2014

A survey of IPV6 address usage in the public domain name system

Clinton R. Carpene
Edith Cowan University

Andrew Woodward
Edith Cowan University

DOI: 10.4225/75/57b666d2343d8

12th Australian Information Security Management Conference. Held on the 1-3 December, 2014 at Edith Cowan University, Joondalup Campus, Perth, Western Australia.
This Conference Proceeding is posted at Research Online.
https://ro.ecu.edu.au/ism/171

A SURVEY OF IPV6 ADDRESS USAGE IN THE PUBLIC DOMAIN NAME SYSTEM

Clinton Carpene, Andrew Woodward
School of Computer Security Science & Security Research Institute
Edith Cowan University, Perth, Australia
c.carpene@ecu.edu.au, a.woodward@ecu.edu.au

Abstract

The IPv6 protocol has been slowly increasing in use on the Internet. The main reason for the development of the protocol is that the address space provided by IPv4 is nearing exhaustion. The pool of addresses provided by IPv6 is 2^{96} times larger than IPv4, and should be sufficient to provide an address for every device for the foreseeable future. Another potential advantage of this significantly large address space is the use of randomly assigned addresses as a security barrier as part of a defence in depth strategy. This research examined the addresses allocated by those implementing IPv6 to determine what method or pattern of allocation was being used by adopters of the protocol. This examination was done through the use of DNS queries of the AAAA IPv6 host record using public DNS servers. It was observed that 55.84% of IPv6 addresses were in the range of 0 to ($2^{32}-1$). For those addresses with unique interface identifier (IID) portions, a nearly equal number of sequential and random IIDs were observed. Hong Kong and Germany were found to have the greatest number of IPv6 addresses. These results suggest that adopters are allocating most addresses sequentially, meaning that no security advantage is being obtained. It is unclear as to whether this is through design or the following of accepted practice. Future research will continue to survey the IPv6 address space to determine whether the patterns observed here remain constant.

Keywords

IP Networks, TCPIP, Internet, Domain Name System, IPv6

INTRODUCTION

IPv6 usage is on the rise. Although slow to gain traction, the protocol has seen an upward trend in global usage over the past few years. Currently IPv6 accounts for approximately 5% of Google's search traffic (Google, 2014) and approximately 0.5% for Akamai Technologies' global content delivery network (Akamai Technologies, 2014). In order for a device to communicate using IPv6, it must have a valid IPv6 address. An interface identifier (IID) is the lower order 64 bits of an IPv6 address (see Figure 1, whilst the network portion (encompassing the subnet id) is the higher order 64 bits of the address (as depicted in Figure 1). This distinction is of interest to the survey paper because IIDs identify hosts, and are not necessarily globally unique. There are also many ways in which a device may be configured with an IID. These ways include stateless means, using self-configuration techniques such as stateless address autoconfiguration (SLAAC) and SLAAC with privacy extensions, or through stateful means, such as the dynamic host configuration protocol version 6 (DHCPv6) or manual address allocation.

```
+-------------------------+-----------+----------------------------+
|         n bits          |   m bits  |       128-n-m bits         |
+-------------------------+-----------+----------------------------+
| global routing prefix   | subnet ID |       interface ID         |
+-------------------------+-----------+----------------------------+
```

Figure 1 - Global unicast IPv6 address structure (R. Hinden & Nordmark, 2003). Figure displays how an IPv6 address can be deconstructed into the global routing prefix, the subnet ID, which when coupled form the network portion of the address, and the interface ID which represents the host portion of the address.

The resulting IIDs of the different address configuration schemes can be used to influence heuristics. As mentioned in Carpene & Woodward (2012), DHCPv6 or manual allocations can result in low entropy sequential IIDs (starting at address 0 and incrementing upward). Alternatively, stochastic generation schemes, such as privacy extensions, random DHCPv6 allocation, or cryptographically generated addresses (CGAs) can result in high entropy, seemingly random IIDs being created. SLAAC addresses that have been constructed without the use of privacy extensions have a predictable pattern of 0xfffe between bits 24 and 42 of the address. Additionally, since the process used to create the IID (the modified extensible unique identifier - or EUI-64 (IEEE Standards Association, nd)) uses the host's MAC address, there are significant privacy issues associated with this method. It has been demonstrated by Carpene et al. (2014) that machine learning-based classification systems can be used to classify interface identifiers into categories that reflect how they were constructed. Having been the subject of much debate regarding privacy and practicality (Dunlop et al., 2011b,a; Carpene & Woodward, 2012), research was undertaken to assess the trends in usage of IPv6 in real-world situations.

The research questions that directed this study are as follows:

1) How are publicly accessible IPv6 addresses configured?
2) Are the interface identifiers (IIDs) that are allocated to public hosts unique?

To this end a survey was conducted to attempt to answer the research questions. The survey involved gathering information about publicly accessible IPv6 records and drawing inferences from the data.

Why is this research of significance, and significant to cyber security in general? IPv6 usage is of interest due to the security implications and the rise of the Internet of things (Giusto et al., 2010). The 2^{128} addresses provided by IPv6 will result in nearly every device being theoretically capable of communicating on the public IPv6 Internet. For example, IPv6 enabled light globes were announced in 2012, and we have subsequently seen an expansion in the range of devices that are Internet addressable (Turiel, 2011). Whilst there is no great degree of risk associated with using an internet enabled light globe, of greater concern is the expansion in use of IPv6 enabled bio sensors and other medical devices (Jara et al., 2010).

Along with this increase in devices which are being networked, is the errant claim that IPv6 is inherently more secure than IPv4, and as such the use of IPv6 for medical devices is therefore more secure (Das, 2008). Whilst IPv6 does introduce some new security features, there is no evidence to suggest that the use of IPv6 over IPv4 provides any security benefits. This is highlighted by research which aims to introduce extra layers of security around the use of IPv6 in medical devices, as there are clearly elevated risks associated with its implementation as highlighted by other research (Jara et al., 2010).

Similar efforts that examine the usage of IPv6 globally have been underway by organisations such as Google (2014) and Akamai Technologies (2014) with respect to the services they offer. These efforts track the usage habits of IPv6 clients when accessing services. Existing research is focused on the proportion of devices using IPv6 for accessing services. This differs from the study presented since this research is concerned with how IPv6 is being implemented, rather than its perpetuation.

INFORMATION SOURCES

In order to obtain the IPv6 addresses for Internet hosts, the author conducted an enumeration of domain name system (DNS) records. For this DNS enumeration to occur a list of potential domain names was required. The author gathered a list of known Internet domain names and subdomains from publicly available information sources.

Of particular interest was the Alexa Top 1 million websites (Alexa Internet, Inc., 2014b). Alexa Internet, Inc. is an organisation that provides access to analytical information and Internet statistics. The company provides a publicly available comma separated values (CSV) file containing a list of Alexa Internet, Inc.'s (2014b) top one million recorded websites. According to Alexa, this statistic is calculated based upon an average of the daily pageviews and unique website visitors (Alexa Internet, Inc., 2014a).

Secondly, an open-sourced project called ipv6-hosts (xslidian & VersusClyne, 2014) exists that provides a comprehensive list of IPv6 addresses and IPv6-enabled domains for users to download. These records correlate

to commonly used Internet services such as YouTube, Facebook, and Google.
Thirdly, a website scraper was programmed to extract domain names from the Australian A-Z list of government websites (Australia.gov.au, 2014). This source contains a list of all of the known Australian government websites. The scraper parsed the URLs gathered from the information source and extracted their domain names.
Finally, a number of domain names were gathered from public sources that were unrecorded at the time of the survey.

Table 1 - Information sources used to provide domain names for the DNS enumeration. Against each source are the total IPv6-enabled subdomains and IPv6 addresses extracted from the survey.

Source Name	Unique IPv6 Addresses gathered	Unique IPv6 enabled subdomains	Description
Alexa top one million websites	151360	109276	The Alexa company posts a list of the top one million websites daily. Each subdomain in this list was queried and the results collected (Alexa Internet, Inc., 2014b).
Unknown	61116	46063	These addresses were collected from sources that were unrecorded at the time the survey was conducted.
IPv6-Hosts list	4032	2176	A public information source providing a list of commonly used internet services and their IPv6 addresses (xslid- ian & VersusClyne, 2014).
Australian Government website	4	4	List of Australian government subdomains (Australia.gov.au, 2014).
Total	216512	157519	

METHODOLOGY

A study was conducted in order to ascertain how IPv6 addresses are being allocated in the public IPv6 Internet. The instrument used to test the research questions was a survey. The intention of the survey was to assess real world usage habits of the IPv6 protocol. The survey was conducted over a period of 454 days. Data was collected through the use of DNS enumeration, which involved querying public DNS servers for IPv6 AAAA host records. These subdomains were obtained from a number of sources (see Table 1) over a total of 969 intervals. The procedure used when conducting the survey is detailed below.

1) Gather list of potential subdomains: the list of subdomains used for the DNS enumeration attempt were first gathered from the sources of subdomains listed in Table 1. Additionally, for each unique root domain name gathered, the following potential subdomains were also compiled into the list: ipv6.<domain>, www.<domain>, mail.<domain>, intranet.<domain> and vpn.<domain>. These prefixes were chosen since they are commonly used for public Internet services.
2) Extract possible IPv6 address(es) from each subdomain: Each subdomain from the list was then queried for an AAAA IPv6 host record using public DNS servers. The IPv6 addresses from valid answers were collected.
3) Store the results: The results were gathered into a database of IPv6 addresses, along with the subdomain the address corresponded to, a timestamp and the source list of domains the result originated from.
4) Analyse the data: Data analysis was then conducted on surveyed data. This included descriptive statistics, classification of data, and correlating geographical information.

The survey was conducted on an eight-core computer using the Linux Mint 14 operating system. A program was created using Python (version 2.7) to parse the subdomain lists, and conduct the DNS enumeration. This program saved data to a locally hosted MySQL database for record storage. Data from the MySQL database were extracted and manipulated using Python programs and libraries. In particular the *numpy* (McKinney, 2010) and *pandas* (McKinney, 2012) libraries were used for data manipulation and transformations. The survey's network transmissions were handled by a Cisco 6504 layer 3 switch.

RESULTS

The survey resulted in many examples of IPv6 addresses being used in public networks globally. The information presented has been anonymised and removed of any identifying factors. Domain names have been removed, as well as complete IPv6 addresses. Only the interface identifiers have been included where necessary, as this information is not globally unique or permanent. Table 1 outlines the sources used to survey information about IPv6 usage. In total, 216,512 unique IPv6 addresses were gathered during the survey phase from the surveyed sources.

The data were then analysed to determine the distribution of observed addresses across the prefixed subdomains. It was observed that 107,944 of the collected IPv6 were associated to a root domain name with no prefixed subdomain. Table 3 exposes the results for the number of IPv6 addresses observed for each of the prefixed subdomains.

Since an IPv6 address can be divided into a network portion and an interface identifier (IID), the data was transformed to remove the network portion and maintain the IID. IIDs are interesting as they are the host portion of an IPv6 address, and can be allocated in numerous ways as mentioned previously. This is contrary to the network portion, which is generally assigned by an authority such as an ISP or other Internet register. Table 2 shows the unique IPv6 IIDs that were gathered from each of the survey sources. Of the 216,512 unique IPv6 addresses gathered, a total of 41,171 unique IPv6 IIDs were extracted.

Table 2 - The results from the IPv6 Survey revealing the number of unique IPv6 IIDs gathered from each information source.

Source name	IIDs Observed
Alexa Top one million websites	29231
Unknown	16560
IPv6-Hosts list	297
Australian Government websites	4
Total unique IIDs collected	41171
Total observed IIDs	216512

Table 3 - Observed frequencies of IPv6 enable domains by subdomain name. Each root domain name extracted from the sources in Table 2 was queried for AAAA records at potential subdomains

Subdomain	Count
root domain (no subdomain)	107944
www.	55811
ipv6.	17604
mail.	15015
intranet.	10106
vpn.	10032

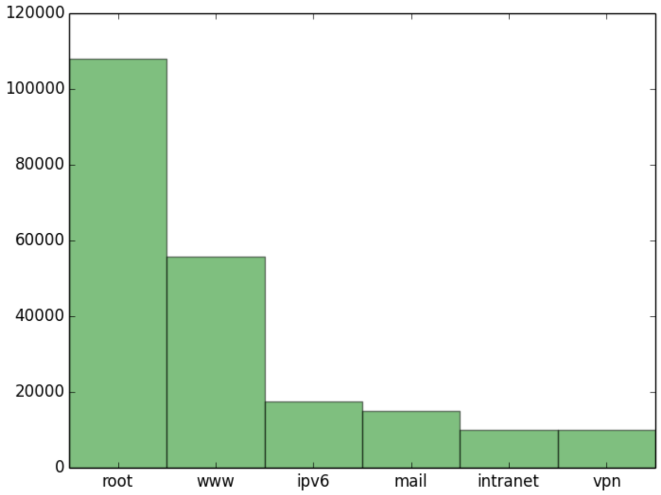

Figure 2 - Distribution of IPv6-enabled subdomains observed in the IPv6 survey. The “root” x label refers to the root

domain name (with no subdomain prefixed), whilst the other labels indicate the prefixed subdomains. The y axis displays the frequency for each subdomain.

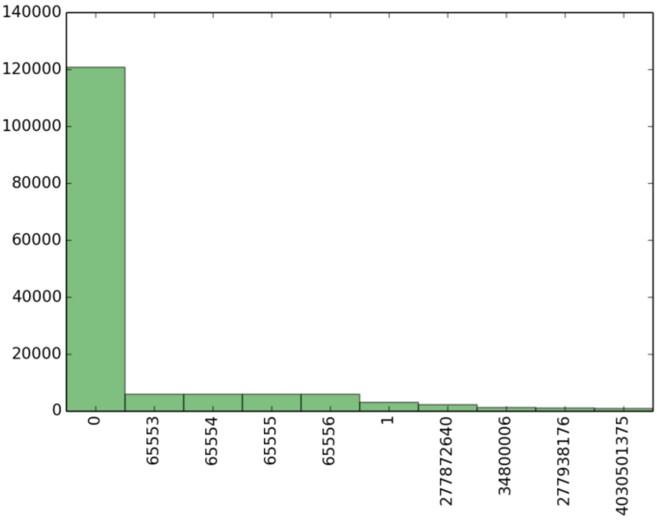

Figure 3 - Chart of IID frequency distribution observed from the IPv6 survey chunked into 2^{32} buckets, each containing 2^{32} IPv6 IIDs. The x axis represents the top ten observed bucket numbers (between 0 and 2^{32}) whilst the y axis represents the frequency of IIDs per bucket. The distribution appears to be weighted heavily toward the 0th bucket (encompassing addresses 0 to $2^{32} - 1$)

Table 4 – Observed IPv6 IIDs classified using an ANN classification system by their construction type.

Classification type	Unique IIDs Observed	All observed IIDs (including duplicates)
Incremental/Manually Assigned	20338	159939
Stochastic	18473	49967
EUI-64	2360	6606

Next the collected IID data were chunked into 2^{32} chunks (or buckets) that each represented 2^{32} IPv6 IIDs. From this a frequency count was conducted to determine the distribution of IIDs over the entire 64 bit address space. It was apparent that the majority of IIDs collected in the survey were contained to the first 32 bits of the 64 bit address space. In fact, as can be seen in Figure 3, over half of the observed IIDs were contained within the first 32 bit bucket of the address spaces (120,898 out of 216,512 total observed IIDs). There appeared to be an even distribution of addresses over the remainder of the space (Figure 3).

After classifying the surveyed data using an artificial neural network (ANN) classification system described in Carpene et al. (2014), it was observed that the majority of sampled IIDs conformed to the incremental (or manually assigned) IID construction type (displayed in Table 4). Separating the unique data, however displayed a different trend. Out of the unique IIDs collected, almost the same amounts of IIDs were distributed between the stochastically generated IID type and the incremental IID type, indicating that the majority of IIDs gathered are not unique (see Figure 4). It was observed that the incrementally allocated addresses contained a significant amount of duplicated data, as only 12.71% of the observed IIDs in this category were unique. The lack of unique IID allocations is further evidenced in Figure 5 where it can be seen that the top 5 out of the top 25 observed IIDs were all observed greater than 1,000 times.

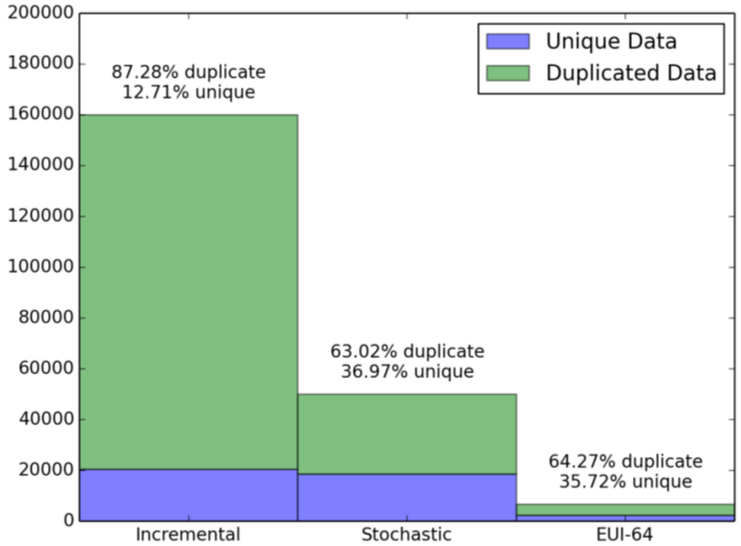

Figure 4 - IID data collected in the survey was classified using an ANN classification system (Carpene et al., 2014) into their relevant construction types. The x axis denotes the classification category, whilst the y axis shows the frequency of observed IIDs in each category. The subplots depict the number of unique IIDs and the total number of IIDs observed in each construction category respectively. The percentage of unique IIDs classified for each category is included at the head of each bar.

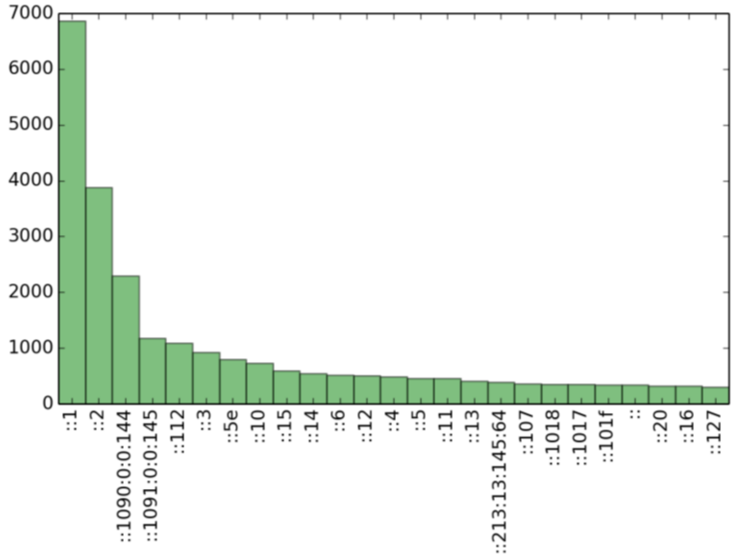

Figure 5 - Histogram of the most observed IPv6 interface identifiers gathered during the survey. The x axis denotes the IPv6 IID (lowest order 64 bits of the IPv6 address), whilst the y axis shows the frequency of the observed IID.

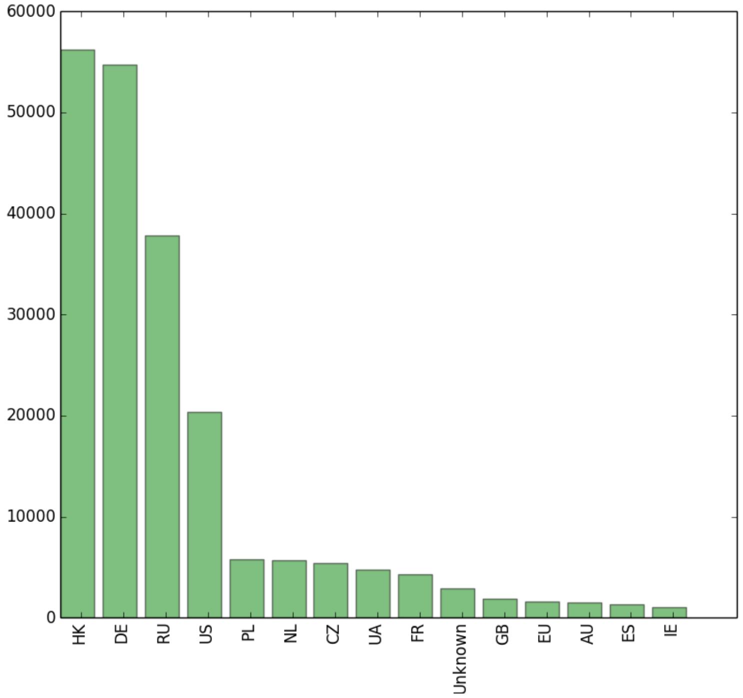

Figure 6 - The top 15 countries observed according to GeoIPv6 correlation. GeoIPv6 correlation was conducted using the MaxMind GeoLite IPv6 Country database (MaxMind, Inc., 2014). The Unknown entity refers to IPv6 addresses that could not resolve to a country using GeoIPv6.

DISCUSSION

The data from Figure 3 clearly shows that the distribution of IID numbers is weighted toward the first 32 bits of the IPv6 IID address space. This information could be used to improve IPv6 host discovery techniques, since an attacker might only need to check the first 32 bits of a network's address space to successfully locate host devices. In such a situation an attacker could then use that information to leverage other methods of attack against the discovered hosts.

One potential security advantage of IPv6 is the use of random allocation of addresses in a /64 address range. In order to discover all addresses in use, it would take 584,555 years to sequentially scan every IP address in this range at a rate of 1,000,000 packets per second, a task that is not practical or viable in a useful timeframe. Although this could be classified as security through obscurity, it could also be argued that it would form part of a defence in depth strategy, as it would create a delay for an attacker (Coole et al., 2012). One key finding of this research is that most organisations are allocating IP addresses sequentially starting at or about the first address in their allocation. This result would appear to indicate that most users of IPv6 addresses have not considered the use of random allocation of IPv6 addresses as part of a defence in depth strategy. This may be due to a number of factors including the general security ethos of not relying on security through obscurity as a control mechanism. Conversely, this finding would also indicate that it would still be temporally viable to enumerate in IP address range of an organisation using sequential methods. Argument can be made that the resources identified during this study relate to public services, and therefore are intended to be locatable. The counter argument to this notion is that there is no net benefit to sequentially allocating addresses, provided rudimentary address management systems are in place. If an organisation uses DHCPv6 to allocate random addresses, it can still populate DNS with those semi-permanent addresses. Then if a resource is taken out of the public domain, a new randomly generated address can replace it.

Additionally, the top 25 observed IIDs are of particular interest. As can be observed in Figure 4, the top two IIDs are incrementally allocated (being integer values '1' and '2'). This is not unusual since it is a natural starting place for administrators to allocate addresses to devices on a network, as it provides for simple memorability of addresses as opposed to when using stochastic means. What is of interest, however, is that the third and fourth place highest frequency addresses are not low-range incremental addresses, and would not be likely candidates for manual address allocation. After analysing the results further, it appears that entities are using registering many domain names to single IPv6[7]addresses in DNS. This is quite peculiar since it is contrary to recommended practices for IPv6, which has enough IPv6 addresses to independently and uniquely address each domain or subdomain. One possible explanation for this observation is that these addresses may be allocated to systems that provide load-balancing capabilities to content delivery networks.

Of interest are the countries that have the greatest use of IPv6 addresses (see Figure 6). It is possible that these countries have moved to IPv6 addressing due to an exhaustion of the IPv4 address space it these regions. This possible increased utilisation is based on the observation that Germany for example is a significant provider of server infrastructure. This is supported by data from the Alexa top one million websites, with Germany being the second largest host of servers. The USA, however, hosts six times the number of servers, which does not correlate with the number of IPv6 addresses as presented in this study.

CONCLUSION

This research presented the results of a survey into the adoption of IPv6 addressing by interrogating AAAA records and public DNS. The finding that the majority of IPv6 addresses are in the first block of 0 to ($2^{32} - 1$) would seem to indicate that established practices of allocating addresses sequentially from address ::1 onward is being followed. Given that the recommendations from Deering & Hinden (1998) are to issue /64 subnetworks, there is the potential to obscure the address of certain servers by allocating them an IPv6 address in a part of the range which would take a considerable amount of time to scan. This would not entirely reduce the risk from the threat of address scanning, but it would serve as valid delay mechanism in a defence in depth strategy. The extensive use of IPv6 in Hong Kong and Europe, with Germany in particular being prominent, would support the notion that the IPv4 range is exhausted in these regions. This finding was further supported by the fact that these regions, whilst significant, are not the greatest server hosts: the US hosts a significantly greater number of servers, yet has a smaller use of IPv6 addresses.

FURTHER WORK

This research forms a part of ongoing research into IPv6 host enumeration strategies. The data gathered during these surveys has been used to influence the development of algorithms that may be used to discover IPv6 nodes on a network. While this survey served as introductory data gathering exercise, it could have been expanded. Future work will make usage of larger subdomain permutations to query in order to widen the search scope and gather more information about the usage habits, and IPv6-enabled subdomains.

Acknowledgements

This research was conducted under sponsorship from Cisco. The survey would not have been possible without the use of supplied Cisco networking equipment and support from Cisco.

REFERENCES

Akamai Technologies (2014). Akamai ipv6 traffic volume.

Alexa Internet, Inc. (2014a). The top 500 sites on the web.

Alexa Internet, Inc. (2014b). Top 1,000,000 sites (updated daily).

Australia.gov.au (2014). A to z list of government sites.

Carpene, C., Johnstone, M., & Woodward, A. (2014). The effectiveness of classification algorithms on ipv6 iid construction. *International Journal of Autonomous and Adaptive Communications Systems*, *Vol. 7*(4 [IN PRESS]).

Carpene, C. & Woodward, A. (2012). Exposing potential privacy issues with ipv6 address construction.

Coole, M., Corkill, J., & Woodward, A. (2012). Defence in depth, protection in depth and security in depth: a comparative analysis towards a common usage language. Das, K. (2008). Medical industry using ipv6 biosensors.

Deering, S. & Hinden, R. (1998). Internet Protocol, Version 6 (IPv6) Specification. RFC 2460 (Draft Standard). Updated by RFCs 5095, 5722, 5871, 6437, 6564, 6935, 6946, 7045, 7112.

Dunlop, M., Groat, S., Marchany, R., & Tront, J. (2011a). The good, the bad, the ipv6. In *Communication Networks and Services Research Conference (CNSR), 2011 Ninth Annual*, (pp. 77–84).

Dunlop, M., Groat, S., Marchany, R., & Tront, J. (2011b). Ipv6: Nowhere to run, nowhere to hide. In *44th Hawaii International Conference on System Sciences (HICSS), 2011*, (pp. 1–10)., Hawaii. Giusto, D., Lera, A., Morabito, G., & Atzori, L. (2010). *The Internet of Things*. Springer.

R. Hinden, Deering. S. & Nordmark, E. (2003). Ipv6 global unicast address format. RFC 3587, Network Working Group.

Google (2014). Ipv6 statistics.

IEEE Standards Association (n.d.). Guidelines for 64-bit global identifier (eui-64) registration authority.

Technical report, IEEE.

Jara, A., Zamora, M., & Skarmeta, A. (2010). An architecture based on internet of things to support mobility and security in medical environments. In *Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE*, (pp. 1–5).

MaxMind, Inc. (2014). Geolite ipv6 country database.

McKinney, W. (2010). Data structures for statistical computing in python. In van der Walt, S. & Millman, J. (Eds.), Proceedings of the 9th Python in Science Conference, (pp. 51 – 56).

McKinney, W. (2012). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc.

Turiel, A. (2011). Ipv6: new technology, new threats. *Network Security*, *2011*(8), 13–15.

van Pelt, P. (2014). Ipv6 toy gallery. xslidian & VersusClyne (2014). ipv6-hosts.