# DNS Measurement Study

Cristofer Jimenez
Colgate University
cjimenez@colgate.edu

Jaanhvi Agarwal
Colgate University
jagarwal@colgate.edu

Oliver Smith
Colgate University
osmith@colgate.edu

## ABSTRACT

In this study, we attempt to replicate a previously conducted research investigation aimed at identifying "How Ready is DNS for an IPv6-Only World?" [6]. Specifically, we focus on determining the percentage of Domain Name Servers that can be resolved exclusively via IPv6. This investigation gains urgency from the fact that it has been five years since the last of the freely available IPv4 addresses were "allocated", underscoring the need for transition strategies. Understanding the current state of IPv6 readiness is crucial for future network designs, given the advantages of IPv6 such as a larger address space, enhanced security features[1], and more efficient routing capabilities that avoid packet fragmentation[4]. While recognizing the significant costs associated with migrating from well-established IPv4 infrastructures, this study does not delve into the issue of DNS operator consolidation, which affects IPv6 resolvability. Instead, our research focuses on actively measuring and analyzing the current landscape of DNS IPv6 readiness, providing insights critical for guiding the evolution of network infrastructures.

## CCS CONCEPTS

• **Networks** → **Network reliability**.

## KEYWORDS

networks, DNS, measurement, reliability

**ACM Reference Format:**
Cristofer Jimenez, Jaanhvi Agarwal, and Oliver Smith. 2024. DNS Measurement Study. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

The rapid expansion of Internet of Things (IoT) devices, from everyday smartphones and laptops to advanced home automation systems like Alexa, has dramatically increased the complexity and scale of data exchanged over the internet. These devices rely on the Internet Protocol (IP) for communication, necessitating unique identifiers known as IP addresses. Due to the cumbersome nature of numeric IP addresses, the Domain Name System (DNS) plays a crucial role by mapping these addresses to more memorable domain names. This system is analogous to a phone book for the internet.


*Conference'17, July 2017, Washington, DC, USA*

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Typically, when a user attempts to access a website such as "www.google.com," the request, or query, is handled through a DNS resolver provided by the user's Internet Service Provider (ISP). This resolver forwards the request to a DNS root name server[3], which then directs it to the appropriate Top-Level Domain (TLD) zone, such as ".com". This hierarchical structure ensures that DNS queries are resolved efficiently and reliably[2].

Our study focuses on the DNS resolution process, particularly in the context of IPv6, the latest Internet Protocol version, which is becoming increasingly important as the availability of IPv4 addresses dwindles. We modified a recursive resolver to not only gather the final response but also to record detailed data at each step of the DNS resolution process. This approach allows us to explore multiple resolution paths within the TLD zone to check for the existence of a feasible IPv6-only path.

The methodology involved enhancing the resolver to capture IPv6 addresses and name server details for each Tranco Top 1M domain. By doing so, we aimed to identify whether these domains are equipped to support communication exclusively over IPv6.

The results indicate that while a significant number of domains are prepared for IPv6, there are still gaps in the universal adoption of this protocol across all zones. This finding underscores the necessity for ongoing adjustments to DNS infrastructure to support IPv6 connectivity fully. In the following sections, we will delve deeper into the conditions necessary for establishing IPv6-only paths and the implications of our findings for the future of Internet communications.

## 2 RELATED WORK

The previous paper identified five main misconfigurations that lead to invalid IPv6 resolution. However, our paper specifically examines whether a domain is resolvable via IPv6 addresses only, meaning that the recursive resolver must identify these IPs while querying for the nameservers and A/AAAA records of the domains. The five misconfigurations mentioned in the previous paper are as follows:

(1) No AAAA records for name servers.
(2) Missing GLUE records (the parent must provide AAAA records in the ADDITIONAL section when returning NS records in the ANSWER section).
(3) No AAAA record for in-bailiwick NS.
(4) Zone of out-of-bailiwick NSes not resolving.
(5) Parent zone is not IPv6 resolvable. [6]

Some misconfigurations cannot occur independently; they must happen together to be considered a single misconfiguration. For example, misconfigurations 1 (No AAAA records for name servers)

and 5 (Parent zone is not IPv6 resolvable) are essentially the same because they both refer to the parent zones having associated AAAA records and being IPv6 resolvable up to the root zone.

The previous authors had access to the Farsight Security Information Exchange (SIE) dataset collected by Farsight Inc. However, we did not have access to this dataset. Instead, we considered any domain that did not return a DNS response (of any records) as unresolvable. In contrast, the previous authors validated whether NS records in zones replied to DNS requests with the A and AAAA records in the Farsight Datasight they had access to.

Furthermore, the paper highlights the centralization of IPv6 readiness, as it is the responsibility of a small number of operators. However, this aspect is not the primary focus of our study, so we do not delve deeply into it. [6]

# 3 METHODOLOGY

This section will detail our methodology for actively measuring DNS IPv6 readiness.
We will now outline our methodology for gathering active data measurements to assess the IPv6 readiness of specific domains in the Tranco Top 1M dataset, as previously mentioned. Our methodology's primary pipeline begins with collecting all records and responses available from our resolver during the querying of each domain. Subsequently, we iteratively analyze our active data collection to determine which domains are IPv6 resolvable according to the configuration criteria we defined earlier. Throughout our active data collection process, we encountered several unexpected challenges, and we will discuss the limitations of our methodology.

# 4 CONDUCTING ACTIVE MEASUREMENTS

## 4.1 Introduction

In our pursuit to collect accurate DNS data, we recognized the need for a specialized tool capable of managing complex DNS queries efficiently. We then attempted to create a custom recursive DNS resolver, tailored to meet our specific research requirements. Our goal was to develop a resolver that not only handles DNS lookups efficiently but also integrates advanced features to enhance data accuracy and prevent common pitfalls such as query loops and redundant requests.

## 4.2 Development of the Custom Resolver

We took inspiration from an existing GitHub project by Rohit Nambiar [5], which served as a baseline, our resolver required significant modifications and enhancements to meet our unique needs:

(1) **Data Collection:** Being that we needed to determine if there was a resolvable path to a given domain in which IPv6 records could be found, properly implemented at each level, we needed to ensure that we covered all reasonable routes to reach the domain. This required altering the structure of the recursive resolver such that all resolvable routes beneath the root servers are attempted. This is a major alteration to the original structure of the code and had severe impacts on resolution speeds. In addition, we added a data structure for data collection with accompanying functions. The data structure we used is a dictionary in which the highest level keys are the domains being resolved, the keys below being an IP that was queried in the process of that domain's resolution, and their values, a simplified form of the response to that query which is also easily traversed.

(2) **Error Handling and Efficiency:**Improving the resolver for mass resolutions included adding functions to handle successive resolutions and adding new data structures to store data as it is collected. We were also required to account for a larger range of errors that might prevent the completion of the data collection. This includes removing CNAME data collection to effectively minimize the chances of the resolver entering redundant loops. We were also forced to add a max recursion depth to prevent loops which caused data collection to terminate. To improve the efficiency of the data collection, we added special cases such that if a specific record query to a specific name server fails due to a timeout, it does not attempt that same query to that same server again.

(3) **Recursive Depth Control:** To prevent the resolver from diving too deep into DNS hierarchies—which can lead to inefficiencies and potential errors—we introduced a maximum recursion depth parameter. This feature allows us to limit the resolver's queries to essential levels, thereby optimizing performance and focus.

(4) **Selective Record Handling:** Given the specific focus of our analysis, we configured the resolver to selectively process DNS records. We excluded MX records, which were irrelevant to our study, and implemented measures to handle CNAME records more cautiously to avoid infinite resolution loops that some domains presented.

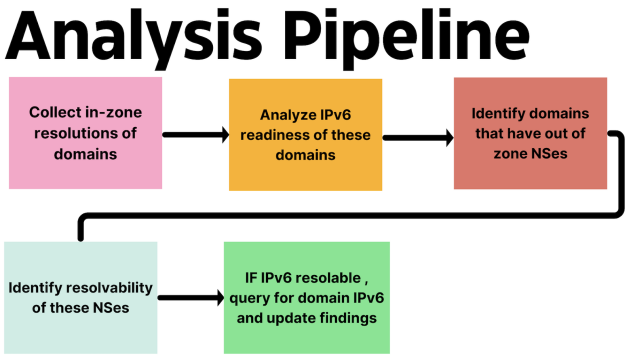
## 4.3 Integration into Active Measurements

The custom resolver was integrated into our active measurement toolkit, enabling us to systematically collect DNS data across various domains. As we deployed the resolver in real-world scenarios, we continuously refined its algorithms based on observed performance and challenges encountered during operations.

We established a routine for exporting the collected data into JSON format, using the domain name as the key. This structured data format facilitates further analysis and integration with other tools in our measurement suite.

The development and deployment of our custom recursive DNS resolver significantly enhanced our capability to conduct active measurements with higher accuracy and efficiency. By tailoring the resolver to our specific needs and continuously refining its performance, we ensured that our DNS data collection efforts were both effective and aligned with our research objectives.

### 4.4 Analysis




'Analysis.py' goes through the active data measurement JSON file and determines which domains are IPv6 resolvable based on whether there is a single path from the root to the domain that contains name servers that possess AAAA records, those within the parent zone being properly advertised through GLUE records.

Our analysis works in stages, covering the initial IPv6 resolvability based on the data collected from the resolver as described. Notably, we lacked the capability, simply due to the means at our disposal, to conduct proper resolutions using IPv6 only. For the sake of our analysis, we were forced to conduct all resolutions on and through IPv4, and while this may be a notable limitation in the quality of our findings, we chose to assess IPv6 capability based solely on the possession of and correct advertising of AAAA records. Down the path of resolution.

The second stage of our analysis is focused on those that did not resolve via IPv6 according to the data gathered but had name servers outside of their parent zone. For these cases, we compiled, for each domain, the name servers that fit this description and attempted to resolve these just as we did for the domains initially. This yields an identically structured set of data in which each name server can be identified as IPv6 resolvable. Should a name server be IPv6 resolvable, we queried that name server for the AAAA records of its respective domain and updated that domain's IPv6 resolvability should this yield results. Although we lack specific data files identifying these cases, they were relatively rare.

## 5 RESULTS

We conducted active measurements to assess DNS IPv6 readiness, identifying domains resolvable via IPv4 and/or IPv6 addresses. Our analysis revealed the following about the current state of IPv6 adoption in DNS infrastructures:

Our pie chart analysis illustrates the distribution of domain resolvability across the internet. According to our data: This distribution highlights the challenges and progress in the transition towards a more robust adoption of IPv6 within DNS infrastructures. Our study's results resonate with some key findings from the paper "How Ready is DNS for an IPv6-Only World?" by Florian Streibelt et al., while also presenting some unique observations. Here we compare our results side-by-side to provide a comprehensive view of DNS IPv6 readiness.

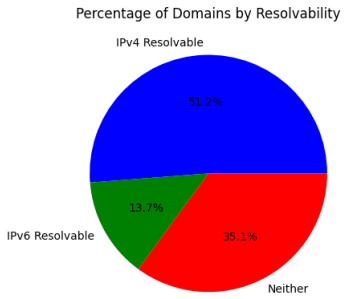



**Figure 1: IPv6 Resolvability Chart**

### 5.1 IPv4 and IPv6 Resolvability

In our analysis, **51.2%** of domains were only resolvable via IPv4 and **13.7%** were resolvable via IPv6. This is consistent with Streibelt et al., who reported a substantial percentage of domains not resolving in IPv6-only environments due to issues in the DNS delegation chain.

### 5.2 Presence of Broken IPv6 Delegations

Our study identified a **35.1%** of domains not resolvable via either IPv4 or IPv6, suggesting significant misconfigurations or inactive domains. Streibelt et al. also discussed the concept of "broken IPv6-delegation," where zones that appear to support IPv6 fail to resolve due to improper configurations. They emphasized that merely having an AAAA record is insufficient without ensuring that the entire delegation chain is IPv6 compliant.

### 5.3 Impact of DNS Centralization

Both studies highlight the centralization of the DNS infrastructure as a critical factor affecting IPv6 readiness. Streibelt et al. noted that 10% of DNS operators control the resolvability of 97.5% of domains not using IPv6. This finding aligns with our observations, where a few large providers dominate the DNS services, significantly impacting the adoption rate of IPv6.

### 5.4 Methodological Differences

Our study exclusively utilized active measurements to assess DNS IPv6 readiness, focusing on directly querying DNS servers to determine the resolvability of domains via IPv4 and IPv6. In contrast, Florian Streibelt et al.'s approach combined these active measurements with passive DNS analysis, which allowed them to leverage historical DNS query data for a longitudinal perspective over an extended period. This fundamental difference in methodology could explain some of the variances we observe between our findings and theirs, particularly in terms of the extent and nature of IPv6 resolvability issues. The use of passive data by Streibelt et al. might have provided them with insights into DNS configurations that

were not directly observable through our active-only approach. Specifically, their method could detect intermittent or historical issues that our real-time queries could not, potentially leading to a more comprehensive assessment of the DNS ecosystem's readiness for IPv6.

**Conclusions Drawn:** The comparison underlines the urgent need for DNS configuration audits and the adoption of IPv6 across the board to ensure future readiness of internet infrastructures. Both studies contribute to a growing body of evidence that while IPv6 adoption progresses, significant challenges remain due to outdated configurations and centralized control of DNS infrastructures.

## 6 LIMITATIONS & FUTURE WORK

Throughout the execution of this research, we encountered many obstacles in the form of errors, bugs, and the like, which severely impacted our time-frame for conducting measurements and analysis. This resulted in some minor but unimportant issues with our process. The major issues we encountered were largely due to the resources at our disposal.

The first notable shortcoming of our research is that we didn't have access, due to Colgate University's lack of support for IPv6, to a method of resolving using solely IPv6. This means that all resolutions or lack thereof came from IPv4 queries. With the world, moving slowly toward IPv6, and that being the exact focus of this project, it is troubling that our method of data collection had no way to account for let alone detect cases of domains that are only IPv6 resolvable. While there are likely not many, information regarding such cases would be invaluable to our findings.

Another issue that must be recognized is related to the large proportion of domains that didn't resolve at all. This data collection and analysis come down to the last moment, we lack the opportunity to correct this, however, we believe the cause may be found in the method of our resolutions. By attempting to resolve a multitude of routes to the same domain in quick succession, we theorize that we may have caused TLD name servers and possibly some lower-level name servers to cease responding to our queries. This is supported by the examples within our data of the root offering us the IPs for TLD name servers, followed by the resolution ending there, meaning that none of the TLD name servers responded to any queries, either for A or AAAA records for our target domain. For the time being, it is possible that this finding is negligible if failed resolutions are not taken into account. Although unsubstantiated, it is unlikely that the domains that failed to resolve as a result of this would be far from random. In effect, if they are relatively random, this could be tentatively viewed as simply having collected less overall data.

In the future this issue could be resolved, or better identified by the usage of multiple different IPs and possibly locations to issue our queries from. This would prevent the possibility of being cut off for over-querying. We could also increase the spread of our collections by coupling this with proper workload distribution across systems or at least threads. Doing so from sites and devices that support IPv6 would also bring new and insightful data to our work.

An important distinction between our study and the study by Florian Streibelt et al. concerns the rate of non-resolvable domains. In our active measurement, approximately **30%** of domains did not resolve, a stark contrast to the **0.43%** reported in their study during similar active measurement conditions. This significant discrepancy could be influenced by our IPv4-only querying capability and possibly by the intensity and frequency of our DNS requests, which might have led to temporary blocking or rate-limiting by some DNS servers. Understanding these differences is crucial for interpreting the impact and accuracy of DNS infrastructure assessments.

## ACKNOWLEDGMENTS

We would like to acknowledge the foundational research conducted by Florian Streibelt and others in their paper *How Ready is DNS for an IPv6-Only World?* We are also grateful for the guidance and support provided by our professor, Aaron Gember-Jacobson, which was invaluable in our attempt to replicate this research.

## REFERENCES

[1] Jailendrasingh Beeharry and Bhissum Nowbutsing. 2016. Forecasting IPv4 exhaustion and IPv6 migration. In *2016 IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech)*. 336–340. https://doi.org/10.1109/EmergiTech.2016.7737362

[2] DNS Design. 2008. What is DNS?

[3] Benoit Donnet and Timur Friedman. 2007. Internet topology discovery: a survey. *IEEE Communications Surveys Tutorials* 9, 4 (2007), 56–69. https://doi.org/10.1109/COMST.2007.4444750

[4] Walter Goralski. 2014. What is the difference between IPv4 and IPv6? Juniper Networks. https://www.juniper.net/us/en/research-topics/what-is-ipv4-vs-ipv6.html

[5] Rohit Nambiar. 2018. Recursive DNS Resolver. https://github.com/rohitn212/Recursive-DNS-Resolver.git

[6] Florian Streibelt et al. 2023. How Ready is DNS for an IPv6-Only World?. In *Passive and Active Measurement. PAM 2023. Lecture Notes in Computer Science*, Vol. 13882. Springer, Cham, Chapter 22. https://doi.org/10.1007/978-3-031-28486-1_22