Characterizing Dark DNS Behavior

Jon Oberheide[1], Manish Karir[2], and Z. Morley Mao[1]

[1] Electrical Engineering and Computer Science
University of Michigan, Ann Arbor MI 48105
{jonojono,zmao}@umich.edu
[2] Networking R&D
Merit Network Inc, Ann Arbor MI 48105
mkarir@merit.edu

Abstract. Security researchers and network operators increasingly rely on information gathered from honeypots and sensors deployed on darknets, or unused address space, for attack detection. While the attack traffic gleaned from such deployments has been thoroughly scrutinized, little attention has been paid to DNS queries targeting these addresses. In this paper, we introduce the concept of *dark DNS*, the DNS queries associated with darknet addresses, and characterize the data collected from a large operational network by our dark DNS sensor. We discuss the implications of sensor evasion via DNS reconnaissance and emphasize the importance of reverse DNS authority when deploying darknet sensors to prevent attackers from easily evading monitored darknets. Finally, we present *honeydns*, a tool that complements existing network sensors and low-interaction honeypots by providing simple DNS services.

Keywords: DNS, reconnaissance, honeypots, sensors, darknets.

1 Introduction

The emergence of sophisticated malware has led security researchers to develop innovative tools to study and combat its malicious activities. Honeypots, intrusion detection systems, and a multitude of other host and network based sensors have aided researchers extensively in their endeavors. These sensors provide a wide range of functionality, from simply responding to network requests, to emulating vulnerable services and operating systems, all the way to simulating entire virtual network and host topologies. Security researchers and network operators commonly deploy honeypots and other sensors on dark, or unused, address space to gather malware, analyze new exploit techniques, and study long-term attack trends.

To maintain their usefulness, it is vital that these sensors be resistant to remote identification and fingerprinting techniques that attackers may employ. As the arms race between malware authors and researchers continues, increasingly sophisticated attacks can utilize evasion techniques in order to avoid detection and identification. By performing reconnaissance to map valuable targets, an attacker can build specific target lists. Additionally, reconnaissance can identify

B. M. Hämmerli and R. Sommer (Eds.): DIMVA 2007, LNCS 4579, pp. 140–156, 2007.

monitoring systems that should be avoided. The use of DNS identifying potential targets is well known. In this paper we describe how the lack of appropriate DNS support for sensor address space can be used by malware to identify darknet monitoring systems. Such DNS reconnaissance utilizes PTR record DNS queries which attempt to resolve an IP address to a hostname. While current honeypots and darknet sensors are effective at analyzing traffic targeted *at* their addresses, they fail to consider out-of-band probes inquiring *about* their addresses, namely DNS queries. We appropriately label these queries as *dark DNS*. Dark DNS queries are not received by the darknet sensors themselves; instead, they are directed at the DNS nameserver that is authoritative for the darknet. Such dark DNS traffic is due to one of these reasons: DNS mapping efforts, backscatter, misconfiguration, or malicious reconnaissance.

In this paper, we measure and characterize dark DNS activity. We obtained DNS authority over two class B (/16) darknets and were able to direct this dark DNS traffic to our sensor for collection. We measured three distinct datasets, each collecting a week's worth of data: one from passively monitoring the incoming queries, one from actively responding with a NXDOMAIN (non-existent domain) error code, and one from actively responding with a valid hostname IP address queries. We present the behavioral patterns of dark DNS via these measurements and provide insight into the origin of such anomalous traffic.

Reverse DNS probing can be an effective technique for evading darknet monitors. Due to the recursive nature of queries and the hierarchical operation of DNS, an attacker can perform reconnaissance on a target network without sending any probe traffic directly to that network and without revealing the attacker's source. These characteristics make DNS reconnaissance a lucrative feature for sophisticated malware and a viable method for evasion. We show how this technique can be used maliciously to evade several large-scale darknet monitoring systems while still maintaining effectiveness against live hosts.

In order to mitigate the threat posed by this evasive technique, we discuss the proper methodology for delegating reverse DNS for darknet sensor deployments. We also present a defensive countermeasure, designed to complement current honeypot systems, to prevent sensor evasion based on DNS reconnaissance. Our tool, *honeydns*, implements a lightweight DNS responder which is able to reply to PTR queries for large darknets with appropriate records. Honeydns can be easily used to complement and properly configure a large-scale honeypot deployment.

To summarize, our work has the following contributions. We present the first detailed study to characterize and illustrate a significant amount of dark DNS traffic consisting of over 1.48M queries over three weeks. A significant portion of the dark DNS traffic is observed to originate from DNS mapping efforts by Akamai. Aside from dark DNS analysis, our work is the first to highlight the importance of properly configuring darknet DNS servers to prevent the use of PTR reconnaissance as an effective evasion vector. Our honeydns tool provides a lightweight and flexible way to build a more complete darknet sensor by facilitating the ability to correctly respond to DNS queries for a particular darknet and making it appear to contain valid hosts.

The rest of this paper is organized as follows: In Section 2, we provide an introduction to the operation of the Domain Name System and describe common queries types and response codes. Next, in Section 3, we discuss recent related work. In Section 4, we discuss our collection and experimentation setup. We then present a thorough analysis of our experiment results in Section 5. In Section 6, we discuss the implications of our results, and finally, in Section 7, we summarize our contributions and describe some future work.

2 Domain Name System

The Domain Name System (DNS), defined in RFCs 1034 [1] and 1035 [2], is a hierarchical, distributed database which provides essential name-resolution services to Internet applications. In order to perform a DNS query, a resolver will traverse the DNS hierarchy to locate the appropriate authoritative server that can answer its query. Given the address of a root nameserver, which resolvers are typically seeded with, the resolver can query for the address of the next level authoritative nameserver. By recursively performing this process through the hierarchical tree, the resolver will eventually reach the nameserver that is authoritative for the specified query. Once that server is identified, the answer to the query is retrieved by the resolver, completing its query. The DNS infrastructure supports many different query types, of which Address (A) and Pointer (PTR) are the most common.

Address (A) Records. Address record lookups perform the translation from a hostname to an IP address and are the most common DNS query performed. When a user connects to a service which is referred to by a domain name, a DNS query is performed to determine the endpoint IP address to connect to.

Pointer (PTR) Records. PTR records provide the reverse translation of A records by mapping an IP address to a hostname. The lookup is performed by transforming the queried IP address into a special, yet legitimate, domain name. For example, the domain name formed for a query for the IP address aa.bb.cc.dd is dd.cc.bb.aa.in-addr.arpa. The ".arpa" portion is a special top-level domain created specifically for these reverse PTR queries. A PTR query operates in the same manner as an A query by starting at the root and traversing the DNS hierarchy. Once the authoritative zone is reached, the authoritative server will return the hostname associated with the queried IP address. PTR queries are commonly used by network services such as SSH and SMTP to validate connecting clients.

Query Responses. For A record query, the DNS server responds with the appropriate IP address and for a PTR query the server responds with the hostname associated with the query IP address. Additionally, there are numerous status codes that are returned for a DNS query, three of the most common ones are SERVFAIL, NXDOMAIN, and NOERROR. If a resolver is able to determine the address of an authoritative server, but the server is not responding, the resolver will return SERVFAIL. On the other hand, if the authoritative server is

responding but does not possess any records that correspond to the query, it will return NXDOMAIN. If all goes well and the authoritative server is able to answer the resolver, the NOERROR code will be sent. If a darknet sensor is configured without the accompanying DNS configuration changes, PTR queries for the darknet would result in NXDOMAIN replies being generated.

3 Related Work

Our work is related to several areas of previous studies on DNS and darknets which we briefly describe here.

DNS has long been a favorite target for attackers due to its critical role in the Internet infrastructure and the inherent lack of security in the operation of DNS. One of the earliest uses of DNS for malicious attacks was described in 1990 and demonstrated how attackers could utilize a weakness in DNS lookups to subvert system security [3]. Given that the purpose of DNS is provide information about hosts, it is not surprising that it could be used for attack reconnaissance. A common way to perform such probing was the use of a zone transfer [4] to obtain the entire set of hosts that a server is authoritative for. Many administrators subsequently started to secure their servers with proper access controls. As zone transfers are rarely effective anymore, attackers have turned to tools such as TXDNS [5] to map the namespace of a domain strictly through brute-force A record queries using a dictionary.

While network and host based intrusion detection have been studied extensively, attack detection by monitoring DNS is just starting to get the attention it deserves. Malware employing spamming services can easily be detected by their emission of a large number of MX queries [6]. Botnet activity can be inferred via DNS query patterns and maintained blacklists [7,8]. Correlation of DNS activity with regular IP traffic can even be used to detect malware scanning and zero-day worm outbreaks [9]. While these techniques have focused on detecting malicious activity by observing DNS queries, in this paper we focus on characterizing DNS activity for unused portions of the IP address space.

Given the very purpose of monitoring darknet is to detect malicious activities, it is critical to avoid revealing the location of the darknet sensors to prevent evasion. Previous work demonstrated the ease of detecting the location of general network sensors [10,11] through active probing. Recent work by Rajab *et al.* [12] describes how evasive techniques can be used by malware to detect honeypots by selective sampling of IP address space. Discrepancy in the behavior in responding to incoming probes can also be exploited for sensor detection, as shown by a honeyd scanner called Winnie [13]. Our proposed honeydns tool provides the darknet sensors with higher resistance to discovery and complements other tools for configuring darknets such as [14].

Our work builds on previous DNS characterization work of both DNS root servers [15,16,17] as well as local resolvers [18]. DNS can be used to estimate network distance between hosts by exploiting a large number of open-resolver DNS servers [19]. Similar to the AS112 Project, which uses separate servers to

answer PTR queries for RFC1918, dynamic DNS updates and other ambiguous addresses, our work focuses on measuring DNS behavior in address spaces with no legitimate live hosts.

4 Methodology

For our experiments we obtained two class B (/16) darknet address blocks and delegated DNS authority for these subnets to our dark DNS collector. We then proceeded to gather three datasets for our experiments, collecting a week of DNS traffic for each dataset. For the first dataset we simply passively recorded all incoming queries to our delegated subnets without any active responses. The goal of this experiment was to obtain an accurate measure of DNS activity for these subnets without any external influence. The second dataset was obtained by repeating the first experiment, but, instead of passively monitoring, replying to the queries with the NXDOMAIN (non-existent domain) error code. NXDOMAIN is the error code usually received when no resource record is found for a query. The third dataset was obtained by replying to incoming PTR queries with a valid hostname response. The format used for this hostname was *host-*{*a-b-c-d*}*.merit.edu* in response to PTR queries for any IP *a.b.c.d* within our darknet. The DNS time-to-live of the responses was set to zero to ensure resolvers would not cache our response. The goal of collecting these three distinct datasets is to examine DNS probe traffic under three common scenarios.

Table 1. The Measurement Datasets

Response Type	A/16	B/16	A/16+B/16	Duration
No Response	NR_A	NR_B	NR_TOTAL	7 days
NXDOMAIN Response	NX_A	NX_B	NX_TOTAL	7 days
Valid Response	VR_A	-	VR_TOTAL	7 days

Table 1 shows our three datasets and the terminology we will use to refer to them throughout the rest of the paper. The first dataset where the DNS server sent out no responses to any queries are called NR_A and NR_B (No Response) respectively for the two /16 darknets A and B. Similarly, the second one where the DNS server responded with NXDOMAIN replies are called NX_A and NX_B. The third dataset, where we replied with valid responses to PTR queries, is called VR_A (Valid Response). Due to an administrative issue, response dataset VR_B was not captured. Each of the three datasets represents 1 week of data collection. We refer to the combined data from both subnets if available as NR_TOTAL, NX_TOTAL, and VR_TOTAL.

During collection periods, our dark DNS sensor archived each incoming query in a SQLite database backend for subsequent analysis. An extensive schema was used to capture various aspects of each DNS query. The information collected includes IP layer details such as the source IP, identification number, and

time-to-live (TTL) value, transport layer details such as the source port, and DNS details such as the type, id, and query.

For the second part of our study, we obtained one day of NetFlow data from a regional ISP to help identify the feasibility of using PTR scanning to detect live hosts on the Internet. We extracted only IP addresses from the NetFlow data where the TCP ACK flag was set. This ensures that SYN scanning or spoofing does not influence our results. For each of these addresses we performed a query to determine whether that particular live host had an associated PTR entry.

5 Data Analysis

Next we describe our analysis of the three datasets illustrating the presence of potentially malicious DNS activities of the monitored darknets, as aside from DNS mapping there should not be any legitimate DNS traffic for such address space. We expect the darknet DNS traffic to be caused by one of these reasons: (1) DNS mapping efforts such as that by Internet Systems Consortium [20], (2) Backscatter [21] due to spoofed darknet traffic triggering subsequent DNS queries by monitoring systems, (3) misconfiguration, (4) PTR reconnaissance by attackers to identify live hosts for attack targeting.

5.1 Basic Statistics

Table 2 illustrates the basic statistics of our three datasets. As described earlier, our first dataset is designed to gather the raw queries that are associated with the addresses of our delegated darknets. The second dataset shows the continued query activity despite correct NXDOMAIN responses. By comparing these two datasets we can establish the primary characteristics of dark DNS queries. We observe an order of magnitude more queries for the first dataset, which we believe is due to query timeout retries. Interestingly, the unique target probed in the first dataset is more than 88% of all the addresses covered by the two /16s monitored, indicating the behavior of PTR scanning for these two address blocks.

Table 2. Basic dark DNS query statistics. Query rates are per 5 minute interval.

Dataset	Queries	Unique Sources	Unique Targets	Avg Query Rate	Max Query Rate
NR_A	714K	11.2K	64.1K	353.70	5501
NR_B	606K	11.8K	52.2K	300.34	2725
NR_Total	1.32M	17.0K	116K	654.8	5553
NX_A	57K	8.59K	28.9K	27.56	552
NX_B	58K	9.09K	29.4K	28.79	560
NX_Total	115K	13.1K	58.4K	57.1	825
VR_A	45K	7.45K	24.2K	22.35	321
VR_B	-	-	-	-	-
VR_Total	45K	7.45K	24.2K	22.35	321

Table 3. Query Type Distribution for NX_Total

Query Code	Query Type	Count	Percentage
1	A	81	0.0704%
6	SOA	683	0.5937%
12	PTR	114214	99.2846%
15	MX	4	0.0035%
33	SRV	32	0.0278%
255	ANY	23	0.0199%

For the third dataset, in which our DNS responder replies with *host-{a-b-c-d}.merit.edu* in response to PTR queries for any IP *a.b.c.d* within our darknet, we observe slightly lower query rate compared to the second setting with NXDOMAIN responses. We conjecture this can be due to resolvers being satisfied with replies likely to indicate legitimate hosts and therefore stop probing early rather than continuing its scanning activity.

Over the course of our three-week experiment, our darknet DNS sensor received over 1.48 million queries. They originate from more than 8000 IP prefixes and 3900 ASes. Table 3 shows the distribution of various query type codes observed in the incoming DNS queries for the second dataset. We observed similar distribution for the other two datasets. The vast majority of these as expected are PTR queries though we do observe an occasional A record request and even a few MX queries.

5.2 Query Rate

Figure 1 presents the query rate that is observed via our dark DNS monitor for the A/16 subnet. The figure shows the number of queries received in 30 minute intervals for the NR_A, NX_A, and VR_A datasets during the course of our experiments. It shows a fairly high rate of queries over the one week measurement time period for all three datasets. There are two distinct bands visible in the data. The lower band represents query rates observed in the NX_A and VR_A datasets. These are significantly lower than the query rates observed in the NR_A dataset. We believe that the higher rates observed for the NR_A dataset are caused by servers attempting to repeatedly retrying to resolve the same IP address in the absence of any reply.

The average number of queries observed in the NX_A and VR_A datasets over a 5 minute interval is 27.5 and 22.3 respectively while the query rate for the NR_A dataset is an order of magnitude greater at 353.7. The maximum rate observed is also significantly different depending on whether our server actively responds to dark DNS queries.

Figure 1 shows a couple of interesting features as well. The first is the sporadic peaks in the query rate observed in the NX_A and VR_A datasets from the lower band. These peaks are roughly a value of 256 above the lower band, or the size of a /24 subnet, caused by a deliberate scan of that subnet. The second

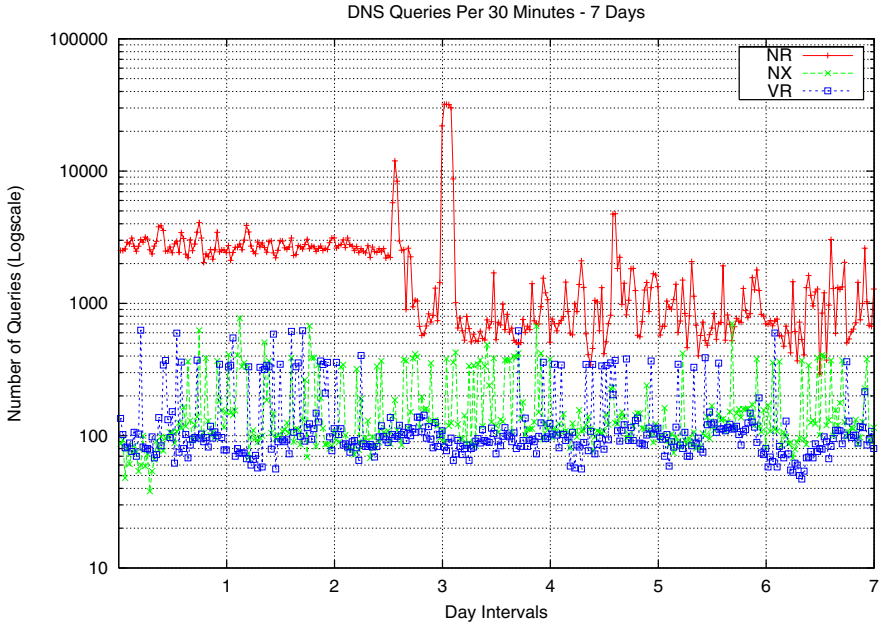

Fig. 1. Number of Queries per 30 Minute Interval - A/16

interesting feature is the significant reduction in the query rate observed in the NR_A dataset around day 4 following a short-lived spike. We will discuss this particular anomaly in greater detail in Section 5.4.

5.3 Query Targets

In order to better understand the nature of dark DNS queries, we analyzed the distribution of IP address that these queries were attempting to resolve. For this analysis we used all three datasets obtained via our collector on the A/16 subnet and for each dataset we computed the number of queries for hosts in each /24 subnet of this address space. A clustering or unusually large number of queries for a particular target address would indicate a bias in query targets. Figure 2 shows the resulting graph from our analysis. The x-axis represents each /24 subnet of our /16 and the y-axis depicts the number of queries. There are no obvious spikes or clusters visible indicating that the queries are roughly randomly distributed across the entire A/16 subnet. The NR_A query rate is clearly much higher than the query rates for the VR_A and NX_A datasets, which exhibit similar behavior.

5.4 Query Sources

One of the most intriguing characteristics of the dark DNS data is the source from which the queries originate. In this section, we discuss and illustrate several important aspects of the source of these dark DNS queries.

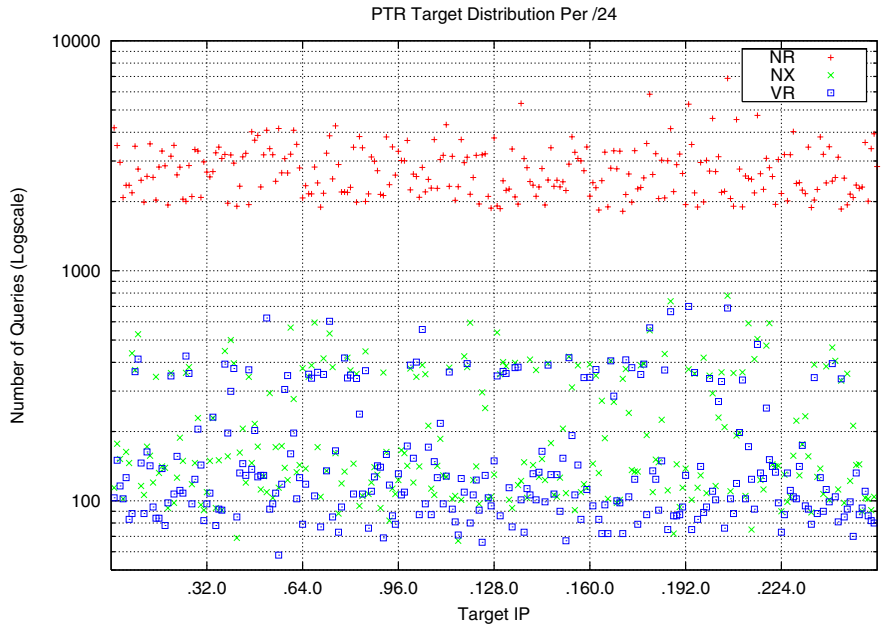

Fig. 2. Query Target IP Address Distribution - A/16

Table 4. Top 10 Sources by Percentage of Total Queries

Rank	Source IP	Percentage
1	69.15.35.X	29.0315%
2	156.45.232.X	1.2431%
3	24.93.41.X	0.5537%
4	65.24.7.X	0.4198%
5	200.169.8.X	0.4172%
6	24.92.226.X	0.4085%
7	212.27.54.X	0.3833%
8	212.27.54.X	0.3712%
9	24.25.5.X	0.3694%
10	216.219.254.X	0.3659%

Top Talkers. Table 4 lists the 10 largest contributors to our dark DNS data measurements. These measurements are based on the NX_TOTAL dataset. What is perhaps the most interesting feature of this data is that a single source IP is responsible for almost 30% of the queries. We believe that this is largely a result of large scale DNS mapping performed during our study. We discuss this characteristic further in the following sections.

Source Distribution. The left graph of Figure 3 shows the number of unique source IP addresses we observed in the NX_TOTAL dataset over time. As the

figure is of linear shape, indicating that over time we are continuing to receive queries from more unique sources that have not queried us before instead of repeated queries from the same set of hosts. This indicates that it is infeasible to block such traffic from the network via simple firewall rules. However, as this figure does not show a completely straight line, given fairly constant query rate over each day, we can conclude that a very small fraction of the total sources are in fact continuing to send queries to our dark DNS collectors over time.

The right graph of Figure 3 shows the percentage of sources as a function of the percentage of total queries. The figure shows a sharp initial increase indicating that a small percentage of the sources are contributing to a large percentage of queries in our NX_TOTAL dataset. The increasing width of the boxes indicate that an increasingly greater percentage of unique sources is needed to account for each additional 5% of the total queries.

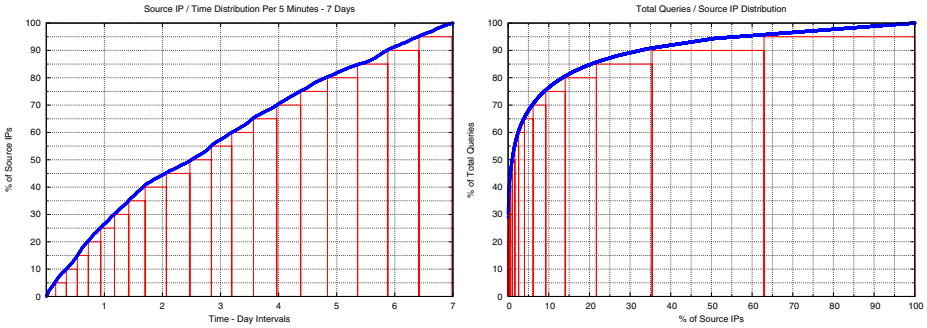

Fig. 3. Source IP distribution: growth over time, query contribution

Table 5. Top 10 Contributing Autonomous Systems by Query Volume

Rank	Query Count (% of total)	ASN	Name
1	33407 (29%)	AS17184	ATL-CBEYOND COMMUNICATIONS
2	4678 (4%)	AS7132	SBIS-AS - SBC Internet Service
3	4140 (4%)	AS12322	PROXAD AS for Proxad/Free ISP
4	2302 (2%)	AS3320	DTAG Deutsche Telekom AG
5	1438 (1%)	AS22773	CCINET-2 - Cox Communications
6	1430 (1%)	AS20170	MARITZFENTONMO - Maritz Inc.
7	1277 (1%)	AS19262	VZGNI-TRANSIT - Verizon Internt
8	903 (1%)	AS3215	AS3215 France Telecom - Orange
9	890 (1%)	AS3269	ASN-IBSNAZ TELECOM ITALIA
10	861 (1%)	AS3352	TELEFONICA-DATA-ESPANA Internet

Autonomous Systems. While there are a number of sources sending PTR queries to our dark DNS sensor, it is helpful to get a high-level view of the organizations that these IPs belong to. In Tables 5 and 6, we have ranked the top contributing organizations with their Autonomous System Number (ASN). Table 5

ranks by the total number of queries received from source IPs owned by the AS, while Table 6 ranks by the unique number of source IPs. Most of these networks are well known ISPs offering DSL and Cable modem services, along with several large ISPs such as Qwest, Deutsche Telekom, and France Telecom. It is surprising that the top query volume contributor CBEYOND accounts for more than 29% of all queries, indicating highly nonuniform source distribution of dark DNS traffic. The distribution for unique source IPs contributed by each AS is less skewed with SBC accounting for more than 4.5% of all sources observed. Also note that both SBC and Deutsche Telekom appear as the top 10 contributing ASes by query rate as well as by unique sources.

Table 6. Top 10 Contributing Autonomous Systems by Number of Unique Sources

Rank	Unique Sources (% of total)	ASN	Name
1	594 (4.5%)	AS7132	SBIS-AS - SBC Internet Service
2	268 (2.0%)	AS3320	DTAG Deutsche Telekom AG
3	214 (1.6%)	AS7018	ATT-INTERNET4 - AT&T WorldNet
4	204 (1.5%)	AS6128	CABLE-NET-1 - Cablevision Systems
5	202 (1.5%)	AS4230	Embratel Brazil
6	194 (1.4%)	AS5617	TPNET Polish Telecom commerce
7	192 (1.4%)	AS209	ASN-QWEST - Qwest
8	190 (1.4%)	AS5089	NTL NTL Group Limited
9	174 (1.3%)	AS21844	THEPLANET-AS - THE PLANET
10	164 (1.2%)	AS577	BACOM - Bell Canada

Operating Systems. Figure 4 shows the distribution of IP header TTL values from the PTR queries. The three distinct clusters signify the three classes of initial TTL values: 64, 128, and 255. As these initial TTL values result from network stack characteristics of different operating systems, we can estimate the operating system distribution of the source IPs. Linux/BSD systems set the initial TTL to 64, Windows systems set it to 128, and Solaris systems set it to 255. Table 7 summarizes the OS distribution percentages observed by our dark DNS collector.

Table 7. Query source OS distribution based on TTL

Operating System	Initial TTL	Unique Sources	Percentage
Linux/BSD	64	10480	72.93%
Windows	128	1043	7.26%
Solaris/Other	255	2846	19.81%

The vast majority of the resolvers appear to be Linux/BSD based systems, followed by a modest percentage that may be Solaris based, and finally a small percentage of Windows based resolvers. This in contrast with the resolver OS percentages reported in a previous study [17] where, of all the sources querying the F-root server, 49% were reported to be Linux/BSD based and almost 40%

were reported to be Windows based. It is clear the queries of dark DNS are not consistent with the behavior expected of a normal DNS system. It is important to note that the origins of of such queries are not necessarily end hosts, but also local resolvers querying on behalf of the end hosts via recursive DNS queries.

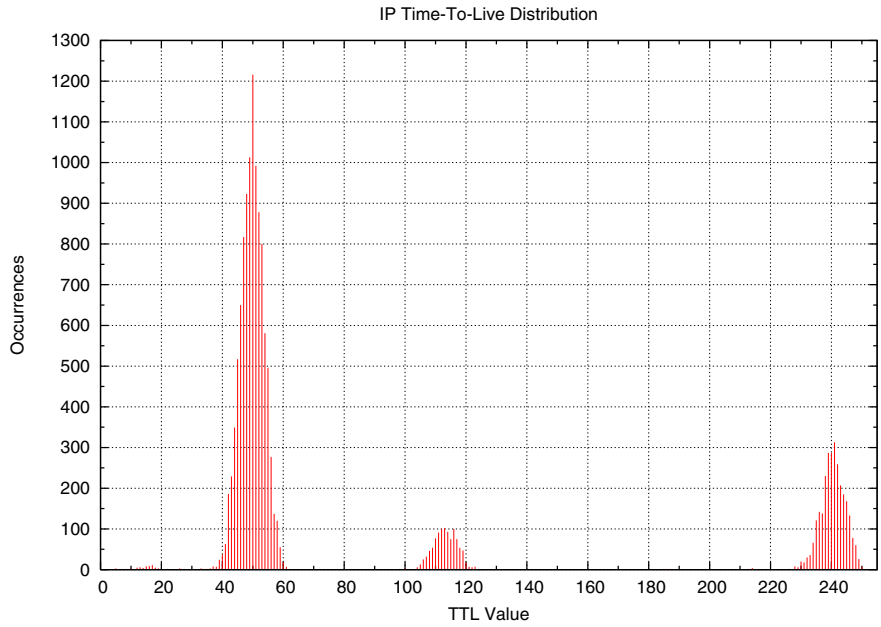

Fig. 4. TTL Distribution in PTR query packets

Akamai Mapping. Akamai is a company that provides a distributed content delivery network (CDN) to accelerate and cache web content. Their platform depends on the ability to determine network locality and distance between hosts. During our experimentation, we noticed that the majority of the top queriers were from hosts deployed by Akamai. These hosts were verified as belonging to Akamai via hostname, Internet routing registries, and SSH banner strings. Our hypothesis that Akamai is using PTR querying to supplement their network locality algorithms is partially confirmed by the DNS-based distance estimation techniques described in previous work [19].

We also determined that 11 distinct Akamai-deployed hosts are responsible for the anomalous spike in the query rate in Figure 1 of Section 5.2. Around day 3 of our NR_A dataset, an order-of-magnitude increase was observed and abruptly followed by an overall decrease in the query rate. By separating out the 11 Akamai hosts from the rest of the source hosts, we are able to more effectively highlight this anomalous behavior.

As shown in Figure 5, the query rate for the Akamai hosts is steady for the first couple days, then drops off briefly, then skyrockets up to 12000 queries

per 30 minutes, then drops off again and is not observed at all for the rest of the dataset collection. Whether this sequence of events represents a potential issue with Akamai's deployments is unknown. More importantly, separating this anomaly from the rest of our dataset demonstrates the relative consistency that all other hosts exhibit in their query rate.

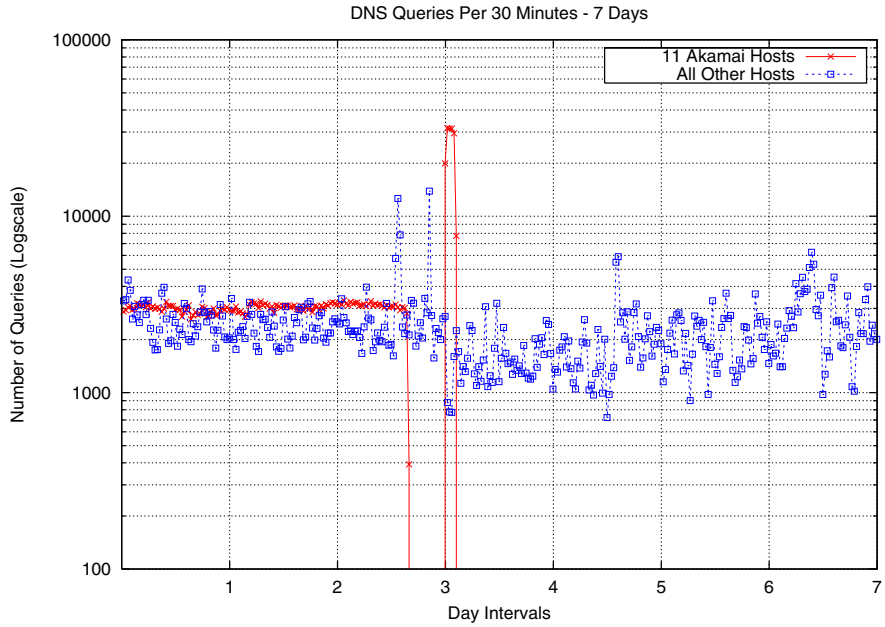

Fig. 5. Query rate distinguishing Akamai hosts from others

6 Discussion

Detailed analysis in the previous section demonstrates significant DNS activities for dark address space, originating from diverse operating systems and a large number of networks, but with a small number of hosts or networks contributing to a large percentage of queries. This data provides preliminary evidence of potentially malicious darknet DNS activities, given the presence of scanning activities and slightly lower query rate for valid response compared to the no domain response. In this section, we discuss the implications of PTR reconnaissance, verify the assumption that most live hosts have valid PTR records, and propose a solution to prevent sensor evasion using PTR reconnaissance.

6.1 PTR Reconnaissance

The PTR query type, as previously discussed, is used to perform the mapping from an IP address to a hostname. Given an address range selected for attack, an

attacker can send a PTR query for each address in the range and note the result. Operating under the assumption that IP addresses with associated PTR records are usually live hosts with potentially exploitable services, an attacker can easily determine whether certain hosts or subnets are worthy of attention. Therefore, if an attacker sees a valid PTR response with an associated hostname, he can continue to attack that target with an increased level of confidence. Otherwise, he can move on to potentially more valuable targets. It is important to note that these PTR queries will not be seen by the sensors monitoring the traffic of the dark address space, but instead by the DNS server authoritative for that address space.

More importantly, an attacker can mask his identity and source IP address while performing this reconnaissance. The DNS infrastructure and its resolvers offer functionality known as recursive querying. If a client requests a recursive query from a resolver with recursive query enabled, that resolver will perform all the necessary communication on behalf of that client and simply return the final result. The other DNS servers involved in the query will have no way of knowing the attacker's true identity as they will only be communicating with the resolver the attacker has chosen. An attacker may choose resolvers located at his local ISP or, for more anonymity, one of the many open resolvers around the Internet that accept recursive queries.

6.2 Validating Usefulness of PTR Reconnaissance

To verify our assumption that most live hosts have associated PTR records and monitored unused address space does not, we performed several measurements.

First, we wanted to determine the distribution of live hosts with PTR records. We obtained 24 hours worth of NetFlow data (of size 268MB compressed) from a large regional provider, containing TCP conversations involving a total of 1,234,842 unique IP addresses. By performing a PTR query on each of these IP addresses, we received a total of 980,835 valid responses, indicating 79.43% of the hosts have associated PTR records. This high percentage confirms our assumption and affirms the effectiveness of PTR reconnaissance.

In addition, given that home users on broadband connections are more frequently targeted by malicious activity, PTR scanning techniques would even more successful against as most ISPs assign PTR records for their addresses. Table 8 shows the PTR record format template for a number of major broadband ISPs obtained from our probes.

We were also able to use PTR reconnaissance to successfully evade several large-scale, distributed systems that monitor dark address space for malicious activity. We actively probed one of these systems, which consists of sensors installed at numerous ISPs around the world monitoring over 17 million routable IP addresses. Of all the various deployments of this sensor network, only a single class C subnet (256 addresses) was configured with reverse DNS and responded to our PTR queries. Utilizing PTR reconnaissance, an attacker would successfully evade 99.9985% of that sensor's darknet monitoring.

Table 8. Common PTR record formats (anonymized)

Organization	PTR Format
AT&T	{ID.detroit-ID-ID}.mi.dial-access.att.net
Belgacom	{A.B-C-D}.adsl-dyn.isp.belgacom.be
Bellsouth	host-{A-B-C-D}.bhm.bellsouth.net
Blueyonder	adsl-{A-B-C-D}.blueyonder.co.uk
Charter	{A-B-C-D}.dhcp.bycy.mi.charter.com
Comcast	c-{A-B-C-D}.hsd1.ma.comcast.net
Earthlink	user-{ID}.cable.earthlink.net
Qwest	{A-B-C-D}.albq.qwest.net
Roadrunner	cpe-{A-B-C-D}.carolina.res.rr.com
Rogers	{ID-ID}.cpe.net.cable.rogers.com
SBC Yahoo	adsl-{A-B-C-D}.dsl.rcsntx.sbcglobal.net
Shawcable	{ID}.vs.shawcable.net
Speakeasy	dsl{A-B-C}.sea1.dsl.speakeasy.net
Telus	d{A-B-C-D}.bchsia.telus.net
Tiscali	{A-B-C-D}.dsl.ip.tiscali.nl
Verizon	pool-{A-B-C-D}.esr.east.verizon.net
XO	{A.B.C.D}.ptr.us.xo.ne

6.3 Honeydns to Combat PTR Reconnaissance

In order to subvert the effectiveness of sensor evasion via PTR reconnaissance, a countermeasure must be deployed. The underlying approach is straightforward: a valid DNS reply must be generated when an attacker performs a PTR query for a sensor address.

Sending responses for an attacker's PTR query requires DNS authority for the targeted address space. Fortunately, as many network sensors have already been delegated permission to monitor dark IP address space, the additional requirement of gaining DNS delegation is not usually a significant technical nor administrative burden. Once DNS authority has been delegated, it becomes possible to reply to an attacker's PTR query with an arbitrary hostname that appears reasonable for a live host.

While such responses can be provided by existing DNS software packages, it is desirable to deploy a solution that decreases deployment complexity and increases functionality and flexibility. DNS servers such as BIND can be cumbersome to configure and deploy as an authoritative server, especially when only a small subset of DNS functionality is required. In addition, many sensor deployments employ sampling and dynamic topologies which require a flexible framework that static configuration files cannot provide.

We kept these design goals in mind when implementing *honeydns*, a simple yet flexible daemon providing PTR response functionality. Honeydns is written in Python and contains less than 200 lines of code. By providing a flexible response framework, honeydns complements the needs of any low-interaction honeypot deployment. Honeydns also provides passive monitoring capabilities

to detect and alert an operator when a malicious attacker is employing PTR reconnaissance techniques.

7 Conclusions and Future Work

Our work is the first detailed study to characterize DNS queries of darknet address space, known as dark DNS. We observe a significant amount of DNS queries to these darknets which are likely due to DNS mapping (*e.g.,* by Akamai), backscatter traffic, misconfiguration, and PTR reconnaissance by attackers. Our work is the first to describe the importance of properly configuring DNS authority for darknet address space to reduce the possibility of sensor evasion. Towards this goal, we develop a lightweight tool called honeydns to provide flexible PTR response functionality in addition to passive DNS traffic anomaly detection capability. As future work, we plan to correlate observed dark DNS traffic with data traffic to the associated darknets to further validate the presence of DNS reconnaissance.

References

1. Mockapetris, P.: RFC 1034: Domain names: concepts and facilities (November 1987), `ftp://ftp.internic.net/rfc/rfc1034.txt`
2. Mockapetris, P.: RFC 1035: Domain names: implementation and specification (November 1987), `ftp://ftp.internic.net/rfc/rfc1035.txt`
3. Bellovin, S.: Using the domain name system for system break-ins. In: Proceedings of the 5th USENIX UNIX Security Symposium (1995)
4. Samwalla, R., Sharma, R., Keshav, S.: Discovering Internet Topology. Unpublished manuscript
5. Silveira, A.: TXDNS: an aggressive multithreaded DNS digger, `http://www.txdns.net/.`
6. Ishibashi, K., Toyono, T., Toyama, K., Ishino, M.: Detecting mass-mailing worm infected hosts by mining DNS traffic data. In: Proceedings of the Special Interest Group on Data Communications (SIGCOMM) (2005)
7. Kristoff, J.: Botnets, detection and mitigation: DNS-based techniques. NU Security Day (2005)
8. Schonewille, A., van Helmond, D.-J.: The Domain Name Service as an IDS: How DNS can be used for detecting and monitoring badware in a network (February 2006), `http://staff.science.uva.nl/delaat/snb-2005-2006/p12/report.pdf`
9. Whyte, D., Kranakis, E., Van Oorschot, P.: DNS-based Detection of Scanning Worms in an Enterprise Network. In: Proceedings of the Network and Distributed Systems Symposium (NDSS) (2005)
10. Bethencourt, J., Franklin, J., Vernon, M.: Mapping Internet Sensors with Probe Response Attacks. In: Proceedings of Usenix Security Symposium (2005)
11. Shinoda, Y., Ikai, K., Itoh, M.: Vulnerabilities of Passive Internet Threat Monitors. In: Proceedings of Usenix Security Symposium (2005)
12. Rajab, M., Monrose, F., Terzis, A.: Fast and Evasive Attacks: Highlighting the Challenges Ahead. Proceedings of the 9th International Symposium on Recent Advances in Intrusion Detection (RAID) (September 2006)

13. Oberheide, J., Karir, M.: Honeyd Detection via Packet Fragmentation. Technical report, Merit Networks Inc. (2006)
14. Sinha, S., Bailey, M., Jahanian, F.: Shedding Light on the Configuration of Dark Addresses. In: Proceedings of NDSS (2007)
15. Brownlee, N.: DNS Root/gTLD Performance Measurements. IETF Meeting (2001), http://www.caida.org/publications/presentations/ietf0112/
16. Nemeth, E.: DNS Damage - Measurements at a Root Server. IETF Meeting (2001), http://www.caida.org/publications/presentations/ietf0112/
17. Wessels, D., Fomenkov, M.: Wow, That's a Lot of Packets. In: Proceedings of Passive and Active Measurement Workshop (September 2003)
18. Jung, J., Sit, E., Balakrishnan, H., Morris, R.: DNS Performance and the Effectiveness of Caching. In: Proc. ACM SIGCOMM Internet Measurement Workshop (2001)
19. Gummadi, K.P., Saroiu, S., Gribble, S.D.: King: Estimating Latency between Arbitrary Internet End Hosts. In: Proceedings of SIGCOMM IMW (2002)
20. Internet Systems Consortium. ISC Internet Domain Survey Background (2006), http://www.isc.org/index.pl
21. Moore, D., Voelker, G., Savage, S.: Inferring Internet Denial of Service Activity. In: Proceedings of the 2001 USENIX Security Symposium (2001)