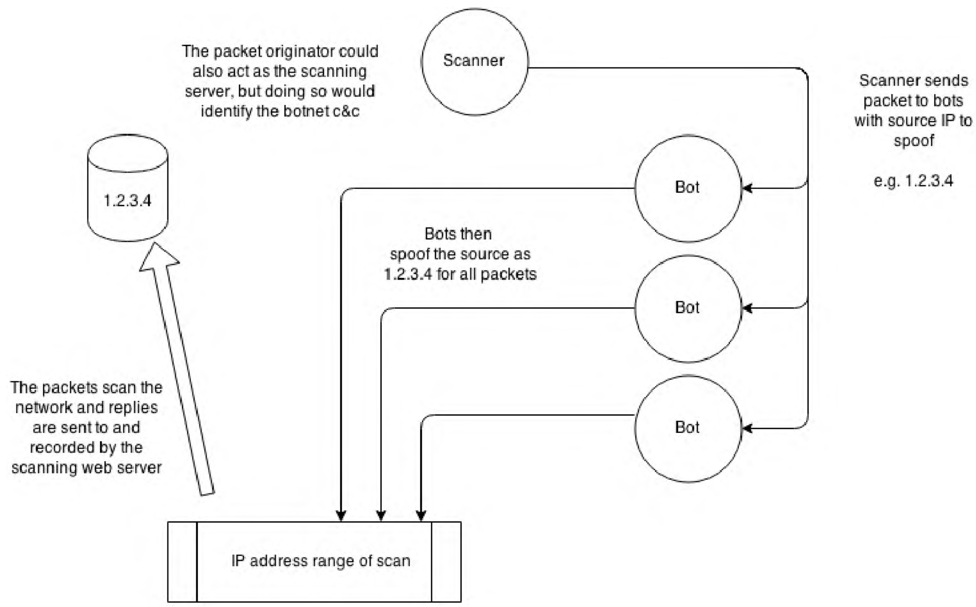


Figure 5.1: Architecture of a distributed DNS scan

Figure 5.1 describes a suggested architecture for a distributed DNS scanning framework. Such a framework would explain the large unique IP TTL presence observed for individual IP addresses within the dataset. The fact that the domains used in the scans are from known attack domains would also suggest that the scanners are looking for open resolvers that have records for, or will respond to queries for, those domains, instead of dropping the queries as a result of filter settings at the open resolver itself. It is believed that the scans showing only a single IP TTL value use the same IP address to send and receive packets, similar to the scanning architecture suggested by Fachkha *et al.* (2014).

5.1.1 Reported attacks

DNS amplification attack responses will only be seen by the targeted server, whose IP address has been spoofed in the query packets (Paxson, 2001). None of the IP addresses in the monitored /24 IPv4 block were the targets of DNS amplification attacks, and as a result only query packets were captured. This research relies on a website that reports on amplification attacks observed on a low bandwidth open DNS server [1]. The domains reported by the aforementioned are used to validate the fact that the captured scans are related to DNS amplification activity on the Internet. The attack dates, as well as attack packet sizes where referenced, are taken from the aforementioned and not recorded in the dataset itself.

5.1.2 Characteristics of captured scans

Table 5.1 outlines the composition of the captured scan traffic. A range of domains as well as target IP addresses were captured in the dataset. In the twenty-one months of data capture, only eleven unique domains appeared as the most frequently queried domain for any given month.

Table 5.1: Captured amplification scans

Month	# of packets	# of domains	# of target IPs	Top domain	% of monthly scans
October 2013	306 364	17	85	30259.info	41.965
November 2013	102 010	19	22	fkfkfkfa.com	14.793
December 2013	37 298	8	13	fkfkfkfa.com	39.183
January 2014	21 818	11	11	fkfkfkfa.com	24.938
February 2014	52 700	13	16	pddos.com	46.556
March 2014	44 505	6	15	ahuyehue.info	45.323
June 2014	6 009	9	15	magas.bslrpg.com	39.241
July 2014	9 952	8	21	wradish.com	42.042
August 2014	13 666	11	25	webpanel.sk	54.544
September 2014	10 902	7	20	webpanel.sk	56.366
October 2014	7 900	16	16	wradish.com	46.684
November 2014	6 148	8	16	wradish.com	24.691
December 2014	6 593	10	25	globe.gov	31.382
January 2015	4 338	8	16	gransy.com	21.047
February 2015	4 199	9	14	pidarasrik.ru	33.746
March 2015	4 974	6	15	defcon.org	45.678
April 2015	2 583	5	7	defcon.org	48.974
May 2015	3 044	9	7	defcon.org	29.928
June 2015	4 095	6	10	defcon.org	68.449
July 2015	2 793	4	5	defcon.org	74.472
August 2015	978	5	5	defcon.org	33.742

October 2013 is the largest amplification traffic source for individual packets, number of unique domains represented as well as overall IP representation. Apart from October and November 2013, none of the subsequent months show a six digit packet influx. The domain repetition seen in the top domain field is also notable, especially the *defcon.org* domain, which is the most commonly seen domain for six consecutive months, and shows the highest individual domain representation percentage in the monthly datasets. A case study of the October 2013 dataset as well as the *defcon.org* domain are given in subsections 5.1.5.1 and 5.1.5.3.

[1] http://dnsamplificationattacks.blogspot.co.za/

5.1.3 Temporal relation between scans and attacks

As the DNS Amplification Attack Observer only reports the day the attack was logged and not the time, all scans that occurred on the same day as the attack will be considered to have happened before the attack was reported. Table 5.2 makes reference to the scans that were recorded the day after the amplification attack using that domain was reported. A table describing the temporal relationship between all the captured scans of the dataset and the reported attacks can be found from tables A.12 to A.32 in the appendix.

Table 5.2: Scans recorded the day after attack was reported

Domain	Reported attack date*	First recorded scan	# of scans after attack	Last recorded scan
30259.info	9 October 2013	10 October 2013	17	22 October 2013
37349.info	15 October 2013	16 October 2013	44	18 October 2013
aa.10781.info	12 October 2013	13 October 2013	4	16 October 2013
babywow.co.uk	11 October 2013	12 October 2013	6	18 October 2013
gtml2.com	19 October 2013	20 October 2013	3	31 October 2013
krasti.us	18 October 2013	19 October 2013	1	19 October 2013
pipcvsemnaher.com	17 October 2013	18 October 2013	2	31 October 2013
cheatsharez.com	11 November 2013	12 November 2013	4	16 November 2013
reanimator.in	1 November 2013	2 November 2013	3	11 November 2013
siska1.com	9 November 2013	10 November 2013	2	17 November 2013
thebestdomainintheworld.cloudns.eu	15 November 2013	16 November 2013	1	16 November 2013
t.pbub.info	6 November 2013	7 November 2013	3	13 November 2013
x.mpnp.info	14 November 2013	15 November 2013	2	17 November 2013
x.slnm.info	17 November 2013	18 November 2013	1	18 November 2013
amp.crack-zone.ru	22 December 2013	23 December 2013	2	27 December 2013
grungyman.cloudns.org	17 December 2013	18 December 2013	2	22 December 2013
saveroads.ru	2 January 2014	3 January 2014	2	15 January 2014
x.xipzersscc.com	24 January 2014	25 January 2014	1	25 January 2014
gerdar3.ru	10 February 2014	11 February 2014	4	25 February 2014
ahuyehue.info	8 March 2014	9 March 2014	8	28 March 2014
www.jrdga.info	1 March 2014	2 March 2014	5	26 March 2014
lalka.com.ru	28 June 2014	29 June 2014	3	30 June 2014
webpanel.sk	23 July 2014	24 July 2014	5	31 July 2014
nlhosting.nl	17 October 2013	18 October 2013	1	19 October 2013
svist21.cz	12 November 2014	13 November 2014	3	20 November 2014

* all attack reports, as previously mentioned, are taken from dnsamplification.blogspot.com

Of the captured scans, 25 domains were scanned, sometimes by multiple IP addresses, the day after the attack was reported. There are many other scans that occurred from days to months after the reported attack, and a minority of scans that occurred shortly before the attacks were reported; as seen in tables A.12 to A.32. The data shows a clear link between attack dates and attack domain scanning behavior, particularly in the cases seen in table 5.2. There are multiple instances of scans being launched shortly after attacks have occurred, indicating that not only are these entities aware of the attacks, but that they attempt to take advantage of the domains used in the attacks themselves.

The relationship between attacks and amplification scanning is important as it offers researchers another method by which to study amplification attacks without having access to pcap files that capture the attack itself. By looking at captured query scans that match the amplification scan profile, they are able to determine with some certainty that the present domain has been, or will be, utilized in a DNS amplification attack.

5.1.4 Target spoofing

A summary of the packet behavior of the IP address that generated the most packets in any given month is outlined in table 5.3. The most anomalous result seen here is the number of unique IP TTL values seen for any given IP address. A large number of unique TTL values is an indicator of IP address spoofing (Jin *et al.*, 2003). IP address spoofing is a necessary part of DRDoS amplification attacks (Paxson, 2001), as this is how reply traffic is directed towards the victim.

Table 5.3: Top monthly spoofed IP behaviour

Month	Top IP	# of packets	# of IP TTLs	# of domains	Top domain	# of destination IPs
Oct 13	198.206.14.130	16 033	48	5	pkts.asia	253
Nov 13	80.82.64.231	16 388	56	9	siska1.com	253
Dec 13	94.102.56.229	9 429	56	3	amp.crack-zone.ru	253
Jan 14	94.102.56.229	6 882	56	6	saveroads.ru	253
Feb 14	46.105.111.230	10 401	54	2	pddos.com	253
Mar 14	46.45.178.250	6 995	240	1	www.jrdga.info	181
Jun 14	178.32.56.245	1 744	2	3	wradish.com	253
Jul 14	178.32.56.245	3 830	5	2	wradish.com	253
Aug 14	178.32.56.245	3 520	4	2	webpanel.sk	253
Sep 14	23.95.82.66	1 771	2	2	wradish.com	253
Oct 14	198.23.213.90	2 530	1	4	wradish.com	253
Nov 14	192.3.186.210	1 265	1	2	wradish.com	253
Dec 14	89.248.172.169	759	1	1	globe.gov	253
Jan 15	162.213.155.176	704	1	3	pidarastik.ru	253
Feb 15	162.251.118.42	1 012	2	3	pidarastik.ru	253
Mar 15	192.3.207.2	1 969	2	2	defcon.org	253
Apr 15	192.3.194.138	759	2	1	defcon.org	253
May 15	167.114.67.106	1 010	1	2	defcon.org	253
Jun 15	167.114.173.202	1 767	3	1	defcon.org	253
Jul 15	151.80.99.219	2 025	3	1	defcon.org	253
Aug 15	104.255.70.245	472	1	3	globe.gov	245

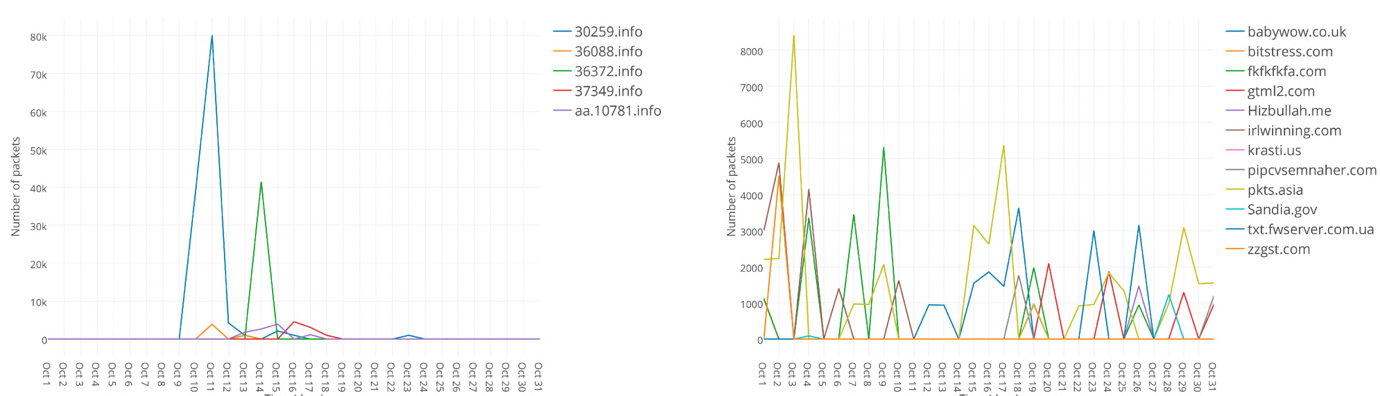
Some of the characteristics seen in table 5.3 point more strongly to scanning behavior than amplification attack behavior. Amplification attacks will almost always target open resolvers (Rossow, 2014) to increase the overall packet throughput to the victim (Fachkha *et al.*, 2014). This makes the traffic observed in the dataset anomalous, as there were not any operating open resolvers in the observed /24 IP block during the traffic capture period. Furthermore, all of the IP addresses in the /24 IP block, not including 196.x.x.0 and 196.x.x.255, were targeted by the captured traffic, suggesting scanning behavior. Scanning for DNS open resolvers is a suggested source of the observed packet traffic, but would usually indicate that the source IP address is not spoofed, so as to gather meaningful data from the reply packets sent from open resolvers (Fachkha *et al.*, 2014). This assumption is not supported by the large range of IP TTLs seen for many of the captured IP addresses.

5.1.5 Case studies

The following sections contain three case studies on the captured amplification query traffic. Subsection 5.1.5.1 gives a more detailed look at the month of October, the month that showed the largest packet influx as well as the presence of the most unique source IP addresses. Subsection 5.1.5.2 looks at traffic related to the *www.jrdga.info* domain, which shows the largest collection of unique TTL values in the dataset, as well as one of the highest individual domain packet counts. Subsection 5.1.5.3 looks at the *defcon.org* domain, which is of interest not only as a result of its popularity as a scanning domain, but because it serves as a legitimate domain which is being exploited, and not a domain under the control of a malicious host.

5.1.5.1 October 2013

October 2013 is the month that showed the largest number of individual packets, as well as the second largest unique domain subset. The dataset is presented in two sections, the first for traffic with the .info ccTLD and the second for all other domains.



(a) .info domains (b) Other domains

Figure 5.2: Timeseries for October 2013 scans by domain

The .info domain timeseries in figure5.2a shows the two highest packet frequencies seen for a single domain in a one day period across the entire dataset. The *30259.info* domain scans represent the largest packet influx and contains 17 unique source IP addresses. The attack was reported on 9 October, after which 39 160 packets were captured on the 10th, 80 159 packets captured on the 11th, and 4 254 and 976 packets captured on 12 and 13 of October 2013. Further scans were recorded on 15, 16 and 23 October, but did not show similar packet values to the two days after the attack.

The *36372.info* domain scans showed the second highest single day packet influx, totaling 41 484 on 14 October, the only day with captured scans. These scans are interesting as they all come the day before the attack was reported, and show eight unique source IP addresses. This seems to indicate that these scans were carried out in preparation for the attack launched the next day, and are not post-attack amplification scans, unlike many of the other captures.

The *37349.info* scans are also notable. As is seen in figure 5.2a , there are only three scans recorded, the first coming the day after the attack was reported. While the packet figures are only 4 547, 3 045 and 945 packets for the three days that the scans were present, these domain scans showed the highest concentration of unique IP addresses in the dataset, totalling 44.

Of the 43 scans pictured in figure 5.2b, only four of them were captured pre-attack. One scan was captured for the *irlwinning.com* domain the day before the attack was reported, while a further two were captured on the day of the attack. One scan was also captured on the day of the attack report for the *pkts.asia* domain; the other 39 were all post-attack amplification scans. The *pkts.asia* domain stands out as it was the most commonly queried domain for the top source IP packet provider as seen in table 5.3, and also have scans that were captured across the entire month, including the first and last day, despite not appearing throughout the rest of the dataset. Overall there were only 11 unique IP addresses that scanned for the *pkts.asia* domain.

5.1.5.2 www.jrdga.info

Figure 5.3 is a timeseries of scanning activity for *www.jrdga.info* during March 2014. This domain was selected for the case study due to the large unique IP TTL presence in the scans, the highest seen at any one time. The first scan, comprising 940 packets, comes one day after the attack was reported, as with many of the attacks mentioned in subsection 5.1.3. After a period without activity, there is another scan on the 9th, followed by a collection of 3 separate IP scans from the 24th to the 26th of the month. The most interesting aspect of this is the scan by 46.45.178.250 on the 24th and 25th. The first scan shows 237 unique IP TTL values for the 6244 packets of the same source IP, indicating a greater botnet size than for most other captured scans. While the scan on the 25th by the same IP address shows only 37 IP TTLs, and results in much fewer packets, there are unique IP TTL values present that were not recorded in the first scan, indicating that the botnet may have used different hosts to launch the second scan.

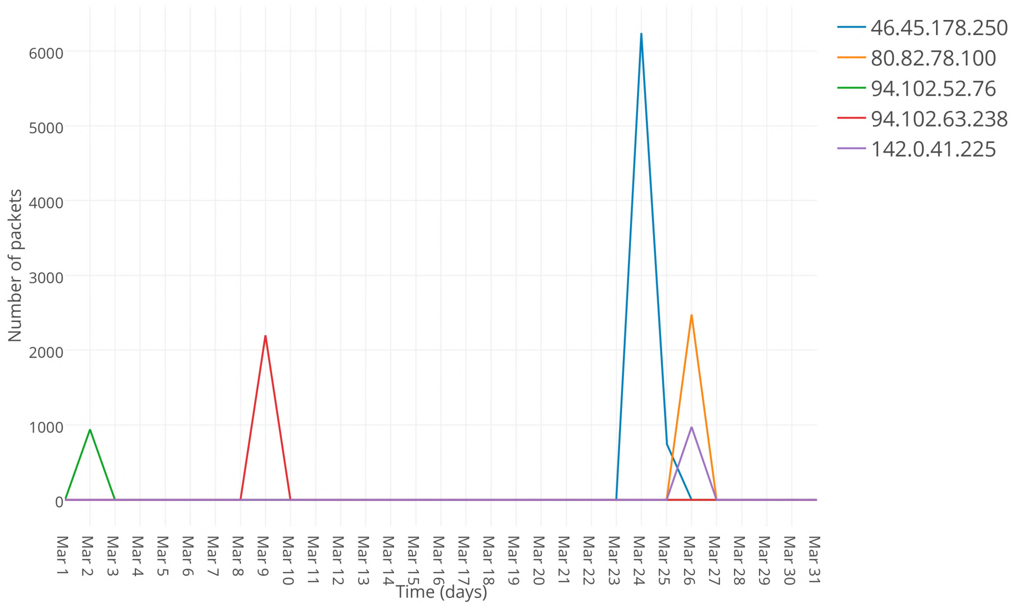


Figure 5.3: Timeseries of www.jrdga.info packets captured during March 2014

A breakdown of source IP activity is given in table 5.4. After the large scans observed in March, two additional but smaller scans were captured in July and September 2014, four and six months after the reported attack respectively. The July scan targetted all 253 IP addresses of the observed /24 IP block over the course of two days, and showed only one IP TTL value for all packets, indicating that the scanning server was both sending and receiving packets. The September scan targetted only 35 IP addresses, most likely as a result of random IP address generation, and showed 11 unique IP TTL values, lower than the values observed in March.

Table 5.4: IP characteristics of www.jrdga.info scans

Date	Source IP	# of packets	# of IP TTLs	# of destination IPs
2 March 2014	94.102.52.76	940	42	25
9 March 2014	94.102.63.238	2199	46	202
24 March 2014	46.45.178.250	6244	237	181
25 March 2014	46.45.178.250	751	37	78
26 Mar 14	80.82.78.100	2478	96	187
26 March 2014	142.0.41.225	976	41	122
3 July 2014	94.102.49.178	84	1	84
4 July 2014	94.102.49.178	169	1	169
17 September 2014	162.212.181.242	35	11	35

It is also worth noting that another domain scan in March 2014 returned 240 unique IP TTL values. A scan on 14 March for *ahuyehue.info* resulted in 5466 packets to 126 IP addresses in the observed /24 IP block. The number of unique TTL values as well as the fact that both scans targetted less IP addresses than the total block comprises suggests that both scans were carried out by the same botnet, despite the source IP address of the scan in question not matching that seen for the *www.jrdga.info* scan.

5.1.5.3 defcon.org

The *defcon.org* scans offer an interesting case study for multiple reasons. The first is that the targeted domain is a legitimate domain, and not a domain controlled by a malicious host. The second is the nature of the scanning captured in the dataset. As seen in table 5.5, the scanning is much more uniform, and there is less evidence of packet clumping as is seen with other domains, which produce thousands of packets in a single day.

Table 5.5: Number of packets received for defcon.org

Day	Dec 14	Jan 15	Feb 15	Mar 15	Apr 15	May 15	Jun 15	Jul 15
1	-	253	48	-	-	-	-	-
2	-	253	-	-	-	-	-	-
3	-	-	-	-	-	-	-	84
4	-	-	-	-	-	-	-	602
5	-	-	-	-	-	-	-	327
6	-	237	253	-	-	-	-	253
7	-	-	-	253	-	253	-	253
8	-	-	-	-	-	-	-	-
9	-	-	-	-	253	-	-	253
10	-	-	-	-	-	103	-	-
11	-	-	-	253	-	-	253	253
12	-	-	-	-	-	-	249	302
13	-	-	-	-	-	-	-	-
14	-	-	-	253	-	-	506	-
15	-	-	228	-	253	-	198	-
16	-	-	253	-	-	-	55	-
17	-	-	-	-	-	-	194	-
18	-	-	-	253	-	-	312	-
19	-	-	-	-	-	51	125	-
20	-	-	-	-	253	253	128	-
21	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	253	-
23	-	-	-	253	-	-	-	-
24	224	-	-	-	-	-	253	-
25	325	-	-	253	253	-	-	-
26	271	-	-	-	-	-	-	-
27	253	-	-	-	-	251	-	-
28	55	-	-	253	-	-	254	-
29	-	-	-	-	253	-	23	-
30	266	-	-	253	-	-	-	-
31	-	-	-	-	-	-	-	-
# of IPs seen	9	3	5	3	2	3	4	3

The scanning behavior is also different from other captured domains in the sense that the same source IP address will perform multiple individual scans on the domain in a given month, as seen by the difference in packet representation between table 5.3 and packet distribution in table 5.5, particularly for June and July 2015. Table 5.6 shows that this behavior is not limited to a single month, as some source IPs scan the same IP block with the same domain in different months as well. IP addresses that appear in multiple months have been bolded.

Table 5.6: IP addresses scanning defcon.org domains

Dec 14	Jan 15	Feb 15	Mar 15	Apr 15	May 15	Jun 15	Jul 15
198.7.63.129	173.242.112.113	96.8.115.114	192.3.207.2	192.3.194.138	**167.114.67.106**	**167.114.173.202**	151.80.99.219
104.218.48.7	141.255.164.162	23.94.1913.82	64.16.211.238	167.114.67.106	178.19.106.10	167.114.210.12	199.168.139.139
46.36.37.81	162.213.115.176	172.245.24.154	167.114.114.98		**167.114.173.202**	63.141.227.10	172.73.123.160
162.251.114.66		209.105.232.87				**167.114.67.106**	
64.6.108.171		**46.19.137.234**					
46.19.137.234							
192.129.201.106							
192.3.34.2							
141.255.166.210							

This is theorized to come about as a result of the *defcon.org* domain being a legitimate domain. Responding servers or open resolvers that choose to drop packets for this domain during attacks may choose to accept them again at a later stage, as legitimate queries for this domain will no doubt be seen. This would also explain the elongated scanning pattern, spanning seven months, as attackers are seeking exploitable servers which may have been reconfigured to allow replies for these queries once more.

5.1.6 Bandwidth amplification factors of domain queries

Table 5.7 shows the bandwidth amplfication factors of some of the scans captured in the dataset. No responses were captured, as none of the scanning IP addresses or attack targets were present in the observed IP block. As such, the response sizes are taken from the attack reports on *dnsamplification.blogspot.com.*

Table 5.7: Bandwidth amplification factor of queried domains

Domain	Query size (bytes)	Response size (bytes)	BAF
30259.info	87 and 99	4 211 [2]	48.402 or 42.535
36372.info	99	4 211 [3]	42.535
37349.info	87 and 99	4 211 [4]	48.402 or 42.535
www.jrdga.info	91	4 112 [5]	45.187
defcon.org	87	4 084 [6]	46.943

Bandwith amplificaiton factor (BAF), is the factor by which bandwidth consumption is increased between the query packet and reply packet. It is calculated by dividing the size of the response packet by the size of the query packet for that response, an equation for which can be seen in section 2.3.2. The BAF recorded here are higher than the average recorded open resolver BAF but lower than the average name server BAF recorded in Rossow (2014). The values are closest to the BAF average recorded for the worst 50% of ANY lookup amplification attacks observed at open resolvers (Rossow, 2014), which would make sense given the large packet response size. The DNS infrastructure used to limit packet size to 512 bytes (Anagnostopoulos *et al.*, 2013), which suggests that these attacks targetted EDNS0 enabled open resolvers (Vixie, 1999) to achieve the 4KB response size; which also explains why the amplification factors are closer to the worst 50% of amplification attacks rather than the overall average (Rossow, 2014).

5.2 Bitflip analysis

This section deals with bitflip identification and analysis in the dataset. Bitflipping and bitsquatting were covered in subsection 2.1.8. The bitflip analysis focuses only on the domains present in the dataset, and does not concern itself with the possible flips seen for other data in the packets.

5.2.1 Bitflip identification approach

First the monthly datasets were filtered so that only a unique list of all seen domains in that month remained. The list of domains was then processed to form a list of binary domains.

```
for line in f:
    h = [bin(ord(ch))[2:].zfill(8) for ch in line]
    for x in h:
        i += x
    i = i[:-5]
    g.write(i+"\n")
    i = ""
```

In Python, *ord()* is a built-in function that returns the value of the byte if the argument is an 8-bit string, given a string of length 1[7]. The *bin()* function is a built-in function that returns a binary string when given an integer. The *bin()* function returns binary in the form 0bxy where x is the highest positive bit and y is any combination of positive and negative bits. As a result `[2:]` is used to strip the first two characters of the bitstring, while `.zfill(8)` pads the front of the binary string until there are 8 bits, to represent a 1 byte character. The `i = i[:-5]` is there to remove the bits generated by the newline character, which in hindsight would have been much more elegantly solved by calling `.strip()` on the line.

The following code was then used to check if there were possible bitflips among the given bitstrings.

[7] https://docs.python.org/2/library/functions.html#ord

```
def is_bitflip(s1, s2):
    if not len(s1) == len(s2):
        raise Exception("Strings are not equal length")
    bitsflipped = 0
    for index in range(len(s1)):
        if s1[index] != s2[index]:
            bitsflipped += 1
        if bitsflipped > 1:
            return False
    return bitsflipped == 1
```

This function works by comparing the values of bits in two given strings. The first test is to see if the two bitstrings are of equal length, and if they are not, to discard the comparison as it is not a possible bitflip. The loop then iterates through both strings using the index value of the characters in the string, comparing the bits. When two non-equal bits are found, the `bitsflipped` counter is increased by one. A second test throws out the strings if there are more than one bit differences present at any given time. The function then returns the boolean value `bitsflipped == 1`, which will evaluate to true if there is a single different bit in the bitstring. The results given by the code were then tested using a completely different algorithm as a sanity check.

```
def rec(x):
    rc = math.log(x,2)
    return (rc == int(rc) and 2**rc == x)
if len(line) != len(q):
    continue
y = int(line,2) ^ int(q,2)
if y == 0:
    continue
z = rec(y)
```

This function receives an integer value, y, which is the integer value of the resulting XOR of the two bitstrings. The function then calls the built-in log() function, which will determine whether or not the XOR'd integer is a power of 2, i.e. a single bitflip, by using log with base 2, and expecting a non-decimal positive number. The first function test is to determine if the log of the power returned an integer, indicating a single flip, while the second test is a sanity test of the log function itself, which would sometimes return integer values to logarithms that were not precisely a power of 2.

Both functions returned the same output on test cases as well as the binary lists generated from the datasets.

5.2.2 False positives and filtering

The bitlfip identification algorithm seen in section 5.2.1 resulted in many false positives as a result of domain naming conventions, DNS infrastructure usage and the structure of certain RRs.

5.2.2.1 Domain names

Many domain names, especially ones that are not used by human clients or do not resolve to user-content hosting IP addresses, will use numbering as a naming convention instead of a name targeted towards consumers. An example of this is *a1163.phobos-apple.com.akadns.net* and *a1063.phobos-apple.com.akadns.net*, which are both valid domains that form part of the **Akamai CDN** infrastructure, but return a positive bitflip as 1063 and 1163 have a 1 bit difference while all the other bits are identical. The *s3-3-w.amazonaws.com* and *s3-1-w.amazonaws.com* domains are another example, both valid **Amazon** domains that register a bitflip as a result of 1 and 3 having a one bit difference. Efforts were made to filter out bitflips generated by naming conventions by filtering the domain dataset through name server resolution of the domain. Domains that were not found to be valid were of greater interest as they constituted a higher chance of being a true bitflip.

5.2.2.2 Use of DNS infrastructure

Services like DNSBL also generated false positives, as they append an IP address to their domain, for example *93.x.x.196.zen.spamhaus.org* and *92.x.x.196.zen.spamhaus.org*, which register as a bitflip due to the 3 of 93 and 2 of 92 having a one bit difference. This is only problematic due to these queries passing through the filter that tested the validation of domains, as queries for IPs will return NXDOMAIN responses, as mentioned in subsection 4.3.2. Efforts were made using pattern matching to filter these false positives from the existing datasets.

5.2.2.3 PTR queries

PTR queries accounted for a significant portion of false positives. For example, the *93.0.168.192.in-addr.arpa* and *91.0.168.192.in-addr.arpa* queries returned a positive bitflip as well as the *93.0.168.192.in-*

addr.arpa and *83.0.168.192.in-addr.arpa* queries. It was decided that the PTR queries would be filtered from the datasets before further processing as a result of the number of false positives generated in the dataset, more so than any other query type. They were filtered through pattern matching and not by RR, to ensure that only numerical flips are removed.

5.2.2.4 Further filtering

A number of legacy domain names were not filtered out by resolving hostnames, as they were no longer valid domains, and as such were removed manually from the datasets.

5.3 Bitflip findings

This section details the observations made during the analysis of possible bitflips left after the filtering stages. Section 5.3.4 showcases some of the possibly squatted domains that have been identified throughout the analysis.

5.3.1 Case insensitive nature of DNS

Some of the possible bitflips captured for domains are domains where one of the characters of the domain is uppercase, i.e. *fsmx.async.org.za* and *fsmx.async.Org.za.* As a result of the case-insensitive nature of DNS (Eastlake, 2006), it becomes difficult to say with certainty whether or not the observed domain difference is as a result of a flipped bit or simply different configuration at the querying host. It also hinders the filtering of possible bitflips through domain validation, as those domains will count as valid due to the nature of DNS. Case-flipped queries that resolved to an active domain were ignored, as mentioned in section 5.2.2.4.

The presence of case-flipped letters was also noticeable in the domain dataset for non-existent or non-resolvable domains. For the unresolved async.org.za domain, the permutations asyNc.org.za, Async.org.za, async.oRg.za, asynC.org.za, async.Org.za , async.org.Za were all captured in June 2014. All of these have a bit difference of one from the original domain. Another case from the same month is moria.org, which resulted in the permutations moria.Org, moRia.org, moriA.org, morIa.org , mOria.org, Moria.org, moria.orG.

For both domains, we see flips appearing in the domain itself, the TLD and the ccTLD. This, coupled with the fact that the original domain is non-resolvable, seems to indicate that it is more likely to be a flipped bit as opposed to different case configurations by end-hosts. A possible explanation for this case inconsistency is given in section 5.3.2.

5.3.2 Recorded IN-ADDR.ARPA flips

It was mentioned in 5.2.2.3 that PTR records had been filtered out, due to the generation of false positives through the presence of an IP address in the domain. While the addresses themselves were problematic, these queries form an interesting case study as there is evidence of case change in these domains as well.

5.3.2.1 Example Flips

166.x.x.196.IN-ADDR.ARPA

167.x.x.196.IN-ADDR.ARPA

These two queried domains registered as a bitflip due to 166 and 167 being one bit apart. This is an example of a falsely identified bitlfip. Almost all of the PTR queries were listed as flips due the the close nature of the queried IPs, which exist in the observed network block. A small subset of the captured queries show interesting domains that are indicative of true bitflipping.

167.x.x.196.IN-ADDR.ARPA

167.x.x.196.IN-ADDR.ARPa

In this example, the bitflip was not detected as a result of IP differences, but because of a case difference in the domain extension. Table 5.8 shows the unique permutations captured for this domain extension during August 2015. There were 39 TLD bitflips recorded for IN-ADDR.ARPA in that month. The flipped bit has been bolded. This is similar to the case inconsistencies for the *exodus.desync.com* domains mentioned in section 4.1.3.

Almost all of the PTR bitflip traffic was captured at the two authoritative servers.

Table 5.8: Permutations of PTR query domain extensions August 2015

Extension	Expected ascii	Captured ascii	Expected bits	Captured bits
IN-AdDR.ARPA	D	d	01000100	01**1**00100
IN-ADdR.ARPA	D	d	01000100	01**1**00100
iN-ADDR.ARPA	I	i	01001001	01**1**01001
In-ADDR.ARPA	N	n	01001110	01**1**01110
IN-aDDR.ARPA	A	a	01000001	01**1**00001
IN-ADDr.ARPA	R	r	01010010	01**1**10010
IN-ADDR.ARpA	P	p	01010000	01**1**10000
IN-ADDR.ArPA	R	r	01010010	01**1**10010
IN-ADDR.aRPA	A	a	01000001	01**1**00001
IN-ADDR.ARPa	A	a	01000001	01**1**00001

This case difference seen for these addresses, among others, was mentioned in section 5.3.1. It is interesting to note that all of the bits are flipped at the same position in the letter byte array. The uniformity of the flipped bit suggests that this bitflip may be caused by a software or hardware related issue in this case, and is not due to happenstance. It is also possible that it is a 0x20 bit hack (Wessels, 2012). Manipulation of the 0x20 bit in the domain was suggested as a security feature to increase the difficulty of cache poisoning attacks (Dagon *et al.*, 2008), as this would force poisoners to guess the correct Capital Sequencing of the domain for pattern matching. This configuration is also most likely the cause of the registered flips mentioned in section 5.3.1. The irony of this is that cache poisoners are increasing hit chances by flipping the 0x20 bit in domain names, which are seen as case-insensitive by the receiving servers (Vixie and Dagon, 2008).

5.3.2.2 Frequency of flips for IN-ADDR.ARPA queries

Table 5.9 looks at the IN-ADDR.ARPA case flips present. The number of packets refers to the number of queries in the dataset that had an uppercase IN-ADDR.ARPA in the domain name, while the number of flips refers to the number of packets with a case-flipped letter, for the months in the dataset. While the number of flipped packets is low, the percentage of captured packets is much higher than bitflip frequency is suggested in other research (Wessels, 2012). The ratio of packets to IP addresses also suggests that these bitflips are the result of a system configuration or error rather than a memory error.

Table 5.9: Flip frequency for IN-ADDR.ARPA packets

Month	# of flips	# of packets	% of packets	# of IPs
Oct 13	22	7 952	0.277	12
Nov 13	42	10 292	0.408	20
Dec 13	24	11 072	0.217	12
Jan 14	14	5 380	0.260	9
Feb 14	14	6 815	0.205	8
Mar 14	23	7 344	0.313	9
Jun 14	19	5 855	0.325	8
Jul 14	34	8 444	0.403	18
Aug 14	28	6 615	0.423	13
Sep 14	32	8 350	0.383	17
Oct 14	50	9 993	0.500	24
Nov 14	58	10 107	0.574	26
Dec 14	35	6 645	0.511	15
Jan 15	52	12 909	0.403	19
Feb 15	53	13 373	0.396	26
Mar 15	91	16 529	0.551	44
Apr 15	173	53 586	0.323	48
May 15	82	9 624	0.852	48
Jun 15	83	11 216	0.740	47
Jul 15	112	13 561	0.826	50
Aug 15	39	11 155	0.350	29
Total	997	246 817	0.404	62

The fact that so few packets were captured in any given month, and also that there was no evidence of domain clumping, suggests that these queries come from systems or end-hosts that have implemented 0x20 bit encoding (Dagon *et al.*, 2008) to increase their DNS cache security, and not as the result of an attempted cache poisoning attack. Such a defense could cause problems however if a numerical character is flipped, in which case the domain would resolve differently.

5.3.3 Recorded bitflips

This section looks at some of the flips captured in the dataset. Some of the flips are categorized as possible typos, and will be discussed. Other captured domains are almost certainly flips as they deviate from DNS naming standards (Mockapetris, 1987b).

5.3.3.1 Possible Typos

Some of the bitflips were more likely a typing error that resulted in the string being one bit different. Moore and Edelman (2010) define a measure of distance called *fat finger distance* for measuring the likelihood of a typo in a domain, and also the likehood of a domain being typosquatted. This distance is one adjacent key on the keyboard from the desired key.

Table 5.10: Bitflipped domains as a result of typos

Domain	Adjacent character
ggogle.com	-
www.facebooj.com	k
www.youtbe.com	-
www.youube.comq	-
www.fqcebook.com	a
wsw.etoro.com	w
www.ru.ac.xa	z
yqhoo.comq	a
sww.saprepschool.com	w
googld.com	e
kingswoodcollegd.com	e
facebokk.com	o

These bitflips are all most likely the result of typing errors instead of a memory error. All but three of the examples have letters within fat-finger distance of the typo. The other three are interesting as they do not follow this pattern. It is suspected that the g key was tapped twice when the domain was inputted in the first case, causing the error. The *youtbe.com* and *youube.com* cases are clearly an error in character ommision, but are mentioned as they were recorded as bitflips for one another.

This adds an interesting dynamic to bitflipping, as it could be used to statistically enhance the chance of traffic to a typosquatted domain if they register a domain that is also a bitsquat of the original domain.

5.3.3.2 Definite Bitflips

Examples of bitflips captured in the dataset are given in table 5.11. The tilde (~) in the *async.org.za* domains is interesting as it is the result of a flip at the 0x04 bit for lower z. Five seperate addresses saw a flip at the 0x04 bit of the letter d. These cases are all examples of bitflips invalidating addresses, as they no longer conform with domain name specifications (Mockapetris, 1987a). This means that some bitflips, due to the nature of the character the flip produces, cannot be bitsquatted.

Table 5.11: Bitflipped domains

Domain	Domain
ns1.async.org.~a	i‘entity.apple.com.akadns.net
ns2.async.org.~a	c‘n.spotxchange.com
kingswoodcollage.com	plus.coogle.com
kingswoodkollege.com	talkomsa.net
ww7.sacschool.com	speampowered.com
www.goocle.com	c‘n.fastclick.net
eray.com	c‘n.spotxchange.com
r‘13p04sa.guzzoni-apple.com.akadns.net	a‘xhm.d.chango.com

Other bitflips are also observed in letter changes, which could be the result of typos. The distance between the letter and its substitute is more than the fat-finger distance on a QWERTY keyboard, making it more likely that these are true bitflips.

5.3.4 Possible Bitsquats

The sites that were filtered from the main dataset were processed so that only the TLDs remained, and those were put through the Linux command `sort -u` to create a list of domains, from which possible bitsquat domains were identified.

The site *barcleys.com*, which is a bitflip for *barclays.com*, displays a blank page when visited. This site is considered as empty. Around 2.7% of bitsquatted sites deliver no content at all (Nikiforakis *et al.*, 2013). The *foogle.com* domain returns 'Coming soon'. The three domains *cmail.com, hotmaal.com* and *watppad.com* are listed as for sale on the site. Domains for sale make up roughly 10% of bitsquatted domain sites (Nikiforakis *et al.*, 2013).

The *fabebook.com* domain is most likely a bitsquat while *webme.com* is most likely a typosquat of webmd. Both redirect traffic to unrelated sites. *The goggle.com* domain is a known typosquat domain, which is coincidentally also a bitsquat. The site tries to persuade visiters to sign up for a £3-per-text quiz competition, offering **Apple** merchandise as prizes[8].

The verixon.net bitsquat is owned by **Verizon**, and relocates to their home page *www.verizon.net*. It is surprising that, of the subset, this is the only squatted domain owned by the organisation that is being squatted.

Table 5.12looks at the bit difference between the squatted domains and the target domains. The flipped bit distribution is greater than the Case-flipped domains. These bitsquatted domains are all squatting well known domains, which confirms past research (Nikiforakis *et al.*, 2013).

Table 5.12: Bitflip seen in Bitsquatted domains

Domain	Expected ascii	Captured ascii	Expected bits	Captured bits
barcleys.com	a	e	01100001	01100**1**01
cmail.com	g	c	01100**1**11	01100011
fabebook.com	c	b	0110001**1**	01100010
foogle.com	g	f	0110011**1**	01100110
goggle.com	o	g	0110**1**111	01100111
watppad.com	t	p	01110**1**00	01110000
webme.com	d	e	01100100	0110010**1**

It seems that, in line with the findings of Nikiforakis *et al.* (2013), while there is a definite bitsquatting presence on the Internet, very few of the bitsquatted domains are tailored to serving malware, and more often than not are simply using the domain to generate revenue through domain parking or sale; or owned by the same owner of the legitimate domain that is being squatted.

5.4 Chapter Summary

This chapter covers three main topics. The first is post-attack amplification scanning activity, which is covered in section 5.1. A lot of work has been done around DNS amplificaition attacks, but to the knowledge of the author this is the first time that amplification scanning behavior has been linked to amplification attacks that have already been carried out, sometimes months before the scans themselves occur. This behavior is important for a number of reasons. Firstly, it allows researchers that have access to darnket packet captures to infer amplification attacks through observed scanning activity. It also indicates that possible attackers will attempt to take advantage of attack domains used by other parties. There were many instances where amplification scans

[8] http://www.theregister.co.uk/2011/10/12/google_v_goggle/

were captured the day after the attack was reported, strongly suggesting a temporal link between attacks and post-attack scanning.

Section 5.2 outlines the code used for possible bitflip detection. Two methods were used as a confirmation of accuracy with respect to bit differences. The largest issue encountered was the number of false positive bitflips encountered during processing. Many domains have numbers attached to identical domain strings as part of their naming convention, including the PTR RR, which resolves an IP address. These domains all registered as bitflips as a result of the one bit difference between the numbers in the domain name. A number of filtering strategies were attempted in order to seperate true bitflips from false positives. First, all of the PTR queries that showed digit flips instead of character flips were filtered out. Second, domains were filtered by their ability to be resolved, in order to preserve genuine digit bitflips while excluding similar but actively registered domains. DNSBL flips that occurred on digits, i.e. of the IP addresses, were also filtered out from the non-resolving domain dataset. Lastly, certain legacy domains had to be manually filtered. The need to manually filter certain flips is also mentioned in Dinaburg (2011), indicating that he encountered a similar problem with automatic filtering algorithms.

The last section, 5.3, describes the results of the bitflip analysis. The first thing to note here is the large presence of flipped 0x20 bits, causing a change in capitalization of the domain. While it is technically a bitflip, it is believed to be the result of software configuration (Dagon *et al.*, 2008) and not the result of random occurrence or machine malfunction. Very few definite bitflips were found considering the number of false positives, although it is possible that some actual bitflips were filtered out accidentally, it is better to err on the side of caution. The example bitflips showed a much wider variety of positions of the flipped bit, which further confirms the assumption that the large number of 0x20 flipped bits are not true bitflips. Of the confirmed bitflips, many domains were malformed, as some of the flipped bits resulted in domains that do not conform to domain name specifications. Finally, some examples of Bitsquatting were identified from the dataset. Similar to the findings reported by Nikiforakis *et al.* (2013), many of the squatted domains were either owned by the legitimate domain company, were parked domains, ad-revenue domains, or pages that redirect to unrelated content.

Chapter 6

Conclusion

This chapter reflects on the goals identified at the start of the thesis, and evaluates to what level they have been achieved. It also discusses some of the important findings of the thesis, as well as why they are important. Concluding remarks are then followed by recommendations for future avenues of study in the field of DNS traffic analysis.

6.1 Reflection on goals

The first goal outlined in the thesis was to gain an understanding of how legitimate services and end-hosts were currently utilizing DNS, in an attempt to better understand DNS packet activity and configurations on the Internet. To achieve this three separate studies were conducted on DNS TTL usage, DNS authoritative latency for local domains and NXDOMAIN presence in the dataset. This goal was achieved in the following ways: The thesis offers extensive analysis of DNS TTL implementation and behavior from a wide variety of authoritative servers, spanning a number of resource records, and representing different domain needs. The geolocation of .za authoritative servers is original work that builds on creating a locally contextualised understanding of the spread of authoritative servers for local domains. The evaluation of DNS-based latency generation also works to achieving this goal, as it creates a clearer understanding of the latency costs implicit in authoritative server choice. The NXDOMAIN analysis sheds light on the cost of host misconfiguration to the network, as well as some of the security threats that are generated by misconfigurations that result in NXDOMAIN responses. The large DNSBL presence found in both the TTL analysis, as well as the NXDOMAIN analysis, covers a large area of current common DNS utilization in third party services. This goal, however, could not be completely realized as the covered sections are not able to fully encompass current DNS practice and utilization. There is simply too much data, with numerous avenues of study, to be included in a single thesis.

The second goal was to investigate instances of malicious DNS behavior and DNS abuse. This goal was met through analysis on two distinct forms of DNS abuse, amplification-attack scanning and bitsquatting. The first step towards this goal was a thorough analysis of the temporal relationship between DNS amplification DDoS attacks and DNS amplification scanning using attack domains. This thesis found a direct correlation between attack reports and scanning behavior, and was able to link amplification scans to attacks that had been reported the day before, as well as scans for attacks that happened months before. The second step to achieving this goal was an analysis of possible presence of bitflips within the dataset. The successful identification of confirmed bitflips, as well as analysis on observed Bitsquatted domains captured in the dataset, shed more light on a relatively new field in DNS security. While the same holds true for both sections in the sense that there is simply too much data to cover all of the captured malicious DNS activity, the author believes that the second goal has been fully completed. The research on this area is original, relevant to current DNS threats, and an important foundation for future work in the case of post-attack amplification scanning.

6.2 Key Findings

This section outlines what the researcher believes to be the most important findings arising from the research conducted, and will discuss the reasons they are considered to be so. The findings on latency relate to the first research goal, and offer solutions for improving DNS operational ability as well as user experience with respect to DNS. The findings on amplification scans and bitsquatting relate to the second goal, and shed light on how malicious entities abuse the DNS infrastructure to achieve their own ends; as well as possible means of combating that abuse.

6.2.1 Latency and its cost

Several papers (Ramsay *et al.*, 1998; Brutlag, 2009; Vulimiri *et al.*, 2012) highlighted the effects that experienced latency has on user perception of and interaction with web-hosted content. The inversely proportional relationship between user experience and experienced latency indicate a clear need for content webhosts to consider latency when implementing their network infrastructure. This assertion becomes relevant when considering the geolocation of authoritative servers for .za domains, as seen in section 4.2. Around 40% of unique authoritative servers were placed in the United States, generating between 200 ms and 300 ms of latency. The percentage of international servers outweighed local authoritative servers over the total subset of data, indicating that many .za webhosts, of which presumably some target ZA residents as a primary consumer base, introduce hundreds of milliseconds of DNS-based latency. By switching to local authoritative servers, they could see an increase in site revenue while also improving end-user experience.

6.2.2 Temporal relationship between amplification attacks and scans

Confirming a temporal relationship between amplification attacks and post-attack amplification scanning, seen in section 5.1, allows for the inference of amplification attack activity from datasets that did not capture the attack itself. This is important from a security perspective as it shows that known attack domains pose a threat to DDoS targets even after an attack has been launched. As a result, the computer security community should move towards not only filtering packets by IP or domain name to prevent DDoS attacks, steps should also be taken to deregister known attack domains, while also decreasing the number of open resolvers present on the Internet. This finding is also relevant from a researcher's perspective as it allows researchers to infer DDoS amplification attack activity without capturing packets of the DDoS itself, but by observing the times and numbers of recorded DNS amplification scans.

6.2.3 Confirmation of other Bitsquatting research

Of the Bitsquatted domains identified in section 2.1.8, all of them fell into the groups identified by Nikiforakis *et al.* (2013). There was also a similar distribution of different bitsquatting behavior to the results delivered in that paper. The confirmation of the results of others is important in relatively new fields of study, of which Bitsquatting is undoubtedly one. The most important finding, however, is that the prevalence of bitsquatted domains is far greater than the prevalence of malware-serving domains. This indicates that the current threat from observed bitsquatted domains is much lower than it could be. While this is a positive result, it leaves a lot of room for future illegitimate activity to develop, and the author believes that action should be taken to prevent or mitigate future threats created as a result of bitflipping.

6.3 Future work

- There is a lot of work that could be done on TXT RR data mining. When DNS was first implemented, TXT records existed to hold generic text or comments about the domain records (Aitchison, 2005). Now however DNS TXT records have a number of uses, among them being used for DNS-Based Service Discovery (Cheshire and Krochmal, 2013). Further analysis into TXT record utilization could create a better understanding of current DNS practice.
- Further work should also be considered in the vein of creating a locally contextualized understanding about our end-host and network interaction with the Internet as a whole. Barnett and Ehlers (2012) and this thesis are but small steps in the direction of a more unified understanding of local and international network interaction. Further study could take a

number of paths, but the author suggests evaluating the presence of local traffic captured by darknets. This would serve to give insight into the global packet presence generated by local infected or misconfigured hosts, but will also allow a comparison between data volumes and packet activity seen from local hosts and international hosts.

- There is also scope for future work in the area of Bitflipping analysis. One interesting study would be to attempt to correlate bitflip frequency with regional temperatures from the geographic location of the observed IP block over an extended period. Another avenue of analysis would be to study the domains that return responses to bitflipped queries, specifically checking whether or not the domain was registered before or after the Dinaburg (2011) paper.

References

Agten, P., Joosen, W., Piessens, F., and Nikiforakis, N. *Seven months' worth of mistakes: A longitudinal study of typosquatting abuse*. In *Proceedings of the 22nd Network and Distributed System Security Symposium (NDSS 2015)*. 2015.

Aitchison, R. *Pro DNS and BIND*. Apress, August 2005. ISBN13: 978-1-59059-494-0.

Anagnostopoulos, M., Kambourakis, G., Kopanos, P., Louloudakis, G., and Gritzalis, S. *DNS amplification attack revisited. Computers & Security*, 39:475–485, 2013.

Andrews, M. *Negative caching of DNS queries (DNS NCACHE)*. RFC 2308 (Proposed Standard), March 1998. Updated by RFCs 4035, 4033, 4034, 6604.
URL `http://www.ietf.org/rfc/rfc2308.txt`

Barnett, R. J. and Ehlers, K. *An Exploratory study into the location and routing of the most popular websites in south africa*. 2012. Accessed on 28/11/2015.
URL http://www.iadisportal.org/digital-library/an-exploratory-study-into-the-location-and-routing-of-the-most-popular-websites-in-south-africa.

Barr, D. *Common DNS Operational and Configuration Errors*. RFC 1912 (Informational), February 1996.
URL `http://www.ietf.org/rfc/rfc1912.txt`

Brutlag, J. *Speed matters for Google web search. Google*, June 2009. Accessed on: 15/10/2015.
URL http://services.google.com/fh/files/blogs/google_delayexp.pdf.

Büscher, A. and Holz, T. *Tracking DDoS attacks: Insights into the business of disrupting the web*. In *Proceedings of the 5th USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET), San Jose, CA, USA*. 2012. Accessed on 01/12/2015.
URL https://www.usenix.org/system/files/conference/leet12/leet12-final26.pdf.

Chapin, L. and McFadden, M. *Reserved top level domain names*. Internet Draft, May 2011.
URL `https://tools.ietf.org/html/draft-chapin-rfc2606bis-00`

Chen, Y., Antonakakis, M., Perdisci, R., Nadji, Y., Dagon, D., and Lee, W. *DNS noise: Measuring the pervasiveness of disposable domains in modern DNS traffic*. In *Dependable Systems and Networks (DSN), 2014 44th Annual IEEE/IFIP International Conference on*, pages 598–609. IEEE, 2014.

Cheshire, S. and Krochmal, M. *DNS-Based Service Discovery*. RFC 6763 (Proposed Standard), February 2013.
URL `http://www.ietf.org/rfc/rfc6763.txt`

Choi, H., Lee, H., Lee, H., and Kim, H. *Botnet detection by monitoring group activities in dns traffic*. In *Computer and Information Technology, 2007. CIT 2007. 7th IEEE International Conference on*, pages 715–720. IEEE, 2007.

Dagon, D., Antonakakis, M., Vixie, P., Jinmei, T., and Lee, W. *Increased DNS forgery resistance through 0x20-bit encoding: security via leet queries*. In *Proceedings of the 15th ACM conference on Computer and communications security*, pages 211–222. ACM, 2008.

Danzig, P. B., Obraczka, K., and Kumar, A. *An analysis of wide-area name server traffic: a study of the Internet Domain Name System*. *ACM SIGCOMM Computer Communication Review*, 22(4):281–292, 1992.

Dinaburg, A. *Bitsquatting: DNS Hijacking without exploitation*. *Proceedings of BlackHat Security*, 2011. Accessed on 30/11/2015.
URL http://media.blackhat.com/bh-us-11/Dinaburg/BH_US_11_Dinaburg_Bitsquatting_WP.pdf.

Eastlake, D. *Domain Name System (DNS) Case Insensitivity Clarification*. RFC 4343 (Proposed Standard), January 2006.
URL `http://www.ietf.org/rfc/rfc4343.txt`

Eastlake, D. *Domain Name System (DNS) IANA Considerations*. RFC 6895 (Best Current Practice), April 2013.
URL `http://www.ietf.org/rfc/rfc6895.txt`

Eastlake, D. and Panitz, A. *Reserved Top Level DNS Names*. RFC 2606 (Best Current Practice), June 1999. Updated by RFC 6761.
URL `http://www.ietf.org/rfc/rfc2606.txt`

Elz, R. and Bush, R. *Clarifications to the DNS Specification*. RFC 2181 (Proposed Standard), July 1997. Updated by RFCs 4035, 2535, 4343, 4033, 4034, 5452.
URL `http://www.ietf.org/rfc/rfc2181.txt`

Ennis, J. *Pure Python GeoIP API*. June 2015. Accessed on: 07/09/2015. URL https://pypi.python.org/pypi/pygeoip/

Fachkha, C., Bou-Harb, E., and Debbabi, M. *Fingerprinting internet dns amplification DDoS activities*. In *New Technologies, Mobility and Security (NTMS), 2014 6th International Conference on*, pages 1–5. IEEE, 2014.

Fontanini, M. *libtins - packet crafting and sniffing library*. March 2015. Accessed on: 14/04/2015. URL http://libtins.github.io

Gao, H., Yegneswaran, V., Chen, Y., Porras, P., Ghosh, S., Jiang, J., and Duan, H. *An empirical reexamination of global DNS behavior*. In *ACM SIGCOMM Computer Communication Review*, volume 43, pages 267–278. ACM, 2013.

Gulbrandsen, A., Vixie, P., and Esibov, L. *A DNS RR for specifying the location of services (DNS SRV)*. RFC 2782 (Proposed Standard), February 2000. Updated by RFC 6335. URL http://www.ietf.org/rfc/rfc2782.txt

Hermanowski, D. *Open source security information management system supporting it security audit*. In *Cybernetics (CYBCONF), 2015 IEEE 2nd International Conference on*, pages 336–341. IEEE, 2015.

Holgers, T., Watson, D. E., and Gribble, S. D. *Cutting through the confusion: A measurement study of homograph attacks*. In *USENIX Annual Technical Conference, General Track*, pages 261–266. 2006.

Holz, T., Gorecki, C., Rieck, K., and Freiling, F. C. *Measuring and detecting fast-flux service networks*. In *Symposium on Network and Distributed System Security (NDSS)*. 2008.

Irwin, B. and Pilkington, N. *High level internet scale traffic visualization using Hilbert Curve mapping*. In *VizSEC 2007*, pages 147–158. Springer, 2008.

Jin, C., Wang, H., and Shin, K. G. *Hop-count filtering: an effective defense against spoofed DDoS traffic*. In *Proceedings of the 10th ACM conference on Computer and communications security*, pages 30–41. ACM, 2003.

Jung, J. and Sit, E. *An empirical study of spam traffic and the use of DNS black lists*. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 370–375. ACM, 2004.

Jung, J., Sit, E., Balakrishnan, H., and Morris, R. *DNS Performance and the Effectiveness of Caching*. *Networking, IEEE/ACM Transactions on*, 10(5):589–603, 2002.

Kambourakis, G., Moschos, T., Geneiatakis, D., and Gritzalis, S. *Detecting DNS amplification attacks.* In *Critical Information Infrastructures Security*, pages 185–196. Springer, 2008.

Krishnamurthy, B., Wills, C., and Zhang, Y. *On the use and performance of content distribution networks.* In *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*, pages 169–182. ACM, 2001.

Kumar, A., Postel, J., Neuman, C., Danzig, P., and Miller, S. *Common DNS Implementation Errors and Suggested Fixes.* RFC 1536 (Informational), October 1993.
URL `http://www.ietf.org/rfc/rfc1536.txt`

Kwon, J., Kim, J., Lee, J., Lee, H., and Perrig, A. *Psybog: Power spectral density analysis for detecting botnet groups.* In *Malicious and Unwanted Software: The Americas (MALWARE), 2014 9th International Conference on*, pages 85–92. IEEE, 2014.

Larson, M. and Barber, P. *Observed DNS Resolution Misbehavior.* RFC 4697 (Best Current Practice), October 2006.
URL `http://www.ietf.org/rfc/rfc4697.txt`

Ling, Z., Luo, J., Wu, K., Yu, W., and Fu, X. *Torward: Discovery of malicious traffic over tor.* In *INFOCOM, 2014 Proceedings IEEE*, pages 1402–1410. IEEE, 2014.

Lottor, M. *Domain Administrators Operations Guide.* RFC 1033, November 1987.
URL `http://www.ietf.org/rfc/rfc1033.txt`

Mauch, J. *Open resolver project.* In *Presentation, DNS-OARC Spring 2013 Workshop (Dublin)*. 2013. Accessed on 24/05/2014.
URL https://indico.dns-oarc.net/event/0/session/0/contribution/24/material/slides/1.pdf.

Metcalf, L. and Spring, J. *Domain parking: Not as malicious as expected.* Technical report, Pittsburgh PA Software Engineering Institute, 2014. (No. CERTCC-2014-57). Carnegie-Mellon University.

Miszalska, I., Zabierowski, W., and Napieralski, A. *Selected methods of spam filtering in email.* In *CAD Systems in Microelectronics, 2007. CADSM'07. 9th International Conference-The Experience of Designing and Applications of*, pages 507–513. IEEE, 2007.

Mockapetris, P. *Domain names - concepts and facilities.* RFC 1034 (INTERNET STANDARD), November 1987a. Updated by RFCs 1101, 1183, 1348, 1876, 1982, 2065, 2181, 2308, 2535, 4033, 4034, 4035, 4343, 4035, 4592, 5936.
URL `http://www.ietf.org/rfc/rfc1034.txt`

Mockapetris, P. *Domain names - implementation and specification.* RFC 1035 (INTERNET STANDARD), November 1987b. Updated by RFCs 1101, 1183, 1348, 1876, 1982, 1995, 1996, 2065, 2136, 2181, 2137, 2308, 2535, 2673, 2845, 3425, 3658, 4033, 4034, 4035, 4343, 5936, 5966, 6604.
URL `http://www.ietf.org/rfc/rfc1035.txt`

Moore, T. and Edelman, B. *Measuring the perpetrators and funders of typosquatting.* In *Financial Cryptography and Data Security*, pages 175–191. Springer, 2010.

Nazario, J. and Holz, T. *As the net churns: Fast-flux botnet observations.* In *Malicious and Unwanted Software, 2008. MALWARE 2008. 3rd International Conference on*, pages 24–31. IEEE, 2008.

Nikiforakis, N., Balduzzi, M., Desmet, L., Piessens, F., and Joosen, W. *Soundsquatting: Uncovering the use of homophones in domain squatting.* In *Information Security*, volume 8783 of *Lecture Notes in Computer Science*, pages 291–308. Springer, 2014.

Nikiforakis, N., Van Acker, S., Meert, W., Desmet, L., Piessens, F., and Joosen, W. *Bitsquatting: Exploiting bit-flips for fun, or profit?* In *Proceedings of the 22nd international conference on World Wide Web*, pages 989–998. International World Wide Web Conferences Steering Committee, 2013.

Oberheide, J., Karir, M., and Mao, Z. M. *Characterizing dark dns behavior.* In *Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 140–156. Springer, 2007.

Olzak, T. *Dns cache poisoning: Definition and prevention.* 2006. Accessed on: 13/09/2015.
URL http://www.infosecwriters.com.

Padmanabhan, V. N. and Mogul, J. C. *Using predictive prefetching to improve world wide web latency. ACM SIGCOMM Computer Communication Review*, 26(3):22–36, 1996.

Pashalidis, A. *A cautionary note on automatic proxy configuration.* In *IASTED International Conference on Communication, Network, and Information Security, CNIS 2003, New York, USA, December 10-12, 2003, Proceedings*, pages 153–158. ACTA Press, 2003.

Paxson, V. *An analysis of using reflectors for distributed denial-of-service attacks. ACM SIGCOMM Computer Communication Review*, 31(3):38–47, 2001.

Prechelt, L. *An empirical comparison of C, C++, Java, Perl, Python, Rexx and Tcl. IEEE Computer*, 33(10):23–29, 2000.

Ramsay, J., Barbesi, A., and Preece, J. *A psychological investigation of long retrieval times on the world wide web. Interacting with computers*, 10(1):77–86, 1998.

Rekhter, Y., Moskowitz, B., Karrenberg, D., de Groot, G. J., and Lear, E. *Address Allocation for Private Internets*. RFC 1918 (Best Current Practice), February 1996. Updated by RFC 6761.
URL `http://www.ietf.org/rfc/rfc1918.txt`

Rossow, C. *Amplification hell: Revisiting network protocols for DDoS abuse*. In *Symposium on Network and Distributed System Security (NDSS)*. 2014.

Sanner, M. F. *Python: a programming language for software integration and development*. *Journal of Molecular Graphics and Modelling*, 17(1):57–61, 1999.

Sarat, S., Pappas, V., and Terzis, A. *On the use of anycast in DNS*. In *Computer Communications and Networks, 2006. ICCCN 2006. Proceedings. 15th International Conference on*, pages 71–78. IEEE, 2006.

Schonewille, A. and van Helmond, D.-J. *The domain name service as an IDS*. *Research Project for the Master System-and Network Engineering at the University of Amsterdam*, 2006.

Singla, A., Chandrasekaran, B., Godfrey, P., and Maggs, B. *The internet at the speed of light*. In *Proceedings of the 13th ACM Workshop on Hot Topics in Networks*, page 1. ACM, 2014.

Smith, N. *Security Threats Against Secure Sockets Layer (SSL)*. Ph.D. thesis, Nova Southeastern University, 2010.

Souders, S. *Velocity and the bottom line*. Online, July 2009. Accessed on 14/08/2015.
URL http://radar.oreilly.com/2009/07/velocity-making-your-site-fast.html.

Stalmans, E. and Irwin, B. *A framework for dns based detection and mitigation of malware infections on a network*. In *Information Security South Africa (ISSA), 2011*, pages 1–8. IEEE, 2011.

Sullivan, D. T. *Survey of malware threats and recommendations to improve cybersecurity for industrial control systems version 1.0*. Technical report, Raytheon Technical Services Company LLC, 2015. Accessed on 15/10/2015.
URL http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA617910.

Teo, L. *Port scans and ping sweeps explained*. *Linux Journal*, 2000(80es):2, 2000.

The Measurement Factory. *ipv4-heatmap: Mapping the ipv4 address space*. March 2015. Accessed on: 12/06/2015.
URL `http://maps.measurement-factory.com/`

van Zyl, I., Rudman, L. L., and Irwin, B. *A review of current DNS TTL practices*. In **Otten, D. F. and Balmahoon, M. R.**, editors, *Southern Africa Telecommunication Networks and Applications Conference (SATNAC) 2015*, pages 131 –136. August 2015.

Vixie, P. *Extension Mechanisms for DNS (EDNS0)*. RFC 2671 (Proposed Standard), August 1999. Obsoleted by RFC 6891.
URL http://www.ietf.org/rfc/rfc2671.txt

Vixie, P. and Dagon, D. *Use of Bit 0x20 in DNS Labels to Improve Transaction Identity*. Internet Draft, March 2008. Accessed on: 02/12/2015.
URL https://tools.ietf.org/html/draft-chapin-rfc2606bis-00

Von Arx, K. G. and Hagen, G. R. *Sovereign domains: A declaration of independence of cctlds from foreign control*. *Rich. JL & Tech.*, 9:4–8, 2002.

Vulimiri, A., Michel, O., Godfrey, P., and Shenker, S. *More is less: reducing latency via redundancy*. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, pages 13–18. ACM, 2012.

Wessels, D. *Observations on Checksum Errors in DNS Queries to VeriSign's Name Servers*. *Verisign Incorporated*, February 2012. Accessed on 17/11/2015.
URL http://www.verisign.com/assets/VRSN_Bitsquatting_TR_20120320.pdf?inc=www.verisigninc.com.

Williamson, C. *Internet traffic measurement*. *Internet Computing, IEEE*, 5(6):70–74, 2001.

Wills, C. and Shang, H. *The contribution of DNS lookup costs to web object retrieval*. Technical report, Worcester Polytechnic Institute, 2000. Technical Report: TR-00-12.

Yadav, S., Reddy, A. K. K., and Ranjan, S. *Detecting algorithmically generated domain-flux attacks with DNS traffic analysis*. *Networking, IEEE/ACM Transactions on*, 20(5):1663–1677, 2012.

Yadav, S. and Reddy, A. N. *Winning with DNS failures: Strategies for faster botnet detection*. In *Security and privacy in communication networks*, pages 446–459. Springer, 2012.

Zdrnja, B. *Security Monitoring of DNS traffic*. *University of Auckland*, 2006. Accessed 13/07/2015.
URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.511.7926&rep=rep1&type=pdf.

Zdrnja, B., Brownlee, N., and Wessels, D. *Passive monitoring of DNS anomalies*. In *Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 129–139. Springer, 2007.

Appendices

Table A.1: Dataset TTL frequency

Rank	1		2		3		4		5		6		7		8		9		10	
Month	TTL	%ofTotal	TTL	%ofTotal	TTL	%ofTotal	TTL	%ofTotal	TTL	%ofTotal	TTL	%ofTotal	TTL	%ofTotal	TTL	%ofTotal	TTL	%ofTotal	TTL	%ofTotal
October 2013	300	24.939	60	14.203	20	12.122	3600	10.445	600	6.445	30	5.402	86400	4.077	7200	3.666	900	3.665	21600	1.990
November 2013	300	26.222	60	14.959	20	11.270	3600	10.554	600	6.655	30	4.964	86400	3.772	7200	3.532	900	3.390	1800	1.845
December 2013	300	20.270	3600	15.068	86400	14.038	600	13.419	60	6.266	7200	5.575	900	3.587	20	2.626	1800	1.946	43200	1.933
January 2014	300	27.648	60	13.172	3600	10.784	600	8.511	20	7.951	86400	5.235	7200	3.894	30	3.594	900	3.406	1800	2.177
February 2014	300	22.943	60	13.477	3600	11.257	20	9.123	86400	7.842	600	6.442	30	4.343	900	4.014	7200	3.738	1800	2.205
March 2014	300	27.499	60	14.749	20	10.019	3600	9.603	600	6.236	30	4.622	86400	4.198	900	3.934	1800	2.151	3200	1.934
June 2014*	300	25.829	60	16.395	20	12.171	3600	9.636	30	5.769	600	5.609	900	3.383	86400	2.816	120	2.331	1800	2.255
July 2014	300	28.091	60	17.232	20	12.806	3600	9.712	30	5.708	600	5.464	900	2.954	86400	2.492	1800	2.145	120	2.092
August 2014	300	30.126	60	14.902	20	11.002	3600	9.744	30	6.358	600	6.111	900	3.271	86400	2.659	1800	2.354	120	1.836
September 2014	300	28.369	60	15.958	20	11.082	3600	8.759	1	6.336	30	5.637	600	4.996	120	2.585	900	2.497	86400	2.172
October 2014	300	31.117	60	16.858	20	10.979	3600	10.173	30	5.312	600	5.247	900	2.839	86400	2.516	120	2.435	1800	1.983
November 2014	300	32.315	60	17.189	20	10.907	3600	9.805	30	5.320	600	5.086	120	3.089	900	2.249	86400	2.254	1800	1.921
December 2014	300	32.811	60	12.890	3600	11.050	20	10.725	600	7.576	30	4.113	86400	3.333	900	3.300	1800	1.648	21600	1.443
January 2015	300	31.356	60	18.741	20	10.889	3600	9.758	600	5.639	30	4.191	900	2.813	86400	2.579	120	1.906	1800	1.549
February 2015	300	32.161	60	18.221	20	10.509	3600	9.657	600	5.365	30	4.574	120	2.773	900	2.458	86400	2.284	1800	1.589
March 2015	300	32.497	60	17.769	20	10.635	3600	9.829	600	5.333	30	4.898	120	3.483	900	2.466	86400	2.216	1800	1.541
April 2015	300	31.691	60	18.040	3600	10.085	20	9.693	600	7.092	30	5.016	900	2.750	86400	2.680	21600	1.537	1800	1.495
May 2015	300	28.438	60	19.665	20	13.689	3600	8.666	30	5.710	600	5.698	120	4.102	900	2.244	86400	1.898	21600	1.395
June 2015	300	27.303	60	18.331	20	12.393	600	8.832	3600	8.675	30	5.230	120	3.226	86400	2.697	900	2.336	1800	1.337
July 2015	300	28.855	60	16.884	20	11.682	3600	9.743	600	6.218	30	4.494	86400	4.109	900	2.632	120	1.661	21600	1.333
August 2015	300	26.346	60	19.468	20	12.470	3600	9.891	600	6.579	30	5.116	86400	3.543	900	3.113	120	1.485	21600	1.463

Table A.2: Top 5 geolocation distribution for .co.za domains

Rank	1		2		3		4		5	
Month	Country	% of IPs	Country	% of IPs	Country	% of IPs	Country	% of IPs	Country	% of IPs
October 2013	US	45.060	ZA	33.044	UK	7.275	DE	5.248	NL	1.376
November 2013	US	45.072	ZA	33.444	UK	6.718	DE	5.574	NL	1.440
December 2013	ZA	39.516	US	39.032	UK	6.989	DE	5.753	CA	1.344
January 2014	US	44.780	ZA	33.442	UK	6.811	DE	6.117	NL	1.468
February 2014	US	44.762	ZA	33.233	UK	6.797	DE	5.821	NL	1.427
March 2014	US	43.665	ZA	34.374	UK	6.610	DE	5.839	NL	1.542
June 2014*	US	43.785	ZA	35.068	UK	6.336	DE	5.851	NL	1.493
July 2014	US	44.524	ZA	34.246	UK	6.039	DE	5.514	CA	1.463
August 2014	US	42.443	ZA	35.502	UK	6.585	DE	5.752	NL	1.428
September 2014	US	45.505	ZA	32.753	UK	6.551	DE	5.401	NL	1.463
October 2014	US	45.344	ZA	33.103	UK	6.844	DE	5.166	NL	1.645
November 2014	US	46.437	ZA	32.833	UK	5.898	DE	4.876	NL	1.568
December 2014	US	42.143	ZA	36.964	UK	5.580	DE	5.045	NL	1.473
January 2015	US	45.535	ZA	34.902	UK	5.854	DE	5.446	NL	1.519
February 2015	US	46.818	ZA	32.050	UK	6.497	DE	4.930	NL	1.608
March 2015	US	46.454	ZA	33.176	UK	6.353	DE	4.975	NL	1.513
April 2015	US	46.480	ZA	34.870	UK	5.517	DE	4.323	NL	1.317
May 2015	US	44.212	ZA	35.463	UK	5.885	DE	5.304	NL	1.510
June 2015	US	44.790	ZA	34.681	UK	5.910	DE	5.521	FR	1.555
July 2015	US	44.876	ZA	34.356	UK	5.745	DE	4.775	NL	1.980
August 2015	US	44.598	ZA	35.804	UK	5.318	DE	4.941	NL	1.508

Table A.3: Observed TTL and RRs for .co.za domains

Rank	1		2		3		4		5		6		7		8		9		10	
Month	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs
October 2013	7200	22.954	14400	15.845	86400	14.447	3600	13.736	600	11.070	300	5.600	7260	2.606	60	1.524	10800	1.492	38400	1.486
November 2013	7200	24.152	14400	14.684	86400	13.800	3600	13.435	600	11.501	300	6.184	7260	2.730	10800	1.647	60	1.570	43200	1.299
December 2013	7200	20.568	86400	15.389	3600	14.519	14400	13.515	600	12.003	300	7.622	7260	2.796	60	1.771	1800	1.553	38400	1.274
January 2014	7200	23.876	3600	13.797	14400	13.626	86400	13.415	600	11.661	300	6.618	7260	2.815	1800	2.156	60	1.800	43200	1.226
February 2014	7200	23.997	86400	13.576	14400	13.441	3600	13.311	600	11.696	300	6.509	7260	3.009	1800	2.664	60	1.456	10800	1.231
March 2014	7200	24.193	86400	14.735	3600	13.313	14400	12.952	600	11.547	300	6.226	1800	2.938	7260	2.783	60	1.389	38400	1.211
June 2014*	86400	24.633	7200	21.399	3600	12.142	14400	11.158	600	10.410	300	5.658	7260	2.503	1800	2.503	60	1.119	10800	1.012
July 2014	7200	24.073	86400	13.678	3600	13.631	14400	12.144	600	11.637	300	6.723	1800	3.505	7260	3.224	10800	1.349	60	1.218
August 2014	7200	22.003	86400	17.600	3600	13.097	600	11.268	14400	11.221	300	6.426	1800	3.893	7260	2.896	28800	1.817	60	1.302
September 2014	7200	24.131	86400	15.289	3600	12.801	14400	11.793	600	11.441	300	6.133	1800	3.356	7260	2.825	28800	2.531	10800	1.205
October 2014	7200	22.094	3600	13.734	86400	13.681	14400	13.265	600	11.620	300	5.905	28800	4.450	1800	3.238	7260	2.481	60	1.517
November 2014	7200	23.239	3600	14.371	14400	12.821	600	12.535	86400	12.443	300	6.393	1800	3.415	7260	2.727	28800	2.402	60	1.443
December 2014	7200	21.241	3600	15.090	600	14.307	14400	12.421	86400	10.557	300	7.731	1800	3.713	7260	2.900	28800	1.655	60	1.610
January 2015	7200	23.717	3600	14.117	600	13.451	14400	13.201	86400	10.565	300	6.925	1800	3.446	7260	3.002	28800	1.870	60	1.526
February 2015	7200	24.532	3600	14.089	14400	13.197	600	12.739	86400	10.388	300	6.848	1800	3.695	7260	3.187	28800	1.763	60	1.564
March 2015	7200	24.423	14400	14.201	3600	13.861	600	12.942	86400	10.088	300	6.492	1800	3.524	7260	3.059	28800	2.011	60	1.599
April 2015	7200	23.884	3600	14.272	600	13.363	14400	12.867	86400	9.809	300	7.051	1800	3.561	7260	3.097	28800	1.987	60	1.712
May 2015	7200	26.074	3600	13.988	600	13.723	14400	11.748	86400	9.267	300	6.723	1800	3.322	7260	3.226	28800	2.059	60	1.4945
June 2015	7200	25.619	600	15.625	3600	13.273	14400	11.803	86400	9.118	300	6.631	1800	3.256	7260	2.702	28800	2.114	60	1.566
July 2015	7200	25.150	600	14.771	3600	13.036	14400	11.889	86400	9.578	300	7.325	1800	3.792	7260	2.628	28800	2.121	60	1.516
August 2015	7200	25.036	600	13.442	3600	12.892	14400	11.854	86400	9.334	300	7.329	1800	4.133	7260	3.329	28800	2.759	60	1.595

Rank	1		2		3		4		5		6		7		8		9		10	
Month	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs
Oct13	A	75.522	MX	13.397	CNAME	6.486	TXT	4.420	PTR	0.066	SOA	0.060	AAAA	0.027	NS	0.016	ANY	0.005	N/A	
Nov13	A	75.345	MX	14.005	CNAME	6.245	TXT	4.361	AAAA	0.028	SOA	0.011	NS	0.006	N/A					
Dec13	A	67.295	MX	21.189	TXT	5.965	CNAME	5.478	SOA	0.031	AAAA	0.031	NS	0.010	N/A					
Jan14	A	74.779	MX	14.403	CNAME	6.210	TXT	4.529	AAAA	0.040	SOA	0.020	PTR	0.013	NS	0.007	N/A			
Feb14	A	75.653	MX	13.565	CNAME	6.514	TXT	4.177	AAAA	0.034	NS	0.023	SOA	0.017	PTR	0.011	SRV	0.006	N/A	
Mar14	A	77.074	MX	12.835	CNAME	6.048	TXT	3.990	AAAA	0.044	SOA	0.017	NS	0.017	PTR	0.006	N/A			
Jun14	A	79.413	MX	11.792	CNAME	5.112	TXT	3.588	SOA	0.028	NS	0.028	AAAA	0.028	SRV	0.011	N/A			
Jul14	A	72.422	MX	18.055	CNAME	5.182	TXT	4.269	AAAA	0.036	NS	0.024	SRV	0.006	SOA	0.006	N/A			
Aug14	A	75.148	MX	14.416	CNAME	6.039	TXT	4.297	AAAA	0.047	NS	0.029	SOA	0.012	PTR	0.012	N/A			
Sep14	A	77.402	MX	11.857	CNAME	7.073	TXT	3.572	AAAA	0.058	SOA	0.019	NS	0.014	PTR	0.005	N/A			
Oct14	A	76.734	CNAME	9.864	MX	9.820	TXT	3.455	AAAA	0.102	NS	0.018	SOA	0.009	N/A					
Nov14	A	75.168	MX	10.772	CNAME	10.244	TXT	3.671	AAAA	0.111	SOA	0.010	PTR	0.010	NS	0.010	SRV	0.005	N/A	
Dec14	A	69.134	MX	19.444	CNAME	6.598	TXT	4.682	AAAA	0.097	NS	0.030	SOA	0.015	N/A					
Jan15	A	74.030	MX	12.191	CNAME	9.389	TXT	4.251	AAAA	0.083	SOA	0.017	NS	0.017	SRV	0.011	PTR	0.011	N/A	
Feb15	A	73.702	MX	11.852	CNAME	10.308	TXT	3.979	AAAA	0.120	SRV	0.015	SOA	0.015	NS	0.010	N/A			
Mar15	A	73.490	MX	11.374	CNAME	10.685	TXT	4.270	AAAA	0.144	SRV	0.014	SOA	0.014	NS	0.010	N/A			
Apr15	A	73.2161	MX	13.218	CNAME	8.259	TXT	5.149	AAAA	0.098	SOA	0.0261	NS	0.0261	SRV	0.007	N/A			
May15	A	77.379	MX	12.487	CNAME	5.555	TXT	4.484	AAAA	0.045	SOA	0.023	SRV	0.011	NS	0.011	PTR	0.006	N/A	
Jun15	A	77.236	MX	12.510	CNAME	5.478	TXT	4.641	NS	0.040	PTR	0.034	AAAA	0.034	SOA	0.017	SRV	0.011	N/A	
Jul15	A	77.841	MX	11.952	CNAME	5.481	TXT	4.627	PTR	0.029	AAAA	0.029	NS	0.017	SRV	0.012	SOA	0.012	N/A	
Aug15	A	75.862	MX	13.556	CNAME	5.392	TXT	5.038	PTR	0.057	AAAA	0.038	SOA	0.025	SRV	0.019	NS	0.013	N/A	

Table A.4: Top 5 geolocation distribution for .org.za domains

Rank	1		2		3		4		5	
Month	Country	% of IPs	Country	% of IPs	Country	% of IPs	Country	% of IPs	Country	% of IPs
October 2013	ZA	44.886	US	34.943	UK	6.818	DE	4.545	CA	3.409
November 2013	ZA	44.648	US	37.003	UK	6.116	DE	3.976	CA	2.446
December 2013	ZA	48.428	US	28.931	UK	8.805	DE	6.918	CA	3.774
January 2014	ZA	46.964	US	31.579	UK	7.287	DE	4.858	CA	3.239
February 2014	ZA	47.727	US	31.169	UK	7.468	DE	5.844	CA	3.247
March 2014	ZA	43.567	US	35.380	UK	7.018	DE	4.971	CA	2.632
June 2014*	ZA	45.614	US	36.842	UK	5.614	DE	4.211	CA	2.105
July 2014	ZA	43.791	US	35.621	DE	5.882	UK	4.575	CA	2.941
August 2014	ZA	42.236	US	37.267	UK	5.590	DE	4.658	CA	4.037
September 2014	ZA	43.223	US	37.084	UK	5.882	DE	5.882	CA	3.069
October 2014	ZA	44.784	US	39.440	UK	4.326	DE	3.817	CA	2.290
November 2014	ZA	44.179	US	37.910	UK	5.671	DE	5.075	CA	2.090
December 2014	ZA	53.333	US	28.571	DE	6.190	UK	5.238	CA	2.857
January 2015	ZA	45.946	US	36.149	DE	5.068	UK	4.730	CA	3.716
February 2015	ZA	47.309	US	35.411	UK	7.082	DE	2.833	CA	2.833
March 2015	ZA	46.392	US	36.856	UK	5.928	DE	3.351	CA	3.093
April 2015	ZA	52.434	US	31.461	DE	5.243	UK	4.120	CA	3.371
May 2015	ZA	49.045	US	32.166	UK	5.732	CA	4.459	DE	3.503
June 2015	ZA	51.118	US	29.393	UK	6.709	CA	4.473	DE	3.834
July 2015	ZA	50.316	US	31.329	UK	6.646	DE	4.430	CA	2.532
August 2015	ZA	51.957	US	28.114	UK	8.185	DE	4.982	CA	3.203

Table A.5: Observed TTL and RRs for .org.za domains

Rank	1		2		3		4		5		6		7		8		9		10	
Month	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs
October 2013	7200	38.916	86400	12.682	3600	11.713	14400	10.939	600	7.841	300	6.196	10800	2.420	7260	2.227	60	0.968	21600	0.968
November 2013	7200	34.829	14400	14.050	86400	12.161	3600	11.806	600	8.501	300	6.730	10800	2.597	7260	2.243	43200	1.653	28800	1.181
December 2013	7200	46.835	3600	11.603	86400	10.127	14400	7.806	300	7.384	600	5.696	7260	2.743	10800	2.2.611743	43200	1.055	1200	0.844
January 2014	7200	42.298	3600	10.574	86400	10.183	14400	9.138	600	7.702	300	6.658	1800	2.872	7260	2.611	10800	1.567	43200	1.044
February 2014	7200	53.086	86400	8.401	3600	8.178	600	6.766	14400	6.543	300	5.576	1800	3.048	7260	1.710	10800	1.264	43200	0.595
March 2014	7200	52.250	3600	9.579	86400	8.999	14400	6.967	600	5.806	300	4.717	1800	4.209	7260	1.451	10800	1.016	1200	0.798
June 2014*	7200	31.447	3600	12.704	86400	10.692	600	10.189	14400	10.063	300	9.937	1800	5.031	7260	2.138	10800	2.138	1200	1.006
July 2014	7200	29.348	3600	11.685	600	11.413	86400	11.277	300	9.375	14400	9.375	1800	5.707	7260	2.038	10800	2.038	43200	1.223
August 2014	7200	32.064	3600	11.916	14400	11.425	86400	9.828	600	8.845	300	8.477	1800	6.511	7260	2.088	10800	1.966	43200	1.106
September 2014	7200	32.059	600	11.373	14400	11.373	3600	10.686	86400	9.804	300	7.353	1800	4.020	7260	2.647	1200	1.667	10800	1.569
October 2014	7200	33.618	14400	12.821	3600	11.491	600	11.396	86400	9.212	300	7.787	1800	2.944	10800	2.659	10800	1.330	1200	1.235
November 2014	7200	32.589	600	11.830	3600	11.719	14400	11.161	86400	9.263	300	8.371	1800	2.344	7260	1.897	43200	1.897	10800	1.563
December 2014	7200	33.095	600	11.807	3600	11.449	14400	9.302	86400	8.945	300	8.408	7260	2.2.372862	1800	2.683	10800	2.147	43200	1.968
January 2015	7200	36.205	14400	10.861	3600	10.737	600	10.487	86400	9.738	300	8.365	7260	2.372	1800	2.247	10800	1.623	43200	1.124
February 2015	7200	34.353	14400	11.927	3600	11.519	600	10.907	86400	8.665	300	8.053	1800	3.976	7260	1.733	43200	1.223	10800	1.019
March 2015	7200	33.301	600	11.900	14400	11.324	3600	10.940	86400	9.309	300	7.869	7260	2.399	1800	2.303	1200	1.536	43200	1.344
April 2015	7200	30.857	3600	12.143	600	11.571	14400	11.571	300	10.142	86400	8.143	1800	2.857	7260	2.714	60	1.286	43200	1.286
May 2015	7200	33.209	3600	12.515	600	12.268	14400	11.152	300	8.426	86400	8.055	7260	2.230	1800	2.230	10800	1.611	60	0.991
June 2015	7200	30.194	600	12.903	3600	12.000	14400	10.839	300	10.326	86400	8.000	1200	2.323	7260	2.065	1800	2.065	10800	1.677
July 2015	7200	29.699	600	12.657	3600	12.531	14400	12.030	300	9.398	86400	7.393	1800	2.757	7260	2.005	1200	1.880	10800	1.629
August 2015	7200	30.381	3600	11.444	600	10.218	14400	9.946	300	9.537	86400	8.038	1800	5.450	7260	2.997	1200	2.316	43200	1.907

Rank	1		2		3		4		5		6		7		8		9		10	
Month	RR.	% of RRs	RR.	% of RRs	RR.	% of RRs	RR.	% of RRs	RR.	% of RRs	RR.	% of RRs	RR.	% of RRs	RR.	% of RRs	RR.	% of RRs	RR.	% of RRs
Oct13	A	81.413	MX	7.938	CNAME	7.067	TXT	3.291	SOA	0.194	NS	0.097	N/A							
Nov13	A	79.221	MX	9.681	CNAME	6.848	TXT	4.132	SOA	0.118	N/A									
Dec13	A	79.114	MX	14.135	TXT	4.641	CNAME	1.899	SOA	0.211	N/A									
Jan14	A	82.637	MX	8.355	TXT	4.439	CNAME	4.439	SOA	0.131	N/A									
Feb14	A	86.022	MX	6.543	CNAME	4.610	TXT	2.602	SOA	0.149	AAAA	0.074	N/A							
Mar14	A	86.865	MX	5.660	CNAME	4.790	TXT	2.612	SOA	0.073	N/A									
Jun14	A	81.132	MX	9.182	CNAME	6.667	TXT	2.893	SOA	0.126	N/A									
Jul14	A	78.804	MX	11.821	CNAME	4.891	TXT	4.212	SOA	0.136	AAAA	0.1136	N/A							
Aug14	A	77.914	MX	10.061	CNAME	8.098	TXT	3.681	SOA	0.123	AAAA	0.123	N/A							
Sep14	A	79.020	MX	9.118	CNAME	8.431	TXT	3.235	SOA	0.0980	AAAA	0.0980	N/A							
Oct14	A	79.582	CNAME	9.402	MX	8.072	TXT	2.754	SOA	0.095	AAAA	0.095	N/A							
Nov14	A	78.125	CNAME	10.156	MX	8.259	TXT	3.125	AAAA	0.335	N/A									
Dec14	A	76.029	MX	14.848	CNAME	4.651	TXT	4.472	N/A											
Jan15	A	79.775	MX	9.738	CNAME	7.491	TXT	2.871	AAAA	0.125	N/A									
Feb15	A	76.962	CNAME	11.009	MX	8.767	TXT	3.262	N/A											
Mar15	A	79.559	CNAME	9.981	MX	7.006	TXT	3.167	AAAA	0.288	N/A									
Apr15	A	79.286	MX	10.143	CNAME	5.429	TXT	5.000	AAAA	0.143	N/A									
May15	A	79.678	MX	9.913	CNAME	5.452	TXT	4.957	N/A											
Jun15	A	81.443	MX	9.278	TXT	4.768	CNAME	4.510	N/A											
Jul15	A	82.206	MX	8.772	TXT	4.762	CNAME	4.261	N/A											
Aug15	A	80.518	MX	10.763	TXT	4.496	CNAME	4.223	N/A											

Table A.6: Top 5 geolocation distribution for .gov.za domains

Rank	1		2		3		4		5	
Month	Country	% of IPs	Country	% of IPs	Country	% of IPs	Country	% of IPs	Country	% of IPs
October 2013	ZA	96.226	US	3.774	-	-	-	-	-	-
November 2013	ZA	94.845	US	3.093	NZ	1.031	DE	1.031	-	-
December 2013	ZA	96.825	US	3.175	-	-	-	-	-	-
January 2014	ZA	94.565	US	4.348	DE	1.087	-	-	-	-
February 2014	ZA	94.898	US	5.102	-	-	-	-	-	-
March 2014	ZA	92.593	US	3.704	UK	0.926	NZ	0.926	DE	0.926
June 2014*	ZA	94.231	US	4.808	NZ	0.962	-	-	-	-
July 2014	ZA	96.000	US	3.000	NZ	1.000	-	-	-	-
August 2014	ZA	94.444	US	4.630	DE	0.926	-	-	-	-
September 2014	ZA	95.833	US	3.333	DE	0.833	-	-	-	-
October 2014	ZA	95.370	US	4.630	-	-	-	-	-	-
November 2014	ZA	95.146	US	4.854	-	-	-	-	-	-
December 2014	ZA	97.260	US	2.740	-	-	-	-	-	-
January 2015	ZA	94.624	US	4.301	DE	1.075	-	-	-	-
February 2015	ZA	93.333	US	5.714	DE	0.952	-	-	-	-
March 2015	ZA	94.898	US	4.082	UK	1.020	-	-	-	-
April 2015	ZA	95.402	US	4.598	-	-	-	-	-	-
May 2015	ZA	94.624	US	5.376	-	-	-	-	-	-
June 2015	ZA	92.771	US	7.229	-	-	-	-	-	-
July 2015	ZA	96.739	US	3.261	-	-	-	-	-	-
August 2015	ZA	96.552	US	3.448	-	-	-	-	-	-

Table A.7: Observed TTL and RRs for .gov.za domains

Rank	1		2		3		4		5		6		7		8		9		10	
Month	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs
October 2013	600	33.334	3600	22.727	86400	17.172	7200	5.556	300	5.051	10800	4.545	60	3.535	38400	2.020	14400	2.020	7260	1.010
November 2013	600	32.418	86400	18.681	3600	18.681	7200	6.044	14400	4.945	300	4.396	10800	4.396	60	3.297	38400	2.198	7260	1.099
December 2013	600	30.275	3600	28.440	86400	14.679	7200	5.505	300	4.587	14400	4.587	10800	4.587	60	2.752	38400	2.752	240	0.917
January 2014	600	26.316	3600	24.561	86400	19.883	7200	6.433	10800	6.433	300	5.848	60	3.509	14400	2.339	38400	1.754	604800	0.585
February 2014	600	26.923	3600	24.519	86400	17.788	10800	7.212	300	5.769	14400	4.327	7200	3.846	38400	1.923	60	1.442	1800	1.442
March 2014	600	33.333	3600	22.857	86400	16.667	300	6.190	7200	4.286	14400	3.333	10800	3.333	60	2.381	38400	2.381	7260	0.952
June 2014*	600	30.928	3600	24.227	86400	18.041	7200	5.155	300	5.155	10800	5.155	60	2.062	38400	2.062	1800	2.062	14400	2.062
July 2014	600	34.211	3600	21.579	86400	18.421	300	5.263	10800	4.211	7200	3.684	60	3.158	1800	3.158	38400	2.105	14400	1.579
August 2014	600	32.105	3600	23.684	86400	13.684	7200	5.789	10800	5.263	60	3.684	14400	3.158	300	2.632	1800	2.632	38400	2.105
September 2014	600	33.061	3600	19.184	86400	16.327	300	6.531	7200	6.122	10800	5.306	14400	3.673	7260	2.041	38400	2.041	60	1.633
October 2014	600	31.197	3600	21.368	86400	16.667	300	5.983	10800	5.983	7200	5.556	60	2.991	38400	2.137	14400	2.137	7260	1.282
November 2014	600	31.429	3600	18.571	86400	15.238	7200	7.143	300	7.143	10800	6.190	60	2.857	14400	2.381	7260	1.429	38400	1.429
December 2014	600	29.861	86400	21.528	3600	17.361	7200	6.944	300	6.250	10800	4.861	14400	3.472	60	2.083	43200	2.083	38400	2.083
January 2015	600	30.811	3600	21.081	86400	16.216	7200	7.027	300	5.946	10800	3.784	38400	3.243	1800	3.243	60	2.162	14400	1.622
February 2015	600	26.500	3600	22.500	86400	17.000	7200	8.000	300	6.000	10800	5.000	14400	3.000	38400	2.500	43200	2.000	1800	2.000
March 2015	600	31.884	3600	21.256	86400	18.357	7200	5.314	300	4.831	14400	4.831	10800	3.865	38400	1.932	7260	1.449	60	1.449
April 2015	600	30.247	3600	17.901	86400	14.198	7200	7.407	43200	7.407	300	6.173	14400	4.321	10800	4.321	38400	3.086	1800	1.852
May 2015	600	33.333	3600	21.637	86400	11.696	10800	6.433	14400	5.848	300	4.678	7200	4.094	43200	4.094	38400	2.339	1800	1.754
June 2015	600	31.169	3600	22.078	86400	14.286	300	7.792	43200	5.195	7200	4.545	14400	4.545	10800	3.247	38400	2.597	60	1.299
July 2015	600	28.070	3600	18.713	86400	11.696	300	7.018	14400	7.018	7200	6.433	43200	5.263	10800	4.678	38400	2.339	1800	2.339
August 2015	600	26.946	3600	23.353	86400	10.778	300	8.982	43200	7.186	14400	5.389	7200	4.192	10800	4.192	60	1.796	38400	1.796

Rank	1		2		3		4		5		6		7		8		9		10	
Month	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs
Oct13	A	80.808	CNAME	14.141	MX	3.030	TXT	1.515	NS	0.505	N/A									
Nov13	A	82.967	CNAME	12.088	MX	3.297	TXT	1.099	NS	0.549	N/A									
Dec13	A	85.321	CNAME	7.339	TXT	3.669	MX	3.669	N/A											
Jan14	A	82.456	CNAME	11.696	TXT	2.924	MX	2.924	N/A											
Feb14	A	80.769	CNAME	12.500	TXT	3.365	MX	3.365	N/A											
Mar14	A	81.905	CNAME	12.857	TXT	2.857	MX	2.381	N/A											
Jun14	A	86.082	CNAME	8.763	TXT	3.093	MX	2.062	N/A											
Jul14	A	83.158	MX	6.842	CNAME	6.842	TXT	3.158	N/A											
Aug14	A	80.000	CNAME	13.684	MX	3.684	TXT	2.632	N/A											
Sep14	A	81.224	CNAME	13.061	TXT	2.857	MX	2.857	N/A											
Oct14	A	81.197	CNAME	13.248	TXT	2.991	MX	2.564	N/A											
Nov14	A	76.190	CNAME	15.238	MX	4.762	TXT	3.810	N/A											
Dec14	A	77.083	CNAME	9.722	MX	8.333	TXT	4.861	N/A											
Jan15	A	79.459	CNAME	10.811	TXT	4.865	MX	4.865	N/A											
Feb15	A	79.000	CNAME	12.000	TXT	4.500	MX	4.500	N/A											
Mar15	A	74.396	CNAME	15.459	TXT	5.314	MX	4.831	N/A											
Apr15	A	75.926	CNAME	12.963	TXT	6.173	MX	4.938	N/A											
May15	A	80.117	CNAME	7.602	TXT	6.433	MX	5.848	N/A											
Jun15	A	81.169	TXT	6.494	CNAME	6.494	MX	5.844	N/A											
Jul15	A	81.287	MX	7.018	TXT	6.433	CNAME	5.263	N/A											
Aug15	A	81.437	TXT	6.587	MX	5.988	CNAME	5.988	N/A											

Table A.8: Top 5 geolocation distribution for .ac.za domains

Rank	1		2		3		4		5	
Month	Country	% of IPs	Country	% of IPs	Country	% of IPs	Country	% of IPs	Country	% of IPs
October 2013	ZA	72.174	US	13.913	DE	4.348	UK	3.478	AU	1.739
November 2013	ZA	77.228	US	13.861	DE	4.950	AU	1.981	UK	0.990
December 2013	ZA	82.432	US	9.459	DE	5.405	UK	2.703	-	-
January 2014	ZA	76.596	US	13.830	DE	4.255	UK	2.128	AU	2.128
February 2014	ZA	75.229	US	15.596	DE	3.670	UK	2.752	AU	1.835
March 2014	ZA	81.308	US	11.215	DE	4.673	UK	1.869	AU	0.935
June 2014*	ZA	78.899	US	12.844	DE	4.587	AU	1.835	UK	0.917
July 2014	ZA	81.731	US	9.615	DE	3.846	UK	1.923	AU	1.923
August 2014	ZA	79.824	US	12.281	DE	3.509	UK	2.632	MU	0.877
September 2014	ZA	78.814	US	13.559	DE	3.390	UK	2.542	AU	1.695
October 2014	ZA	82.300	US	10.620	DE	3.540	UK	2.655	AU	0.885
November 2014	ZA	73.913	US	14.783	UK	4.348	DE	4.348	AU	1.739
December 2014	ZA	76.000	US	16.000	DE	5.333	UK	2.667	-	-
January 2015	ZA	75.893	US	15.179	DE	4.464	UK	2.679	AU	1.786
February 2015	ZA	73.984	US	17.073	UK	3.252	DE	3.252	AU	1.626
March 2015	ZA	79.464	US	12.500	DE	3.571	UK	2.679	AU	1.786
April 2015	ZA	79.245	US	16.038	DE	3.774	AU	0.943	-	-
May 2015	ZA	76.190	US	15.238	UK	3.810	DE	3.810	AU	0.952
June 2015	ZA	78.641	US	13.592	DE	3.883	UK	1.942	AU	1.942
July 2015	ZA	71.560	US	19.266	UK	3.670	DE	3.670	AU	1.835
August 2015	ZA	72.222	US	18.519	UK	3.704	DE	3.704	AU	1.852

Table A.9: Observed TTL and RRs for .ac.za domains

Rank	1		2		3		4		5		6		7		8		9		10	
Month	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs
October 2013	86400	47.263	3600	17.518	10800	9.854	900	4.927	7200	2.372	300	2.190	600	1.825	14400	1.825	60	1.460	259200	1.460
November 2013	86400	48.904	3600	14.474	10800	11.404	900	5.482	300	2.412	7200	2.193	600	2.193	259200	1.974	172800	1.535	8600	1.316
December 2013	86400	50.000	3600	14.286	10800	9.375	900	4.465	600	2.679	300	2.679	172800	2.679	7200	2.232	60	2.232	259200	2.232
January 2014	86400	46.868	3600	14.903	10800	9.719	900	6.0475	600	4.104	7200	2.376	259200	2.376	172800	1.944	300	1.728	60	1.5119
February 2014	86400	45.520	3600	17.025	10800	7.527	900	6.631	600	3.943	300	2.688	259200	2.688	7200	2.330	172800	1.971	8600	1.434
March 2014	86400	48.639	3600	15.971	10800	8.711	900	4.900	600	4.900	300	3.085	7200	2.722	259200	1.452	172800	1.452	14400	1.270
June 2014*	86400	45.088	3600	17.719	60	5.263	900	4.211	600	4.211	1800	3.333	10800	3.333	300	3.158	5	2.105	259200	2.105
July 2014	86400	49.825	3600	19.825	10800	4.561	900	4.386	600	3.509	1800	2.982	300	2.632	259200	2.281	7200	1.930	60	1.579
August 2014	86400	50.635	3600	19.056	900	5.263	10800	4.356	1800	3.086	600	2.722	172800	2.722	300	2.541	259200	2.178	14400	1.996
September 2014	86400	53.800	3600	17.288	900	5.514	1800	4.173	10800	3.428	600	2.832	300	2.236	7200	1.490	60	1.490	14400	1.490
October 2014	86400	52.191	3600	16.467	900	6.906	10800	5.976	1800	3.718	300	2.523	600	2.390	7200	1.859	172800	1.461	60	1.328
November 2014	86400	51.598	3600	17.047	900	7.610	10800	6.697	1800	4.110	300	2.435	7200	1.979	600	1.674	60	1.370	259200	0.761
December 2014	86400	41.812	3600	26.829	900	5.575	1800	5.226	10800	4.181	300	3.484	60	2.439	7200	2.091	259200	2.091	600	1.394
January 2015	86400	44.301	3600	22.243	900	7.537	10800	4.412	1800	3.309	300	3.125	259200	2.757	7200	2.390	600	1.838	172800	1.287
February 2015	86400	49.394	3600	19.650	10800	7.133	900	6.326	1800	2.423	300	2.019	7200	1.884	60	1.750	172800	1.346	259200	1.211
March 2015	86400	51.122	3600	16.708	900	9.601	10800	7.357	1800	2.618	300	2.244	7200	1.870	172800	1.746	600	1.122	259200	1.122
April 2015	86400	50.432	3600	18.135	10800	7.254	900	6.563	1800	3.282	300	2.764	7200	2.073	600	1.727	172800	1.554	60	1.036
May 2015	86400	50.307	3600	16.973	10800	5.726	900	5.317	1800	3.885	600	2.863	300	2.249	7200	2.045	60	1.636	172800	1.636
June 2015	86400	48.283	3600	18.240	900	5.150	10800	4.721	1800	4.506	600	2.575	300	2.575	7200	2.361	172800	2.146	60	1.502
July 2015	86400	43.927	3600	18.826	10800	6.680	7200	4.656	1800	4.656	900	4.453	300	3.036	259200	2.024	600	1.822	172800	1.822
August 2015	86400	46.723	3600	20.507	900	5.920	1800	5.074	10800	4.652	7200	2.748	600	2.537	300	2.537	172800	1.480	60	1.268

Rank	1		2		3		4		5		6		7		8		9		10	
Month	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs
Oct13	A	74.635	CNAME	16.971	MX	3.832	TXT	3.102	AAAA	1.460	N/A									
Nov13	A	76.974	CNAME	13.816	MX	3.947	TXT	3.289	AAAA	1.974	N/A									
Dec13	A	78.125	MX	8.036	TXT	5.804	CNAME	5.804	AAAA	1.786	SOA	0.446	N/A							
Jan14	A	75.162	CNAME	15.119	MX	4.536	TXT	3.672	AAAA	1.512	N/A									
Feb14	A	73.118	CNAME	17.384	TXT	3.763	MX	3.763	AAAA	1.971	N/A									
Mar14	A	74.229	CNAME	16.878	MX	3.448	TXT	3.267	AAAA	1.996	SOA	0.181	N/A							
Jun14	A	77.895	CNAME	12.456	TXT	3.333	MX	3.333	AAAA	2.632	SOA	0.175	NS	0.175	N/A					
Jul14	A	75.789	CNAME	13.860	MX	4.211	TXT	3.333	AAAA	2.456	SRV	0.175	NS	0.175	N/A					
Aug14	A	75.906	CNAME	13.768	MX	4.529	TXT	3.442	AAAA	2.174	NS	0.181	N/A							
Sep14	A	71.833	CNAME	19.821	TXT	2.981	MX	2.981	AAAA	2.235	NS	0.149	N/A							
Oct14	A	65.073	CNAME	27.888	TXT	2.523	MX	2.390	AAAA	1.992	NS	0.133	N/A							
Nov14	A	63.470	CNAME	28.767	MX	2.892	TXT	2.588	AAAA	1.979	SOA	0.152	NS	0.152	N/A					
Dec14	A	73.519	CNAME	12.195	MX	6.969	TXT	3.833	AAAA	3.484	N/A									
Jan15	A	70.221	CNAME	20.772	MX	3.493	TXT	3.125	AAAA	1.838	SOA	0.368	NS	0.184	N/A					
Feb15	A	59.892	CNAME	32.167	MX	3.365	TXT	2.826	AAAA	1.480	SOA	0.135	NS	0.135	N/A					
Mar15	A	56.663	CNAME	35.990	MX	2.989	TXT	2.740	AAAA	1.494	NS	0.125	N/A							
Apr15	A	62.586	CNAME	28.103	MX	3.621	TXT	3.276	AAAA	2.069	SOA	0.172	NS	0.172	N/A					
May15	A	79.141	CNAME	11.043	TXT	3.885	MX	3.476	AAAA	2.249	NS	0.204	NS	N/A						
Jun15	A	78.326	CNAME	10.730	TXT	4.077	MX	3.863	AAAA	2.790	NS	0.215	N/A							
Jul15	A	80.972	CNAME	7.894	MX	4.656	TXT	3.846	AAAA	2.632	N/A									
Aug15	A	75.687	CNAME	13.531	MX	4.017	TXT	3.594	AAAA	2.960	NS	0.211	N/A							

Table A.10: Top 5 geolocation distribution for .other za domains

Rank	1		2		3		4		5	
Month	Country	% of IPs	Country	% of IPs	Country	% of IPs	Country	% of IPs	Country	% of IPs
October 2013	ZA	82.222	US	6.667	UK	4.444	MU	4.444	DE	2.222
November 2013	ZA	80.000	US	8.889	UK	4.444	MU	4.444	DE	2.222
December 2013	ZA	78.571	US	10.714	UK	3.571	MU	3.571	DE	3.571
January 2014	ZA	80.851	US	8.511	UK	4.255	MU	4.255	DE	2.128
February 2014	ZA	73.810	US	11.905	MU	9.524	UK	2.381	DE	2.381
March 2014	ZA	77.778	US	11.111	MU	4.444	DE	4.444	UK	2.222
June 2014*	ZA	84.000	US	8.000	MU	6.000	UK	2.000	-	-
July 2014	ZA	67.164	US	8.955	MU	4.478	UK	2.985	DE	1.493
August 2014	ZA	86.765	US	7.353	MU	4.412	UK	1.471	-	-
September 2014	ZA	77.778	US	9.259	MU	9.259	UK	1.852	NL	1.852
October 2014	ZA	84.314	US	9.804	UK	1.961	NL	1.961	MU	1.961
November 2014	ZA	82.258	US	8.065	UK	3.226	MU	3.226	NL	1.613
December 2014	ZA	79.545	US	9.091	MU	6.818	UK	2.273	NL	2.273
January 2015	ZA	83.333	US	6.250	MU	6.250	UK	2.083	DE	2.083
February 2015	ZA	82.222	US	6.667	MU	6.667	UK	2.222	DE	2.222
March 2015	ZA	77.083	US	10.417	MU	4.167	DE	4.167	UK	2.083
April 2015	ZA	77.778	MU	6.667	DE	6.667	US	4.444	UK	2.222
May 2015	ZA	77.083	US	8.333	MU	6.250	DE	4.167	UK	2.083
June 2015	ZA	76.316	MU	7.895	US	5.263	DE	5.263	UK	2.632
July 2015	ZA	77.778	US	8.889	MU	6.667	UK	2.222	NL	2.222
August 2015	ZA	78.261	US	10.870	MU	4.348	UK	2.174	NL	2.174

Table A.11: Observed TTL and RRs for other .za domains

Rank	1		2		3		4		5		6		7		8		9		10	
Month	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs	TTL	% of TTLs
October 2013	3600	26.042	84600	23.958	86400	20.833	7200	10.417	600	10.417	28800	3.125	300	2.08333	900	1.042	38400	1.042	14400	1.042
November 2013	84600	25.000	86400	20.833	3600	18.056	7200	11.111	600	11.111	28800	6.9444	300	2.778	900	1.389	1800	1.389	14400	1.389
December 2013	84600	25.000	7200	19.444	86400	16.667	3600	13.889	600	11.111	900	2.778	300	2.778	28800	2.778	1800	2.778	14400	2.778
January 2014	84600	25.000	86400	20.589	3600	17.647	600	11.765	7200	10.294	28800	5.882	300	2.941	900	1.471	38400	1.471	1800	1.471
February 2014	84600	32.222	3600	20.000	86400	16.667	7200	12.222	600	10.000	28800	5.556	300	2.222	900	1.111	N/A			
March 2014	84600	26.761	3600	21.127	86400	18.310	7200	12.676	600	8.451	28800	5.634	300	2.817	900	1.408	6000	1.408	14400	1.408
June 2014*	600	27.419	84600	22.581	86400	16.129	3600	14.516	7200	11.290	300	6.452	14400	1.613	N/A					
July 2014	84600	22.500	3600	22.500	600	18.750	86400	13.750	7200	10.000	300	7.500	14400	2.500	1800	1.250	1200	1.250	N/A	
August 2014	84600	26.389	3600	23.611	86400	12.500	600	12.500	300	6.944	38400	1.389	1800	1.389	14400	1.389	N/A			
September 2014	84600	28.986	86400	18.841	600	18.841	3600	15.942	7200	10.145	300	4.348	14400	2.899	N/A					
October 2014	84600	25.974	86400	23.377	600	18.182	7200	12.987	3600	12.987	300	5.195	14400	1.299	N/A					
November 2014	84600	28.947	86400	18.421	3600	18.421	600	15.789	300	9.211	7200	6.579	38400	1.316	14400	1.316	N/A			
December 2014	84600	26.531	3600	20.408	600	18.367	86400	12.245	7200	12.245	300	4.082	1800	4.082	38400	2.041	N/A			
January 2015	84600	29.688	600	21.875	3600	15.625	86400	10.938	7200	9.375	300	7.813	14400	3.125	38400	1.563	N/A			
February 2015	84600	21.918	3600	21.918	86400	19.178	600	16.438	7200	6.849	300	6.849	38400	2.740	28800	1.370	1800	1.370	14400	1.370
March 2015	84600	26.389	3600	20.833	86400	19.444	600	16.667	7200	9.722	300	2.778	14400	2.778	38400	1.389	N/A			
April 2015	86400	34.286	600	18.571	84600	15.714	3600	15.714	7200	11.429	38400	1.429	300	1.429	14400	1.429	N/A			
May 2015	86400	26.761	84600	18.310	600	16.901	3600	16.901	7200	8.451	38400	4.225	300	4.225	14400	4.225	N/A			
June 2015	86400	24.242	3600	22.727	84600	21.212	600	15.152	7200	6.061	38400	4.545	300	3.031	601	1.515	14400	1.515	N/A	
July 2015	86400	26.389	84600	19.444	600	16.667	3600	16.667	7200	11.111	300	4.167	38400	2.778	1800	1.389	14400	1.389	N/A	
August 2015	86400	27.143	600	20.000	84600	14.286	3600	14.286	7200	5.714	38400	5.714	300	4.286	14400	2.857	7260	1.429	43200	1.429

Rank	1		2		3		4		5		6		7		8		9		10	
Month	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs	RR	% of RRs
Oct13	A	79.167	MX	11.458	CNAME	8.333	SOA	1.042	N/A											
Nov13	A	78.667	MX	10.667	CNAME	8.000	SRV	1.333	SOA	1.333	N/A									
Dec13	A	81.081	MX	16.216	CNAME	2.703	N/A													
Jan14	A	75.362	MX	15.942	CNAME	5.797	TXT	2.899	N/A											
Feb14	A	84.444	MX	7.778	CNAME	4.444	TXT	2.222	SOA	1.111	N/A									
Mar14	A	86.301	MX	6.849	CNAME	4.110	TXT	1.370	SOA	1.370	N/A									
Jun14	A	85.135	MX	6.757	SOA	4.054	CNAME	2.703	SRV	1.351	N/A									
Jul14	A	82.418	MX	13.187	SOA	2.198	CNAME	1.099	AAAA	1.099	N/A									
Aug14	A	82.796	MX	8.602	CNAME	5.376	SOA	3.226	N/A											
Sep14	A	82.143	CNAME	5.952	MX	4.762	SOA	3.571	TXT	1.190	SRV	1.190	AAAA	1.190	N/A					
Oct14	A	84.8834	CNAME	5.814	SOA	4.651	MX	4.651	N/A											
Nov14	A	84.211	MX	5.263	CNAME	5.263	SOA	3.158	TXT	1.053	SRV	1.053	N/A							
Dec14	A	68.657	MX	19.403	SOA	4.478	SRV	2.985	CNAME	2.985	AAAA	1.493	N/A							
Jan15	A	86.585	MX	7.317	SOA	2.439	CNAME	2.439	NS	1.220	N/A									
Feb15	A	80.822	MX	9.589	SOA	4.110	CNAME	4.110	TXT	1.370	N/A									
Mar15	A	81.944	MX	8.333	CNAME	5.556	SOA	2.778	TXT	1.389	N/A									
Apr15	A	81.429	MX	8.571	CNAME	5.714	TXT	2.857	SOA	1.429	N/A									
May15	A	74.648	MX	12.676	CNAME	7.042	TXT	5.634	N/A											
Jun15	A	72.727	MX	18.182	TXT	4.545	CNAME	3.030	SOA	1.515	N/A									
Jul15	A	81.944	MX	11.111	TXT	4.167	CNAME	2.778	N/A											
Aug15	A	80.000	MX	11.429	CNAME	4.286	TXT	2.857	SOA	1.429	N/A									

Table A.12: Temporal relationship between attacks and scans October 2013

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
30259.info	9 October 2013	0	10 Oct 13	17	22 Oct 13
36088.info	11 October 2013	3	11 Oct 13	1	13 Oct 13
36372.info	15 October 2013	8	14 Oct 13	0	14 Oct 13
37349.info	15 October 2013	0	16 Oct 13	44	18 Oct 13
aa.10781.info	12 October 2013	0	13 Oct 13	4	16 Oct 13
babywow.co.uk	11 October 2013	0	12 Oct 13	6	18 Oct 13
bitstress.com	21 September 2013	0	1 Oct 13	1	1 Oct 13
fkfkfkfa.com	23 September 2013	0	1 Oct 13	5	26 Oct 13
gtml2.com	19 October 2013	0	20 Oct 13	3	31 Oct 13
Hizbullah.me	28 July 2013	0	26 Oct 13	1	26 Oct 13
irlwinning.com	2 October 2013	3	1 Oct 13	3	10 Oct 13
krasti.us	18 October 2013	0	19 Oct 13	1	19 Oct 13
pipcvsemnaher.com	17 October 2013	0	18 Oct 13	2	31 Oct 13
pkts.asia	1 October 2013	1	1 Oct 13	11	31 Oct 13
Sandia.gov	28 September 2013	0	4 Oct 13	2	28 Oct 13
txt.fwserver.com.ua	18 October 2013	0	23 Oct 13	4	26 Oct 13
zzgst.com	9 September 2013	0	2 Oct 13	1	2 Oct 13

Table A.13: Temporal relationship between attacks and scans November 2013

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
36088.info	11 October 2013	0	17 Nov 13	1	17 Oct 13
bitchgotraped.cloudns.eu	21 November 2013	1	19 Nov 13	0	19 Nov 13
cheatsharez.com	11 November 2013	0	12 Nov 13	4	16 Nov 13
eschenemnogo.com	19 November 2013	1	19 Nov 13	4	25 Nov 13
fkfkfkfa.com	23 September 2013	0	5 Nov 13	3	30 Nov 13
hccforums.nl	10 November 2013	1	10 Nov 13	1	29 Nov 13
Hizbullah.me	28 July 2013	0	4 Nov 13	1	4 Nov 13
krasti.us	18 October 2013	0	13 Nov 13	4	24 Nov 13
lrc-pipec.com	14 November 2013	0	16 Nov 13	1	16 Nov 13
pkts.asia	1 October 2013	0	3 Nov 13	3	7 Nov 13
reanimator.in	1 November 2013	0	2 Nov 13	3	11 Nov 13
Sandia.gov	28 September 2013	0	6 Nov 13	4	20 Nov 13
siska1.com	9 November 2013	0	10 Nov 13	2	17 Nov 13
stopdrugs77.com	27 November 2013	1	27 Nov 13	0	27 Nov 13
thebestdomainintheworld.cloudns.eu	15 November 2013	0	16 Nov 13	1	16 Nov 13
t.pbub.info	6 November 2013	0	7 Nov 13	3	13 Nov 13
x.mpnp.info	14 November 2013	0	15 Nov 13	2	17 Nov 13
x.privetrc.com	19 November 2013	1	19 Nov 13	1	21 Nov 13
x.slnm.info	17 November 2013	0	18 Nov 13	1	18 Nov 13

Table A.14: Temporal relationship between attacks and scans December 2013

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
adrenalinessss.cc	7 December 2013	0	20 Dec 13	1	20 Dec 13
amp.crack-zone.ru	22 December 2013	0	23 Dec 13	2	27 Dec 13
datburger.cloudns.org	8 December 2013	0	22 Dec 13	1	22 Dec 13
dnsamplificationattacks.cc	4 December 2013	0	6 Dec 13	3	20 Dec 13
fkfkfkfa.com	23 September 2013	0	5 Dec 13	5	30 Dec 13
grungyman.cloudns.org	17 December 2013	0	18 Dec 13	2	22 Dec 13
ilineage2.ru	6 December 2013	0	8 Dec 13	3	21 Dec 13
krasti.us	18 October 2013	0	2 Dec 13	2	19 Dec 13

Table A.15: Temporal relationship between attacks and scans January 2014

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
bitchgotraped.cloudns.eu	21 November 2013	0	12 Jan 14	1	12 Jan 14
dnsamplificationattacks.cc	4 December 2013	0	19 Jan 14	1	19 Jan 14
fkfkfkfa.com	23 September 2013	0	2 Jan 14	5	10 Jan 14
gtml2.com	19 October 2013	0	13 Jan 14	1	13 Jan 14
krasti.us	18 October 2013	0	1 Jan 14	2	12 Jan 14
pddos.com	5 January 2014	0	7 Jan 14	3	16 Jan 14
Sandia.gov	28 September 2013	0	9 Jan 14	2	12 Jan 14
saveroads.ru	2 January 2014	0	3 Jan 14	2	15 Jan 14
txt.fwserver.com.ua	18 October 2013	0	19 Jan 14	1	19 Jan 14
x.xipzersscc.com	24 January 2014	0	25 Jan 14	1	25 Jan 14
Zong.Zong.Co.Ua	8 January 2014	0	10 Jan 14	1	10 Jan 14

Table A.16: Temporal relationship between attacks and scans February 2014

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
evgeniy-marchenko.cc	23 August 2013	0	17 Feb 14	2	18 Feb 14
fkfkfkfr.com	10 February 2014	0	13 Feb 14	3	17 Feb 14
gerdar3.ru	10 February 2014	0	11 Feb 14	4	25 Feb 14
gtml2.com	19 October 2013	0	1 Feb 14	2	15 Feb 14
Hizbullah.me	28 July 2013	0	4 Feb 14	1	4 Feb 14
krasti.us	18 October 2013	0	9 Feb 14	2	15 Feb 14
nlhosting.nl	17 October 2013	0	8 Feb 14	1	8 Feb 14
pddos.com	5 January 2014	0	1 Feb 14	7	10 Feb 14
Sandia.gov	28 September 2013	0	15 Feb 14	1	15 Feb 14
saveroads.ru	2 January 2014	0	9 Feb 14	1	9 Feb 14
supermegatrue.mcdir.ru	10 October 2013	0	18 Feb 14	1	18 Feb 14
thebestdomainintheworld.cloudns.eu	15 November 2013	0	15 Feb 14	1	15 Feb 14
txt409.tekjeton.com	10 October 2013	0	18 Feb 14	1	18 Feb 14

Table A.17: Temporal relationship between attacks and scans March 2014

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
admin.blueorangecare.com	14 March 2014	0	22 Mar 14	2	29 Mar 14
ahuyehue.info	8 March 2014	0	9 Mar 14	8	28 Mar 14
ddosforums.pw	5 April 2014	1	29 Mar 14	0	29 Mar 14
fkfkfkfr.com	10 February 2014	0	29 Mar 14	1	29 Mar 14
thebestdomainintheworld.cloudns.eu	15 November 2013	0	26 Mar 14	1	26 Mar 14
www.jrdga.info	1 March 2014	0	2 Mar 14	5	26 Mar 14

Table A.18: Temporal relationship between attacks and scans June 2014

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
ahuyehue.info	8 March 2014	0	21 Jun 14	1	21 Jun 14
bangtest.zong.co.ua	28 June 2014	2	7 Jun 14	1	30 Jun 14
ddosforums.pw	5 April 2014	0	25 Jun 14	2	30 Jun 14
doleta.gov	16 October 2014	2	24 Jun 14	0	29 Jun 14
gtml2.com	19 October 2013	0	30 Jun 14	1	30 Jun 14
lalka.com.ru	28 June 2014	0	29 Jun 14	3	30 Jun 14
magas.bslrpg.com	14 May 2014	0	7 Jun 14	8	15 Jun 14
webpanel.sk	23 July 2014	1	30 Jun 14	0	30 Jun 14
wradish.com	27 April 2014	0	17 Jun 14	2	21 Jun 14

Table A.19: Temporal relationship between attacks and scans July 2014

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
energystar.gov	13 October 2014	1	4 Jul 14	0	6 Jul 14
gtml2.com	19 October 2013	0	1 Jul 14	2	15 Jul 14
krasti.us	18 October 2013	0	2 Jul 14	3	13 Jul 14
lalka.com.ru	28 June 2014	0	6 Jul 14	7	25 Jul 14
svist21.cz	12 November 2014	3	14 Jul 14	0	23 Jul 14
webpanel.sk	23 July 2014	0	24 Jul 14	5	31 Jul 14
wradish.com	27 April 2014	0	1 Jul 14	6	26 Jul 14
www.jrdga.info	1 March 2014	0	3 Jul 14	1	4 Jul 14

Table A.20: Temporal relationship between attacks and scans August 2014

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
bangtest.zong.co.ua	28 June 2014	0	25 Aug 14	1	25 Aug 14
ddosforums.pw	5 April 2014	0	16 Aug 14	1	16 Aug 14
doleta.gov	16 October 2014	2	25 Aug 14	0	25 Aug 14
energystar.gov	13 October 2014	6	20 Aug 14	0	31 Aug 14
gtml2.com	19 October 2013	0	25 Aug 14	2	26 Aug 14
lalka.com.ru	28 June 2014	0	2 Aug 14	2	9 Aug 14
magas.bslrpg.com	14 May 2014	0	31 Aug 14	1	31 Aug 14
svist21.cz	12 November 2014	2	21 Aug 14	0	25 Aug 14
thebestdomainintheworld.cloudns.eu	15 November 2013	0	24 Aug 14	1	24 Aug 14
webpanel.sk	23 July 2014	0	1 Aug 14	15	31 Aug 14
wradish.com	27 April 2014	0	1 Aug 14	5	31 Aug 14

Table A.21: Temporal relationship between attacks and scans September 2014

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
doleta.gov	16 October 2014	4	5 Sep 14	0	25 Sep 14
energystar.gov	13 October 2014	4	1 Sep 14	0	17 Sep 14
sema.cz	11 July 2013	0	30 Sep 14	1	30 Sep 14
svist21.cz	12 November 2014	1	18 Sep 14	0	18 Sep 14
webpanel.sk	23 July 2014	0	1 Sep 14	14	30 Sep 14
wradish.com	27 April 2014	0	8 Sep 14	1	8 Sep 14
www.jrdga.info	1 March 2014	0	17 Sep 14	1	17 Sep 14

Table A.22: Temporal relationship between attacks and scans October 2014

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
bitchgotraped.cloudns.eu	21 November 2013	0	7 Oct 14	1	10 Oct 14
bmw.digmehl.cu.cc	16 October 2014	0	19 Oct 14	4	28 Oct 14
datburger.cloudns.org	8 December 2013	0	10 Oct 14	1	10 Oct 14
dnsamplificationattacks.cc	4 December 2013	0	8 Oct 14	1	8 Oct 14
doleta.gov	16 October 2014	0	31 Oct 14	1	31 Oct 14
energystar.gov	13 October 2014	2	5 Oct 14	3	30 Oct 14
evgeniy-marchenko.cc	23 August 2013	0	10 Oct 14	1	10 Oct 14
grungyman.cloudns.org	17 December 2013	0	7 Oct 14	1	8 Oct 14
guessinfosys.com	13 October 2014	0	15 Oct 14	1	15 Oct 14
nlhosting.nl	17 October 2013	0	18 Oct 14	1	19 Oct 14
notthebestdomainintheworld.cloudns.org	28 November 2013	0	7 Oct 14	1	9 Oct 14
supermegatrue.mcdir.ru	10 October 2013	0	8 Oct 14	1	10 Oct 14
svist21.cz	12 November 2014	1	18 Oct 14	0	18 Oct 14
thebestdomainintheworld.cloudns.eu	15 November 2013	0	8 Oct 14	1	9 Oct 14
txt409.tekjeton.com	10 October 2013	0	8 Oct 14	1	10 Oct 14
wradish.com	27 April 2014	0	5 Oct 14	6	31 Oct 14

Table A.23: Temporal relationship between attacks and scans November 2014

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
bmw.digmehl.cu.cc	16 October 2014	0	1 Nov 14	3	19 Nov 14
doleta.gov	16 October 2014	0	6 Nov 14	4	25 Nov 14
gransy.com	1 January 2015	1	27 Nov 14	0	27 Nov 14
krasti.us	18 October 2013	0	27 Nov 14	1	27 Nov 14
nlhosting.nl	17 October 2013	0	1 Nov 14	3	24 Nov 14
non.digmehl.cu.cc	25 Novermber 2014	0	26 Nov 14	1	26 Nov 14
svist21.cz	12 November 2014	0	13 Nov 14	3	20 Nov 14
wradish.com	27 April 2014	0	20 Nov 14	2	22 Nov 14

Table A.24: Temporal relationship between attacks and scans December 2014

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
basjuk.com.ru	9 December 2014	2	8 Dec 14	0	9 Dec 14
defcon.org	22 December 2014	0	24 Dec 14	9	31 Dec 14
doleta.gov	16 October 2014	0	7 Dec 14	2	16 Dec 14
energystar.gov	13 October 2014	0	19 Dec 14	1	19 Dec 14
free-google-2.cloudns.org	8 December 2014	0	13 Dec 14	1	13 Dec 14
globe.gov	17 December 2014	2	15 Dec 14	6	31 Dec 14
gransy.com	1 January 2015	2	2 Dec 14	0	25 Dec 14
maximumstresser.net	17 December 2014	2	15 Dec 14	1	21 Dec 14
pizdaizda.com.ru	13 December 2014	0	15 Dec 14	1	15 Dec 14
svist21.cz	12 November 2014	0	21 Dec 14	3	31 Dec 14

Table A.25: Temporal relationship between attacks and scans January 2015

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
defcon.org	22 December 2014	0	1 Jan 15	3	6 Jan 15
doleta.gov	16 October 2014	0	8 Jan 15	1	8 Jan 15
energystar.gov	13 October 2014	0	21 Jan 15	1	21 Jan 15
globe.gov	17 December 2014	0	7 Jan 15	3	21 Jan 15
gransy.com	1 January 2015	1	1 Jan 15	3	5 Jan 15
nlhosting.nl	17 October 2013	0	23 Jan 15	1	23 Jan 15
pidarastik.ru	24 February 2015	2	14 Jan 15	0	23 Jan 15
uzuzuu.ru	9 February 2015	2	23 Jan 15	0	29 Jan 15

Table A.26: Temporal relationship between attacks and scans February 2015

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
cdnmyhost.com	24 February 2015	2	18 Feb 15	0	20 Feb 15
defcon.org	22 December 2014	0	6 Feb 15	5	16 Feb 15
globe.gov	17 December 2014	0	15 Feb 15	1	15 Feb 15
gransy.com	1 January 2015	0	1 Feb 15	3	23 Feb 15
inboot.co	17 December 2014	0	19 Dec 15	1	20 Feb 15
pidarastik.ru	24 February 2015	5	11 Feb 15	0	20 Feb 15
svist21.cz	12 November 2014	0	18 Feb 15	1	18 Feb 15
uzuzuu.ru	9 February 2015	1	9 Feb 15	0	9 Feb 15
vlch.net	17 December 2014	0	7 Feb 15	2	18 Feb 15

Table A.27: Temporal relationship between attacks and scans March 2015

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
cdnmyhost.com	24 February 2015	0	3 Mar 15	5	30 Mar 15
defcon.org	22 December 2014	0	7 Mar 15	3	30 Mar 15
fkfkfkfa.com	23 September 2013	0	7 Mar 15	1	7 Mar 15
gransy.com	1 January 2015	0	8 Mar 15	3	18 Mar 15
hccforums.nl	10 November 2013	0	28 Mar 15	1	28 Mar 15
viareality.cz	24 February 2015	0	9 Mar 15	2	12 Mar 15

Table A.28: Temporal relationship between attacks and scans April 2015

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
cdnmyhost.com	24 February 2015	0	4 Apr 15	2	8 Apr 15
defcon.org	22 December 2014	0	9 Apr 15	2	29 Apr 15
hccforums.nl	10 November 2013	0	10 Apr 15	1	10 Apr 15
pidarastik.ru	24 February 2015	0	17 Apr 15	1	17 Apr 15

Table A.29: Temporal relationship between attacks and scans May 2015

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
defcon.org	22 December 2014	0	7 May 15	3	28 May 15
domenamocy.pl	23 October 2014	0	16 May 15	1	16 May 15
energystar.gov	13 October 2014	0	4 May 15	2	23 May 15
freeinfosys.com	25 November 2014	0	24 May 15	1	24 May 15
globe.gov	17 December 2014	0	16 May 15	1	16 May 15
gransy.com	1 January 2015	0	15 May 15	4	25 May 15
magas.bslrpg.com	14 May 2014	0	16 May 15	1	16 May 15
svist21.cz	12 November 2014	0	16 May 15	2	31 May 15
viareality.cz	24 February 2015	0	16 May 15	2	26 May 15

Table A.30: Temporal relationship between attacks and scans June 2015

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
cdnmyhost.com	24 February 2015	0	22 Jun 15	1	22 Jun 15
defcon.org	22 December 2014	0	11 Jun 15	4	29 Jun 15
energystar.gov	13 October 2014	0	6 Jun 15	2	26 Jun 15
globe.gov	17 December 2014	0	7 Jun 15	2	23 Jun 15
svist21.cz	12 November 2014	0	6 Jun 15	1	6 Jun 15
vlch.net	17 December 2014	0	14 Jun 15	1	14 Jun 15

Table A.31: Temporal relationship between attacks and scans July 2015

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
067.cz	11 November 2014	0	10 Jul 15	1	10 Jul 15
defcon.org	22 December 2014	0	3 Jul 15	3	10 Jul 15
energystar.gov	13 October 2014	0	8 Jul 15	2	31 Jul 15
svist21.cz	12 November 2014	0	4 Jul 15	2	8 Jul 15

Table A.32: Temporal relationship between attacks and scans August 2015

Domain	Reported attack date*	# of scans before attack	First recorded scan	# of scans after attack	Last recorded scan
defcon.org	22 December 2014	0	2 Aug 15	3	29 Aug 15
energystar.gov	13 October 2014	0	4 Aug 15	1	4 Aug 15
globe.gov	17 December 2014	0	4 Aug 15	1	4 Aug 15
gransy.com	1 January 2015	0	2 Aug 15	2	7 Aug 15
svist21.cz	12 November 2014	0	4 Aug 15	1	4 Aug 15