Advancing in Reverse: A Comprehensive Characterization of IN-ADDR.ARPA Deployment

Alfred Arouna
alfred@simula.no
SimulaMet / OsloMet

Ioana Livadariu
ioana@simula.no
SimulaMet

Roland van Rijswijk-Deij
r.m.vanrijswijk@utwente.nl
University of Twente

Mattijs Jonker
m.jonker@utwente.nl
University of Twente

ABSTRACT

Reverse DNS (rDNS) is often used as a reliable data-source for critical services, such as mail, security appliances or geolocation services. Unlike forward DNS, rDNS remains understudied, especially from a deployment perspective. In this paper, we take steps towards closing the gap, starting at regional Internet registries, down to network operators in the lower hierarchy. To this end, we use public and complementary data sources and find that around 40% of allocated IPv4 address space has well-configured rDNS entries. We highlight regional differences as rDNS deployment is driven by mail and infrastructure providers in the developed world, while national Internet registries and national ISPs are drivers in the developing world. We study the use of classless delegation and the prevalence of configuration errors breaking DNS resolution. Finally, we observe that multi-regional organizations such as CDNs and mail providers actively invest effort towards improving rDNS deployment.

CCS CONCEPTS

• **Networks → Naming and addressing**.

KEYWORDS

rDNS, PTR, RIR

ACM Reference Format:
Alfred Arouna, Ioana Livadariu, Roland van Rijswijk-Deij, and Mattijs Jonker. 2023. Advancing in Reverse: A Comprehensive Characterization of IN-ADDR.ARPA Deployment. In *Asian Internet Engineering Conference (AINTEC '23), December 12–14, 2023, Hanoi, Vietnam.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3630590.3630595

1 INTRODUCTION

The distributed Domain Name System (DNS) is organized among Top Level Domains (TLDs) below a `root`. In addition to well-known TLDs, the DNS also comprises: reserved zones such as `.example` (RFC2606 [1]); special-use names such as `.test` (RFC6761 [8]); and infrastructure zones such as `.arpa` (RFC3172 [20]). Contrary to *forward* DNS (fDNS), *reverse* DNS (rDNS) relies on a single TLD (`.arpa`), but follows the same resolution process. Hostnames can be assigned to IP addresses by publishing pointer (`PTR`) records in rDNS. Much like fDNS, rDNS is subject to DNS risks, including the threat of abuse, misconfigurations and outages. For instance, RIPE NCC's rDNS service experienced issues for about a day in 2012 [22]. Due to the DNS caching mechanisms, the impact was limited, but the root cause of this incident was not found.[1] This illustrates the lack of systematic studies or thorough investigations on rDNS, even though rDNS data is widely used for security mechanisms, service discovery, troubleshooting, logging, geolocation, topology discovery as well as routing naming [16]. However, abuse of these services can create a significant burden on the `.arpa` TLD name servers, impacting its availability. For example, RFC7208 recommend that the `PTR` mechanism should not be use for Sender Policy Framework (SPF) validation since it causes high load on `.arpa` servers [23]. Thus, inconsistencies or unexpected records in rDNS or unavailability of the reverse DNS can lead to misinterpretation or render threat mitigation solutions ineffective. The Common Weakness Enumeration (CWE) 350 captures a vulnerability that affected security software relying on rDNS.[2] Additionally, without hostname resolution, mail servers may consider all incoming emails as forgeries and discard them, which can have significant operational and business implications. Similarly, solutions against Business Email Compromise (BEC) may falter even though BEC has caused over $2.4 billion in losses in 2022 [14].

Despite its importance, rDNS has been studied considerably less than fDNS. To the best of our knowledge, only few previous works have reported on rDNS deployment aspects, and a characterisation of rDNS across its entire hierarchy, i.e., from Regional Internet Registry (RIR) delegations to the network operator practices, is currently missing. In this paper, we take steps towards closing this gap, and further advance the understanding of IPv4 rDNS deployment. We use active measurement data and consolidated data from passive sources such as RIR zone files to this end.
We make the following contributions:

- We study rDNS deployment, starting at the top at the RIRs, down to network operators at the bottom. We characterize configuration practices along the way and show that only 43% of the IPv4 allocated space is well-configured , i.e., responding with `NOERROR`.

AINTEC '23, December 12–14, 2023, Hanoi, Vietnam

ACM ISBN 979-8-4007-0939-5/23/12…$15.00
https://doi.org/10.1145/3630590.3630595

[1] https://labs.ripe.net/author/dfk/conclusions-drawn-from-reverse-dns-event/
[2] https://cwe.mitre.org/data/definitions/350.html

Therefore, rDNS based security tools accuracy and coverage may be limited.
- We consider regional differences in deployment and find that mail and infrastructure providers drive rDNS deployment in the developed world. This is consistent with the increasing centralisation of the Internet on these regions. At the same time, we find that National Internet Registries (NIRs) and national ISPs are the drivers in other regions of the world.
- We analyze the use of `CNAMEs` within as well as outside of the context of classless delegation (i.e., RFC2317) across the hierarchy and quantify the extent to which syntax errors are the cause of resolution errors.
- We dicuss possible incentives for organizations to deploy rDNS and find that some multi-regional organizations consistently strive for full and functional rDNS configuration.

The remainder of this paper is structured as follows. We provide background information on IPv4 address allocation and rDNS delegation in Section 2. Section 3 presents the related work. In Section 4, we introduce our data sources, and present our results in Sections 5, 6 and 7. Finally, we conclude our paper in Section 8 and also outline future work.

2 BACKGROUND

2.1 IPv4 Address Space Delegation

The Internet Assigned Numbers Authority (IANA) manages the allocation of IPv4 addresses and maintains a list with the status of all 256 IPv4 `/8` address blocks [21]. Prior to the inception of the regional registries, IANA was allocating `/8` blocks directly to organizations. These allocations are now referred as *legacy* allocations. Over the span of a few years, five RIRs that geographically covered the entire world were established: the Réseaux IP Européens Network Coordination Centre (RIPE NCC) in 1992; the Asia Pacific Network Information Centre (APNIC) in 1993; the American Registry for Internet Numbers (ARIN) in 1997; the Latin American and Caribbean Internet Addresses Registry (LACNIC) in 2002; and the African Network Coordination Centre (AFRINIC) in 2005. After the establishment of a regional system, the management of IP address space changed. IANA allocated `/8` blocks to RIRs, which in turn distributed this address space regionally, according to their regional policies. RIR allocations are stored in their `WHOIS` databases as well as dumps of their reverse DNS zones. As of 2023, 220 `/8` blocks are allocated and 36 are labelled as *reserved*. The reserved space is meant for non-global use, whereas allocated addresses can be used on the public Internet.

2.2 IPv4 Reverse DNS Delegation

IANA also maintains the rDNS zone `in-addr.arpa`, which delegates the `/8` IPv4 address blocks. According to RFC8499 [18], delegation relies on `NS` records in the parent zone for the children, i.e., subdomains. Simple delegation involves an `NS` record in the parent zone, and a Start Of a zone of Authority (`SOA`) record in the child zone. Reverse delegation follows the Internet addressing structure (RFC1035 [28]), i.e., names under `in-addr.arpa` have up to four labels with the `in-addr.arpa` suffix with each label representing IP address octets as decimal values. For example, L3 in Listing 1 contains the prefix `100/8` in its reversed form `100.in-addr.arpa`, and delegates authority to ARIN. Additional labels can be used to delegate on `/16` or `/24` IPv4 network boundaries. Four octets identify a specific address (see L12-16). Delegating prefixes this way brings challenges: a provider that allocates more specific IP address blocks than `/24` to its customers has to maintain the reverse DNS for these IP blocks, i.e., such customers are not able to manage their own reverse zone autonomously.

RFC2317 [13] suggests the use of `CNAME` records to delegate on non-classful boundaries for more flexibility. For example, L9 in Listing 1 delegates `164.215.39.192/26` to `ns1.shellit.org` and L10 aliases the IP address `164.215.39.193` to a label under this `/26`. Other records for this `/26` are therefore under control of the `shellit.org` name server (see L14 & L15). Similarly, an adjacent `/26` is delegated to a `ficolo.net` name server instead. In these examples, the dash (`-`) is used to delineate blocks, but RFC2317 does not restrict using specifically this character.

IANA delegates *reserved* prefixes such as `10/8` (see L2). According to RFC6305[27], unsolicited lookups for private network addresses [29] are handled by `blackhole-{1,2}.iana.org` and should return `NXDOMAIN`. IANA also allocates *multicast* prefixes to operators (see L4). Although, the `100/8` address block is delegated to ARIN, RFC6598 specifies that rDNS queries for the *shared address space addresses* i.e., `100.64.0.0/10` must not be forwarded to the global DNS infrastructure and is handle by IANA [2] (see L6). However, these are, out of scope of this paper and thus excluded from our analysis.

3 RELATED WORK

Relatively speaking, reverse DNS has received much less attention compared to forward DNS. Studies involving rDNS often retrieve information from hostnames, for example to infer geolocation [11, 12, 26], characterize router-level infrastructure [19, 39, 40], learn aspects of end-user connections [24], or to detect and combat various forms of abuse [9, 17, 31, 36]. Other works have used rDNS to seed other data sets [15].

To the best of our knowledge, very few works exist that study and characterize rDNS deployment. We trace early efforts to the 2017 RIPE DNS Measurement Hackathon, during which participants found, among others, that for a limited number of selected prefixes (100) from RIPE, fewer than 25% of addresses had hostnames assigned.[3] Phokeer *et al.* [33] looked at the African region and studied lame rDNS delegations, finding that 45% of all reverse domains in said region are lame. Van der Toorn *et al.* looked at rDNS deployment through a privacy lens, demonstrating concerning operational practices that lead to leaking highly privacy-sensitive information of network clients [38]. Fiebig *et al.* studied rDNS in terms of its utility towards Internet measurement research [16] using passively and actively collected rDNS data, covering the full IPv4 space. Overall, they showed that rDNS is not as poorly delegated as the common opinion among operators suggests. However, the authors did not consider regional differences and did not use zone files data to infer RIR-level rDNS zone delegation. Borgolte *et al.* [5] discuss how rDNS for IPv6 can be partially enumerated by relying on denial-of-existence properties of DNSSEC. Their work

[3] https://blog.apnic.net/2017/05/25/investigating-status-reverse-dns/

```
; IANA level: in-addr.arpa. zone
10.in-addr.arpa.                          86400   IN  NS      blackhole-1.iana.org.
100.in-addr.arpa.                         86400   IN  NS      r.arin.net.
224.in-addr.arpa.                         86400   IN  NS      a.iana-servers.net.
; RIRs or member level: for example 164.in-addr.arpa. zone
127.100.in-addr.arpa.                     86400   IN  NS      b.iana-servers.net.
128-191.39.215.164.in-addr.arpa.          86400   IN  NS      dns1.ficolo.net.
129.39.215.164.in-addr.arpa.              86400   IN  CNAME   129.128-191.39.215.164.in-addr.arpa.
192-255.39.215.164.in-addr.arpa.          86400   IN  NS      ns1.shellit.org.
193.39.215.164.in-addr.arpa.              86400   IN  CNAME   193.192-255.39.215.164.in-addr.arpa.
; IP PTR
129.128-191.39.215.164.in-addr.arpa.      3600    IN  PTR     ria-ge100-fc1132_315-kiv.ulv.fi.ficolo.net.
161.128-191.39.215.164.in-addr.arpa.      86400   IN  PTR     ria-ge100-fc1132_316-kiv.ulv.fi.ficolo.net.
193.192-255.39.215.164.in-addr.arpa.      10756   IN  PTR     gw-164-1.shellit.org.
194.192-255.39.215.164.in-addr.arpa.      10800   IN  PTR     gw-164-2.shellit.org.
96.97.98.99.in-addr.arpa.                 86400   IN  PTR     99-98-97-96.lightspeed.tukrga.sbcglobal.net.
```

Listing 1: Example of reverse DNS zones (snippets from IANA, RIR-level and member-level zones).

Table 1: Number of responses in daily active rDNS measurement data of the IPv4 address space (medians).

	NOERROR	NXDOMAIN	SERVFAIL	**Timeout**
PTR	1.2B			
Others	4.5M	8.3M	9M	11M

Table 2: Number of zones, subdomains (Subd.) and RR types in collected daily zone files data (medians).

	RIPE	**APNIC**	**ARIN**	**LACNIC**	**AFRINIC**	**IANA**
Zones	96	151	146	47	45	1
Subd.	57K	522K	676K	22K	37K	230
NS	51K	522K	670K	22K	37K	230
CNAME	6K		6K			
A	5		5			
SOA			92		6*	

highlights the difficulties of reliable and exhaustively enumerating the entire rDNS name space for IPv6.

We use publicly available RIR zone files as a means to reliably learn the upper part of rDNS delegation chains, which we complement with existing active rDNS measurement data to go down to network operators. Different from previous works, we study regional differences and delegation across the entire hierarchy. We also take a look at hierarchical consistency.

4 DATASETS

In this section we describe the active and passive DNS data sources that we use for this paper.

4.1 IPv4 Reverse DNS Measurements

Various organisations perform reverse DNS measurements at various levels of granularity. Two organisations in particular collect regularly reverse DNS snapshots: Rapid7 Labs[4] and OpenINTEL. Out of these two, we chose to work with OpenINTEL data for two reasons: 1) OpenINTEL data collection is more fine-grained, collecting snapshots every 24 hours, versus Rapid7 collecting data once per week, and 2) it has become challenging for researchers to gain access to Rapid7's datasets. OpenINTEL uses several heuristics, including delegation searching through SOA records with RFC8020 pruning, to only target prefixes that are actually delegated, thereby keeping NXDOMAIN collection much lower compared to approaches that forcibly query the entire IPv4 address space. Table 1 provides summary statistics for responses to PTR queries. Each measurement run takes at most a day and covers a median of over 1.6 B terminal records (43% of allocated IPv4 space), which represents an increase of about 33% compared to results from 2018 (six days of data collected by Fiebig *et al.* [16]). The *Others* category covers non-PTR response types to PTR queries: DNAME, CNAME and no reply (timeout). Note that numbers have been rounded in all tables. Since the recorded responses are consistent over time, meaning the number of resolvable hostnames is relatively stable, we consider the daily median values for all our analysis.

4.2 Reverse DNS Zone Files

We obtain the consolidated RIR-level rDNS zone files from the rir-data.org project [3]. The project collects publicly available rDNS zone files from the five RIRs as well as IANA data for `in-addr.arpa`. RIRs provide snapshots of their zones[5] and IANA-managed zones are publicly available via zone transfer from the Internet Corporation for Assigned Names and Numbers (ICANN)[6] or via Web download at InterNIC.[7] We collect longitudinal data over a nine-month period (Nov 1, 2022 — Jul 31, 2023). Table 2 lists the median number of zone files and subdomains (child zones) per day over our measurement period. As expected, the data for IANA is limited to one zone (i.e., `in-addr.arpa`) per day, whereas RIRs have multiple zones daily (e.g., allocated `/8` separately). However, some `/8` blocks such as `7/8` are not delegated in IANA's `in-addr.arpa`, thus for IANA we collect on median 230 domains. Analysing the RIR data, we observe that the highest number of zones and subdomains are within the ARIN and APNIC regions, followed by RIPE. This is consistent with the fact that ARIN, APNIC and RIPE are the 1st, 2nd and 3rd largest RIRs in terms of IPv4 prefixes allocation by IANA [21].

[4] https://opendata.rapid7.com/sonar.rdns_v2/

[5] http://ftp.[rir].net/pub/zones/

[6] https://www.dns.icann.org/services/axfr/

[7] https://www.internic.net/domain/

4.3 WHOIS and RIR Delegation data

RIRs make information regarding resources allocated to registrants – such as IP address blocks and Autonomous System Numbers (ASNs) – available through WHOIS [10]. Such data can be obtained in bulk from the registries. However, the WHOIS records format is not consistent across RIRs and at times not even within one region. Furthermore, these records can contain incomplete data. The consolidated RIR-level WHOIS inetnum objects from the rir-data.org project [3] addresses these challenges. Therefore, we leverage this dataset in our analysis. Published daily, these files contain allocated addresses and ASNes, along with country and allocation data. With this consolidated IPv4 address allocation data, we can map each rDNS query from OpenINTEL to a specific prefix and RIR. Furthermore, we link operator activity to organization and/or country, and compare rDNS from RIR and network operator perspectives.

5 RECORDS AND RESPONSE TYPES

We further assess the commonality of different types of rDNS resources and responses and follow delegations from RIRs down to the network operator level.

5.1 Regional Internet Registries

Using the rDNS data, we first analyse the deployment at the RIR level and find that rDNS prefixes largely follow octet boundaries. Table 2 shows resource record types and usage at the IANA and RIR level. The number of zones and domains per RIR appears to follow IPv4 address allocation. ARIN and APNIC dominate rDNS delegations in parallel with their shares of allocations in the IPv4 address space. We expect to observe NS and CNAME records for classful and classless delegation. However, in addition to these record types, we also observe marginal use of SOA and A records as well. Surprisingly, we find that with the exception of ARIN, most RIRs do not publish SOA records. This leads us to believe that RIRs filter them from the zone snapshots.[8] Moreover, our analysis revealed that AFRINIC published SOA records for all six /8 allocated to them, for a period of only five consecutive days in the middle of the nine-month period of our study. This brief appearance is reflected in Table 2 with the sign * attached to the six /8 SOA records. Although the root cause is unknown to us, the brief appearance is peculiar at the very least, and perhaps an operational mistake. Analyzing the A records, we observe consistent use by RIPE and ARIN, for addresses covered by two /24 prefixes. We speculate that these records were either created before CNAME adoption for classless delegation, or that they are part of an experimental test-bed. The latter hypothesis is supported by a common naming pattern in the subdomains.

Figure 1 shows NS and CNAME usage in RIR zone dumps, broken down by RIR and prefix length. We find that RIR-level delegations follow IPv4 address space allocation, with ARIN and APNIC having the most address space. When focusing on delegations, we notice that in the developing world, a significant part of the IPv4 address space is delegated to major operators. For instance, in AFRINIC, APNIC and LACNIC, delegations largely include NIRs, national ISPs, and infrastructure providers. Under RIPE, /24 reverse delegation largely involves a limited number of DNS providers. For classless delegation (i.e., CNAME usage), we observe frequently involvement of a limited number of CDNs, cloud, datacenter or hosting providers that operate in the developed world.[9] We hypothesis that IPv4 address exhaustion most likely also contributes to the classless delegation usage within RIPE and ARIN. With APNIC and LACNIC following NIR-based IPv4 allocation, the need for CNAMEs is minimized at the RIR-level and is expected on lower level of the hierarchy.

5.2 Network Operators

We focus further on the response types and status codes per rDNS domain down at the members per RIR. To this end, we leverage the collected WHOIS data and associate allocated prefixes with OpenINTEL PTR measurement data points. Specifically, to verify if a network operator uses CNAME records according to RFC2317, we identify, on the basis of the reverse query name, the longest prefix match and RIR.

Our results reveal relatively few errors such as timeouts, NXDOMAIN and SERVFAIL (28.3 M combined). This result is most likely due to the overall number of 1.6 B NOERROR responses in our data. The fractions of NXDOMAIN and SERVFAIL in OpenINTEL data are both 0.4%, which we attribute to the measurement's refined pruning heuristics (see Section 4.1). Figure 2 shows the median number of responses to PTR queries, broken down in type, status, and operator regions. More than 98% of all responses were successful (NOERROR) and contain PTR data. Overall, much less common are CNAME (0.4%). However, CNAME records are used by operators in all regions. They are more prevalent on the operator level than at RIRs and in some regions more common. The highest numbers of operators using classless delegation are in regions of RIRs that themselves use classless delegation. Unexpectedly, we observe DNAME [34] responses from operators located in the APNIC, LACNIC and RIPE regions. SERVFAIL responses are negligible, and we find that three parties are responsible for most of these responses. Evidently, NXDOMAIN responses are more likely to follow CNAME and DNAME answers, meaning the resolution chain is broken during expansion of these records. This can be attributed to record configurations at the network operator level which introduce resolution issues. Table 3 shows DNAME naming practices and the contribution thereof to response status, breaking down the use of alphabetical, numeric, alphanumeric (AN), and special characters (includes AN), as well as in-addr.arpa suffix presence (S.) and inadvertent repetition thereof (R.). We find no relation between naming practices and success ratio. Therefore, we relate the large number of NXDOMAIN responses following DNAME answers to absent PTR records.

Figure 3 shows that operators from the developed world use CNAME more often than operators from the developing world. We hypothesis that this is due to the IP address space relatively recent fragmentation. RIRs use CNAME records to classlessly delegate exclusively prefixes more specific than /24s (see Fig. 1). Network operators, in contrast, also delegate (allocated) prefixes less specific than /24s. This shows that CNAME records are used out of the context of RFC2317 [13]. Countries in the APNIC region are both developed and developing countries. Operators in the APNIC region use CNAME for both classless and classful delegation.

[8] Similarly, DNSSEC records are filtered out.

[9] We identify the NS records of theses PTR records and manually classify each providers using publicly available classification tools such as bgp.tools [6].

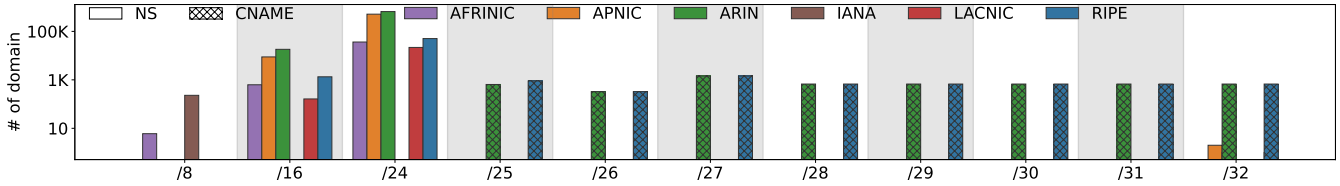

Figure 1: Number of NS and CNAME records (daily median) in daily RIR data, broken down in prefix lengths, record type and RIR (log scale). RIRs use NS records, mostly for /16 and /24 delegations on classful network boundaries. RIPE and ARIN also use CNAME records for classless delegation of prefixes more specific than /24s.

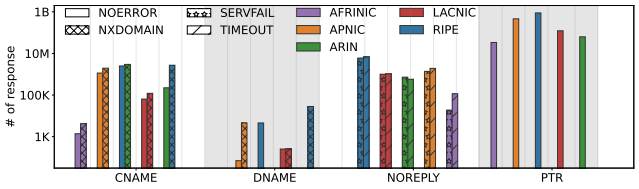

Figure 2: Daily median number of response types (log scale). CNAME records are more often followed by NXDOMAIN responses later in the resolution chain.

Table 3: DNAME naming practices per region. We use ~ for negligible (<0.1%) usage or ratio (rounded numbers).

Region	S.	R.	Format	NOERROR	NXDOMAIN
LACNIC	✗	✗	numeric		**256 (50%)**
	✓	✗	numeric	255 (49.8%)	1 (0.2%)
RIPE	✗	✗	alpha		7 (~)
	✗	✗	alpha-numeric	1 (~)	84 (0.2%)
	✗	✗	numeric		256 (0.7%)
	✗	✓	special + AN	5K (13.1%)	**28K (80.1%)**
	✓	✗	numeric	16 (~)	2K (5.8%)
APNIC	✗	✗	alpha		7 (0.2%)
	✓	✗	numeric	70 (1.5%)	**5K (98.3%)**

Implications: The large ratio of NOERROR suggest a good hygiene in rDNS deployment. Although CNAME records for large prefixes are not compliant with RFC2317, network operators can further delegate part of their IPv4 address space as classless delegation to their customers, which can result in additional CNAME records. Moreover, CNAME can minimise operational complexity, e.g., by using the same configuration template for classful and classless delegation.

6 CNAME AND NON-EXISTENCE

We showed that CNAME usage is popular and that operators rely on it outside of the RFC2317 context. Moreover, we observed that CNAME records are more often followed by NXDOMAIN responses later in the resolution chain. Thus, we now investigate the relation between its use and NXDOMAIN occurrences.

6.1 RIRs and CNAMEs

RFC2317 proposes using the special characters / and - for classless delegation. Thus, we extract any special character combinations from CNAME records collected at the RIR and operator levels. The only RIRs that use classless delegation are RIPE and ARIN. To the best of our knowledge, there is no publicly available policy on CNAME naming. Using RIR data, we find that both RIRs: a) combine numbers and exclusively the special character - [4]; b) end all CNAMEs with the `in-addr.arpa.` suffix; and c) are compliant with RFC2317. Moreover, for both RIRs we find similar absolute numbers of CNAME records. Table 4 lists these results in the last row.

6.2 Network Operators and CNAMEs

Using the approach described in the previous section, we observe, in the active measurement data, that network operators use a variety of characters in CNAME records, including alphanumeric and special characters. In addition to the prevalent special characters - and /, we observe several combinations of other characters, but their use is negligible. Network operators thus seem to follow the examples provided in RFC2317. However, the use of the same combination of special characters is not consistent across regions. As expected, network operators from developed regions lead special character use. Recall that RIRs from developed regions reached IPv4 depletion earlier than those in developing regions. Thus, RIRs from developed regions require more classless delegation compared to other regions. Table 4 also contains characters that can be tied to typographical and encoding errors. For example, $, a comma, or \032 (decimal for space). However, the contribution of typographical and encoding errors to NXDOMAIN appears negligible. We further dissect CNAME records in Table 5, which shows the presence of the `in-addr.arpa` suffix, repetition thereof, and use of alphanumeric and possibly special characters. The use of special characters contributes the most to NXDOMAIN responses and we therefore expect the absence of PTR records[10] to be the leading cause of NXDOMAIN responses rather than typographical errors in CNAMEs. Note however that naming errors do occur as we observe that repetition of the suffix has a low correlation with NXDOMAIN responses. Furthermore, we notice CNAME records containing typo-containing suffixes such as `in-addr.arp` or `inaddr.arpa`. Interestingly, the high NXDOMAIN percentages within the AFRINIC region may be due to the absence of lame delegation checks. Indeed, the RIR Comparative Policy Overview 2022-Q4 report [32] shows that, although lame delegation policies exist in all RIRs, they are not enforced by AFRINIC and ARIN. This could explain the high rate of resolution errors in CNAME chains that both regions exhibit, whereas RIPE has a lower rate of NXDOMAIN due to lame delegation enforcement.

[10] Special characters lead to NXDOMAIN, i.e., badly-configured PTR records.

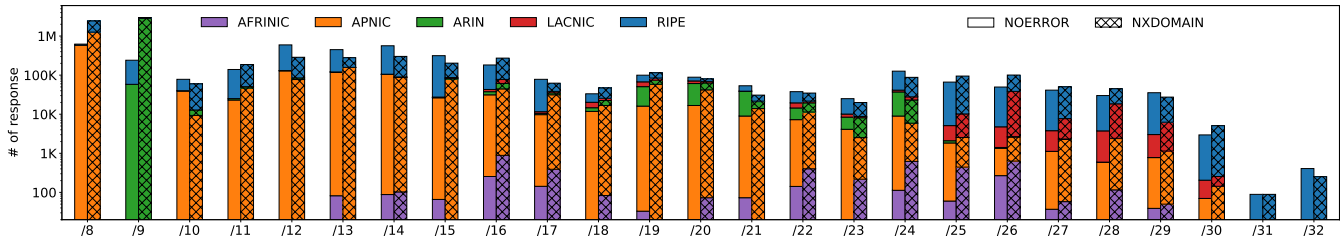

Figure 3: Daily median number of CNAME records in the active rDNS data, per prefix length and region (log scale). Contrary to RIRs, network operators use CNAME records to classlessly delegate prefixes less specific than /24s.

Table 4: Special character (S.C.) combinations (daily median) in CNAME records at RIR-level and network operators-level. RIRs use the character - exclusively while operators rely primarily on / and on - second.

	S.C.	Total	RIPE	APNIC	LACNIC	AFRINIC	ARIN	NOERROR	NXDOMAIN
Operators	$	1	1 (~)					1 (~)	
	$ -	2	2 (~)						2 (~)
	, /	198	64 (~)	128 (~)	6 (~)				198 (~)
	-	2M	**1M (13.7%)**	**374K (4.6%)**	**167K (2.1%)**	2K (~)	**208K (2.5%)**	585K (7.1%)	**1M (15.8%)**
	- /	806	484 (~)	56 (~)	135 (~)	1 (~)	130 (~)	120 (~)	686 (~)
	- \ /	1	1 (~)						1 (~)
	- \	300	46 (~)	254 (~)					300 (~)
	- _	104	104 (~)					38 (~)	66 (~)
	/	6M	**2M (23.9%)**	**1M (18.3%)**	11K (0.1%)	884 (~)	**3M (34.6%)**	485K (5.9%)	**6M (71.1%)**
	\ /	1					1 (~)		91 (~)
	/_	20		20 (~)					20 (~)
	_	3K	1K (~)	1K (~)	32 (~)		15 (~)	1K (~)	1K (~)
RIRs	-	12K	**6K (51%)**				**6K (49%)**		

Implications: The use of CNAME in rDNS at the RIR-level is part of a strategy to grapple with IPv4 depletion. The character '-' is preferred for historical reasons.[11] Interestingly, lame-delegation policy (without enforcement) is not sufficient to motivate for a good rDNS hygiene. Indeed, DNS management best practises (RFC6168) can help to mitigate NXDOMAIN responses due to the use of special characters for CNAME.

7 RDNS DEPLOYMENT IN THE WILD

Having observed that 1.6 B rDNS records have NOERROR responses, we further assess the deployment of rDNS in the wild and characterise its adoption by different providers. To this end, we use CAIDA's IPv4 prefix-to-AS data [7] to map prefixes in the active rDNS measurement data to AS numbers. Next, we leverage data from *bgp.tools* as well as data on mail providers from Liu *et al.* [25] to classify autonomous systems into: *Infrastructure Providers*, *Popular Websites (Alexa)*, and *Mail Providers*. We classify organizations that do not fall in any of these categories as *other*.

7.1 Deployment of rDNS for the IPv4 Space

We evaluate the number of response types per prefix length, category and region. We also tally NOERROR PTR responses in prefix categories to calculate the ratio of well-configured rDNS records and overall deployment. Considering all categories, we observe successful PTR record resolutions (i.e., NOERROR) responses for a little over 40% of the allocated IPv4 addresses. However, LACNIC has a high ratio of well-configured PTR records for their relatively small the pool of allocated IP address blocks. This is in contrast with regions with higher blocks allocation. Figure 4 breaks down the cumulative ratio of NOERROR responses per prefix length, network operator region, and category. Note that we use stacked bars, which means that PTR deployment ratios from operators in different regions can add up beyond 100% for a given prefix length. We plot the results for large drivers of PTR deployment, which are infrastructure and mail providers. Such organizations do not use allocations more specific than /27, while organizations of the category *other* do (these are not shown in the figure).

Mail Providers: Mail provider networks generally configure hostname on a few allocated IP addresses; they account for 22.5% of the overall PTR deployment. With most mail providers registered in the ARIN region, we expect to find a large contribution of mail providers to regional rDNS deployment. However, we find that this is less than 5% within ARIN. In contrast, RIPE and APNIC share the highest contribution (26%) of mail service providers to PTR deployment in their respective region. As mail provisioning is becoming increasingly centralized [25], the adoption of modern architecture contributes to lower PTR deployment compared to other categories, with 76.5% of domains operating with a maximum of two MX [35].

[11] The character / is considered as directory boundary in Unix systems.

Table 5: Network operator `CNAME` naming conventions (daily median) per region, observed in the active rDNS measurement data. Breaks down the use of alphabetical (A), numeric (N), alphanumeric (AN) and special characters (includes AN), as well as `in-addr.arpa` suffix presence (S.) and possible (inadvertent) repetition thereof (R.).

Region	S.	R.	Format	NOERROR	NXDOMAIN	Region	S.	R.	Format	NOERROR	NXDOMAIN
RIPE	✗	✗	alpha	39K (0.7%)	363 (~)	**RIPE**	✗	✗	alphanumeric	93K (1.7%)	157K (3.0%)
	✗	✗	numeric		259 (~)		✗	✗	special + AN	77K (1.5%)	126K (2.4%)
	✗	✓	special + AN	8K (0.2%)	31K (0.6%)		✓	✗	alphanumeric	28K (0.6%)	41K (0.8%)
	✓	✗	numeric	2M (32.7%)	82K (1.5%)		✓	✗	special + AN	561K (10.7%)	**2M (43.6%)**
	✓	✓	special + AN	20 (~)	730 (~)						
APNIC	✗	✗	alpha	382 (~)	30 (~)	**APNIC**	✗	✗	alphanumeric	7K (0.2%)	20K (0.6%)
	✗	✗	numeric		4K (0.1%)		✗	✗	special + AN	5K (0.2%)	9K (0.3%)
	✗	✓	special + AN	305 (~)	2K (~)		✓	✗	alphanumeric	284K (9.2%)	151K (4.9%)
	✓	✗	numeric	604K (19.5%)	150K (4.8%)		✓	✗	special + AN	243K (7.8%)	**2M (52.2%)**
	✓	✓	special + AN		288 (~)						
LACNIC	✗	✗	alpha	154 (~)	9 (~)	**LACNIC**	✗	✗	alphanumeric	3K (1.5%)	2K (0.9%)
	✗	✗	numeric		256 (0.1%)		✗	✗	special + AN	41 (~)	744 (0.4%)
	✗	✓	special + AN	335 (0.2%)	2K (1.2%)		✓	✗	alphanumeric	1K (0.8%)	3K (1.4%)
	✓	✗	numeric	516 (0.3%)	131 (~)		✓	✗	special + AN	60K (31.9%)	**115K (61.3%)**
	✓	✓	special + AN		11 (~)						
AFRINIC	✗	✗	alpha		3 (~)	**AFRINIC**	✗	✗	alphanumeric	83 (1.5%)	519 (9.6%)
	✗	✗	special + AN	159 (2.9%)	356 (6.6%)		✗	✓	special + AN		1 (~)
	✓	✗	alphanumeric	14 (0.3%)	5 (0.1%)		✓	✗	numeric	479 (8.9%)	**2K (29.6%)**
	✓	✗	special + AN	638 (11.8%)	**2K (28.5%)**		✓	✓	special + AN		1 (~)
ARIN	✗	✗	alpha	996 (~)	5 (~)	**ARIN**	✗	✗	alphanumeric	103K (3.2%)	36K (1.1%)
	✗	✗	numeric		1 (~)		✗	✗	special + AN	11K (0.4%)	5K (0.1%)
	✗	✓	special + AN	18K (0.6%)	2K (~)		✓	✗	alphanumeric	209 (~)	812 (~)
	✓	✗	numeric	5K (0.1%)	18K (0.6%)		✓	✗	special + AN	85K (2.6%)	**3M (92.1%)**
	✓	✓	special + AN		1 (~)						

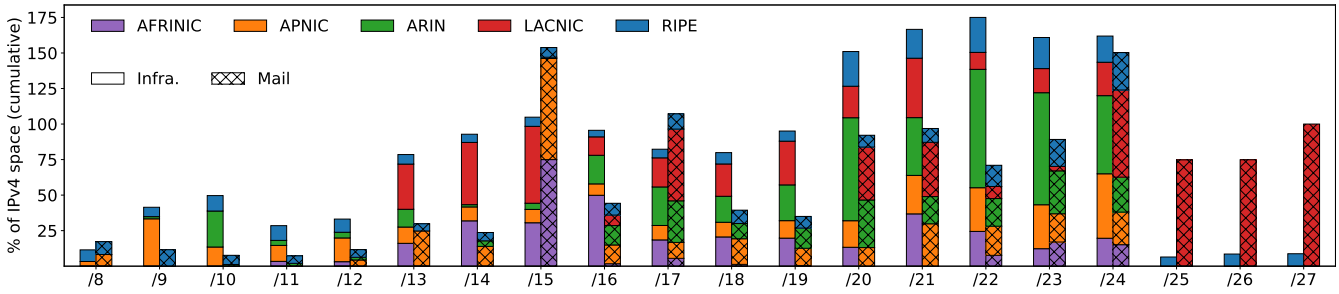

Figure 4: Ratio of `PTR` responses (daily median) in active rDNS data (`NOERROR`), broken down per region and prefix length, for *infrastructure* and *mail* providers. Deployment is driven in larger networks by infrastructure providers in developing world, while mail providers are drivers of rDNS deployment in smaller allocations.

As expected, mail providers have a lower presence in AFRINIC and LACNIC, confirming the increasing centralisation of mail service providers in the most developed regions [25].

Infrastructure Providers: With a 17% contribution to global `PTR` deployment, infrastructure providers tend to configure many of their addresses. As many services and infrastructures are hosted in the ARIN and RIPE regions, nearly half of the functional rDNS deployment in the ARIN region is driven by *infrastructure* providers. APNIC, AFRINIC and RIPE follow, while less than 5% of `PTR` deployment in the LACNIC region can be linked to *infrastructure* providers. A closer look at these organizations reveals that NIRs in LACNIC and national ISPs in AFRINIC are the main drivers behind `PTR` deployment. Although DNS traffic is increasingly centralised by cloud service providers [30], NIRs and ISPs continue to play their - initial - critical role in their respective regions.

Popular Websites (Alexa): Organizations included in the popular Websites category account for 11% of `PTR` deployment. As expected, AFRINIC has the lowest contribution from the *Alexa* category. Moreover, while *Alexa* sites have the lowest contribution in the developed world, their contribution in LACNIC and APNIC is

over 14%, suggesting new trends in: a) user browsing behavior; and b) development of Internet infrastructure in these regions. Indeed, as of 2022, less than 1.5 of the 5 billion total Internet users are from the developed world [37].

Other: Finally, our results show that organizations classified as *other* jointly account for half of all reverse DNS deployment. Unsurprisingly, as the *other* category comprises a large variety of organisations which are distributed across all the five regions. AFRINIC, LACNIC, RIPE and ARIN PTR deployment is dominated by *other*, at rates of 85%, 82%, 54% and 41% respectively. Moreover, with APNIC having the lowest ratio of *other* (35%), we see similar contributions of all four providers categories to regional deployment.

7.2 Multi-regional Autonomous Systems

Our results particularly highlight that part of the organizations driving rDNS deployment are multi-regional, i.e, they operate under – and manage resources from – multiple RIRs. Specifically, we consider mail and CDNs as organizations with the highest rDNS deployment in at least three regions across the world. With CDNs sometimes providing mail services [25], we consider popular activities by these organizations for the purpose of this analysis.

While large part of the organizations operating CDNs are from ARIN region, they tend to deploy rDNS for prefixes allocated by other registries. Recall that ARIN does not enforce lame-delegation checking. We suspect that business incentives and/or organisational complexity drive rDNS deployment, with the same organisation having different rDNS policies and hygiene in different regions. Thus, enforcing an rDNS lame delegation policy is important but not sufficient. As with fDNS, where the DNS industry is driven by commercial motivations, the deployment of rDNS is driven by economic interests.

For example, we observe that AS19551 (Incapsula) rDNS coverage in the developed world exceeds 75%, while for AS14907 (Wikimedia), a non-profit organisation, rDNS deployment is less than 25% in the same region. For critical services such as mail, AS19551 must maintain an SLA to its various customers while AS14907's email delivery is limited to operational purposes.

Amazon sibling ASes AS14618 and AS16509 registered in ARIN, follow different rDNS deployment practices per region, with none of them having rDNS deployment higher than 5% in their home registry. Reverse DNS deployment by AS14618 is almost half that of AS16509. Although AS16509 appears to be dedicated to Amazon Web Services (AWS) and operate similar number of availability zones[12] in ARIN and RIPE, rDNS is significantly more deployed in RIPE. However, Akamai sibling ASes AS32787 and AS16625 registered in ARIN, exhibit similar high rDNS deployment in their home registry. Recall that LACNIC, RIPE and APNIC are enforcing an rDNS lame delegation policy. Therefore, AS19551's, AS16509's and Akamai's high rDNS deployment is most likely due to business interest.

Implications: From an operational perspective, rDNS is still in line with its initial design, whereas for fDNS studies have shown notable operational changes, e.g., the rise of CDNs [30]. The rapid development of Internet infrastructure in less developed regions may reshape rDNS deployment; increasing the role of NIRs, national ISPs and to some extend multi-regional organisations. Therefore, business incentive may drive more rDNS deployment.

[12]https://aws.amazon.com/about-aws/global-infrastructure/regions_az/

8 CONCLUSIONS & FUTURE WORK

Many critical services, including mail and security appliances, rely on reverse DNS. Surprisingly, rDNS has not been subjected to extensive study to date. Our work characterizes global rDNS deployment across the full hierarchy for the first time, starting at the top, down to the network operators. We find that 43% of the allocated IPv4 space is mapped to configured and resolvable rDNS entries. Moreover, lame delegation policy enforcement has an impact on rDNS deployment. While ARIN is less strict in rDNS hygiene, we record a high ratio of well-configured PTR records from the same organizations in other regions, suggesting the importance of business incentives in full and functional rDNS deployment. Thus, although most services and infrastructure are provided by multi-regional organizations registered initially in ARIN, they appear to apply different rDNS policies per region. As future work, we plan to continue our analysis and add a longitudinal perspective, which will allow us to investigate rDNS deployment changes over time and per prefix length.

ACKNOWLEDGMENTS

This work was partially internally funded by SimulaMet. This research was made possible by OpenINTEL, a joint project of the University of Twente, SURF, SIDN, and NLnet Labs; and operational support from the Twente University Centre for Cybersecurity Research (TUCCR).

REFERENCES

[1] Donald E. Eastlake 3rd and Aliza R. Panitz. 1999. Reserved Top Level DNS Names. RFC 2606. https://doi.org/10.17487/RFC2606
[2] Mark P. Andrews. 2016. Adding 100.64.0.0/10 Prefixes to the IPv4 Locally-Served DNS Zones Registry. RFC 7793. https://doi.org/10.17487/RFC7793
[3] Alfred Arouna, Ioana Livadariu, and Mattijs Jonker. 2023. Lowering the Barriers to Working with Public RIR-Level Data. In *Proceedings of the Applied Networking Research Workshop*. 24–26.
[4] Doug Barton. 2012. RFC 2317 Delegations for IPv4 Blocks Less Than /24. https://www.dougbarton.us/DNS/2317.html
[5] Kevin Borgolte, Shuang Hao, Tobias Fiebig, and Giovanni Vigna. 2018. Enumerating active IPv6 hosts for large-scale security scans via DNSSEC-signed reverse zones. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 770–784.
[6] CAIDA. [n. d.]. bgp.tools:Browse the Internet ecosystem. https://bgp.tools/
[7] CAIDA. 2023. Routeviews Prefix to AS mappings Dataset (pfx2as) for IPv4 and IPv6. https://www.caida.org/catalog/datasets/routeviews-prefix2as/
[8] Stuart Cheshire and Marc Krochmal. 2013. Special-Use Domain Names. RFC 6761. https://doi.org/10.17487/RFC6761
[9] Gordon V Cormack et al. 2008. Email spam filtering: A systematic review. *Foundations and Trends® in Information Retrieval* 1, 4 (2008), 335–455.
[10] Leslie Daigle. 2004. WHOIS Protocol Specification. RFC 3912. https://doi.org/10.17487/RFC3912
[11] Ovidiu Dan, Vaibhav Parikh, and Brian D Davison. 2018. Distributed reverse DNS geolocation. In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 1581–1586.
[12] Ovidiu Dan, Vaibhav Parikh, and Brian D Davison. 2021. IP geolocation through reverse DNS. *ACM Transactions on Internet Technology (TOIT)* 22, 1 (2021), 1–29.
[13] Havard Eidnes, Paul A. Vixie, and Geert Jan de Groot. 1998. Classless IN-ADDR.ARPA delegation. RFC 2317. https://doi.org/10.17487/RFC2317
[14] FBI. 2022. Business Email Compromise and Real Estate Wire Fraud. FBI 2022 Congressional Report on BEC and Real Estate Wire Fraud. https://www.fbi.gov/file-repository/fy-2022-fbi-congressional-report-business-email-compromise-and-real-estate-wire-fraud-111422.pdf/view
[15] Tobias Fiebig, Kevin Borgolte, Shuang Hao, Christopher Kruegel, and Giovanni Vigna. 2017. Something from nothing (There): collecting global IPv6 datasets from DNS. In *International Conference on Passive and Active Network Measurement*. Springer, 30–43.

[16] Tobias Fiebig, Kevin Borgolte, Shuang Hao, Christopher Kruegel, Giovanni Vigna, and Anja Feldmann. 2018. In rDNS we trust: revisiting a common data-source's reliability. In *International Conference on Passive and Active Network Measurement*. Springer, 131–145.

[17] Kensuke Fukuda and John Heidemann. 2015. Detecting malicious activity with DNS backscatter. In *Proceedings of the 2015 Internet Measurement Conference*. 197–210.

[18] Paul E. Hoffman, Andrew Sullivan, and Kazunori Fujiwara. 2019. DNS Terminology. RFC 8499. https://doi.org/10.17487/RFC8499

[19] Bradley Huffaker, Marina Fomenkov, and KC Claffy. 2014. DRoP: DNS-based router positioning. *ACM SIGCOMM Computer Communication Review* 44, 3 (2014), 5–13.

[20] IAB. 2001. Management Guidelines & Operational Requirements for the Address and Routing Parameter Area Domain ("arpa"). RFC 3172. https://doi.org/10.17487/RFC3172

[21] IANA. 2022. IANA IPv4 Address Space Registry. https://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xhtml

[22] Daniel Karrenberg. 2012. Conclusions Drawn from Reverse DNS Event. https://labs.ripe.net/author/dfk/conclusions-drawn-from-reverse-dns-event/

[23] Scott Kitterman. 2014. Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1. RFC 7208. https://doi.org/10.17487/RFC7208

[24] Youndo Lee and Neil Spring. 2017. Identifying and analyzing broadband internet reverse DNS names. In *Proceedings of the 13th International Conference on emerging Networking EXperiments and Technologies*. 35–40.

[25] Enze Liu, Gautam Akiwate, Mattijs Jonker, Ariana Mirian, Stefan Savage, and Geoffrey M Voelker. 2021. Who's got your mail? characterizing mail service provider usage. In *Proceedings of the 21st ACM Internet Measurement Conference*. 122–136.

[26] Matthew Luckie, Bradley Huffaker, Alexander Marder, Zachary Bischof, Marianne Fletcher, and K Claffy. 2021. Learning to extract geographic information from internet router hostnames. In *Proceedings of the 17th International Conference on emerging Networking EXperiments and Technologies*. 440–453.

[27] William F. Maton and Joe Abley. 2011. I'm Being Attacked by PRISONER.IANA.ORG! RFC 6305. https://doi.org/10.17487/RFC6305

[28] P. Mockapetris. 1987. Domain names - implementation and specification. RFC 1035. https://doi.org/10.17487/RFC1035

[29] Robert Moskowitz, Daniel Karrenberg, Yakov Rekhter, Eliot Lear, and Geert Jan de Groot. 1996. Address Allocation for Private Internets. RFC 1918. https://doi.org/10.17487/RFC1918

[30] Giovane CM Moura, Sebastian Castro, Wes Hardaker, Maarten Wullink, and Cristian Hesselman. 2020. Clouding up the internet: How centralized is dns traffic becoming?. In *Proceedings of the ACM Internet Measurement Conference*. 42–49.

[31] Jon Oberheide, Manish Karir, and Z Morley Mao. 2007. Characterizing dark dns behavior. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer, 140–156.

[32] Number Resource Organization. 2023. RIR Comparative Policy Overview 2022-Q4. https://www.nro.net/wp-content/uploads/RIR-Comparative-Policy-Overview-2022-Q4.pdf

[33] Amreesh Phokeer, Alain Aina, and David Johnson. 2016. DNS Lame delegations: A case-study of public reverse DNS records in the African Region. In *International Conference on e-Infrastructure and e-Services for Developing Countries*. Springer, 232–242.

[34] Scott Rose and Wouter Wijngaards. 2012. DNAME Redirection in the DNS. RFC 6672. https://doi.org/10.17487/RFC6672

[35] Jukka Ruohonen. 2020. Measuring Basic Load-Balancing and Fail-Over Setups for Email Delivery via DNS MX Records. In *2020 IFIP Networking Conference (Networking)*. 815–820.

[36] Fernando Sanchez, Zhenhai Duan, and Yingfei Dong. 2011. Blocking spam by separating end-user machines from legitimate mail server machines. In *Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*. 116–124.

[37] Statista. 2022. Number of internet users worldwide as of 2022, by region. https://www.statista.com/statistics/249562/number-of-worldwide-internet-users-by-region/

[38] Olivier van der Toorn, Roland van Rijswijk-Deij, Raffaele Sommese, Anna Sperotto, and Mattijs Jonker. 2022. Saving Brian's privacy: the perils of privacy exposure through reverse DNS. In *Proceedings of the 22nd ACM Internet Measurement Conference*. 1–13.

[39] Ming Zhang, Yaoping Ruan, Vivek S Pai, and Jennifer Rexford. 2006. How DNS Misnaming Distorts Internet Topology Mapping.. In *USENIX Annual Technical Conference, General Track*. 369–374.

[40] Zesen Zhang, Alexander Marder, Ricky Mok, Bradley Huffaker, Matthew Luckie, Kimberly C Claffy, and Aaron Schulman. 2021. Inferring regional access network topologies: methods and applications. In *Proceedings of the 21st ACM Internet Measurement Conference*. 720–738.