Capstone project: Proposal (Due: Wed, Mar 27 @ 11pm)

COSC 465, Computer Networks, Spring 2024

Team Members: Jaanhvi Agarwal, Oliver Smith, Cristofer Jimenez
GitHub Repo Link: git@github.com:Jaanhvi18/COSC465_Capstone_Project.git

Project objective and motivation —

- *In this project, we will:*
    - Investigate DNS readiness for an IPv6-only environment by extending the networking research paper " How Ready is DNS for an IPv6-Only World? "
    - Identifying the implications of major corporations, i.e. third-party hosting providers, dominating DNS management on the adoption of IPv6 and the overall robustness of DNS.
    - Identify important reasons why DNS needs to work with IPv6, especially as IPv4 addresses run out and more people use IPv6.
    - Measure / Assess the readiness of DNS for a scenario where IPv6 is the sole internet protocol in use.
    - Study the problems and risks of a few companies controlling most DNS hosting.
    - Develop a detailed understanding of DNS's capability to support a future where IPv6 is the predominant internet protocol
    - Identify third-party hosting providers that can resolve domains to IPv6 addresses, particularly when domains fail to resolve via IPv6.

Survey of related prior artifacts —

- Artifact 1: research paper "How Ready is DNS for an IPv6-Only World? "
- Artifact 2: "Something from Nothing (There): Collecting Global IPv6 Datasets from DNS"
- Artifact 3: (Potential dataset): The latest Tranco list
- Artifact 4: https://github.com/mutax/dns-v6-readyness

Project Plan:

- *Milestone 1 (Wed, Apr 10 @ 11pm):*
    - Downloading the dataset ← Utilize the Farsight SIE dataset or a similar comprehensive DNS dataset that includes IPv6 resolution data.
    - Setting up the resolver ← Configure a DNS resolver that supports both IPv4 and IPv6 to test resolution paths and identify broken delegations.
    - Figuring out how to load and view original study findings (must be un-zipped/viewed by line, we are not yet certain on how to achieve this)
    - Deciding on the scope of domain analysis, including Top-Level Domains (TLDs) and possibly Second-Level Domains (SLDs) or sub-level domains, to understand their IPv6 delegation status.
    - Identifying and categorizing scenarios where IPv6 resolution fails or is suboptimal, highlighting the need for correct IPv6 glue records and delegation paths
    - Querying/Data Collection sourcing from Tranco Top 10K (possibly more of the domains if we have the time and resources) rather than Alexa Top 1M (which is depreciated)
        - DNS resolutions with solely IPv6

        - (Not sure if we need to do basic DNS resolutions trying to pull IPv6 informations to see is domains/name servers have ATTEMPTED to set up IPv6, even if they haven't properly implemented it, or have full access up to root)
        - Bailiwick information (not sure if we need this or not)
        - Zoning information (not sure if we need this or not)
        - AS organization owner (for information about which ASes/ AS orgs are covering/not covering IPv6 implementation)
    - Possibly conducting identical data collection from a different location to identify potential variation in IPv6 standardization (assuming we utilize 3rd part methods of conducting data collection which are capable of such efforts)

- *Milestone 2 (Wed, Apr 24 @ 11pm):*
    - Cleaning up the data ← Filter and preprocessing the dataset to focus on relevant records, such as NS, AAAA, and A records, to identify potential issues in IPv6 delegation
    - Identify the Operators ← Determine the major DNS operators, i.e. the 3rd party hosting providers, and their impact on IPv6 resolution.
    - Do final data analysis ← we are not yet certain how much of the research we will be replicating, so this section is a bit open-ended as of right now
    - Generate visuals from data
    - Produce write-up/paper
        - Code/dependencies/files/etc.
        - Methods
        - Analyzed data
        - Supporting visuals
        - Explanation of findings
        - Limitations
        - Possible future directions section

Artifacts to deliver *— What artifacts (e.g., paper, source code, datasets, etc.) will your project produce?*

- Paper
- Source code for:
    - Data collection
    - Data cleanup & analysis
    - Generating data visualizations
    - Scripts for execution
- Raw dataset(s)
- Cleaned dataset(s)
- Final datasets
- Visuals