UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

DEMÉTRIO FRANCISCO FREITAS BOEIRA

DNS Measurement: Explorando e Investigando o Cenário de DNS na Internet

Monografia apresentada como requisito parcial para a obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Lisandro Zambenedetti Granville
Co-Orientador: Luciano Zembruzki

Porto Alegre
2023

AGRADECIMENTOS

Gostaria de expressar minha gratidão a todas as pessoas que me ajudaram durante este processo de conclusão do meu curso. Em primeiro lugar, gostaria de agradecer a minha mãe, Rosângela, pelo apoio incondicional, incentivo e amor que me proporcionou. Nunca mediu esforços para me ajudar, muitas vezes me colocando acima dela mesmo como prioridade.

Também gostaria de agradecer a meu pai, Carlos, e minha madrasta, Vanise, por sempre me apoiarem em minhas decisões e estarem presentes em momentos importantes da minha trajetória. Com toda a certeza, o apoio dado no final deste trabalho foi essencial para que o mesmo fosse concluído com sucesso.

Não poderia deixar de mencionar meus colegas Pedro, Lorenzo e Gustavo. Agradeço por estarem ao meu lado durante todo o processo, dividindo conhecimentos e experiências, ajudando em momentos de dificuldades e incentivando quando precisei. Sem vocês, com certeza a jornada teria sido mais árdua e menos enriquecedora.

À meu orientador, Professor Lisandro, obrigado por confiar e acreditar em mim, mesmo quando as coisas pareciam que não aconteceriam. Teus conselhos e conhecimentos passados foram fundamentais.

Luciano, aluno de doutorado do instituto de informática da UFRGS, que esteve lá tirando as minhas dúvidas em vários momentos, me dando dicas e direcionando para os melhores caminhos.

Por fim, gostaria de expressar meu agradecimento a todas as pessoas que, de alguma forma, contribuíram para a minha formação e crescimento pessoal e profissional. Obrigado!

RESUMO

O serviço de DNS (Sistema de Nomes de Domínio - *Domain Name System*) é um dos pilares da internet. Esse serviço permite que os usuários acessem sites na internet por meio de nomes de domínio fáceis de lembrar, em vez de digitar endereços IP numéricos complexos. O DNS atua como um diretório que traduz o nome de domínio que o usuário insere no navegador em um endereço IP correspondente, permitindo que a comunicação entre computadores em diferentes redes aconteça. O sistema de DNS é essencial para a navegação na internet e para a resolução de problemas de conectividade em redes.

A concentração dos provedores de serviço de DNS na internet é um problema que afeta a segurança e a privacidade dos usuários, bem como a acessibilidade da rede. A dependência de um pequeno número de grandes provedores de DNS pode levar a riscos de violação de dados e à interrupção do serviço em caso de falhas. Para lidar com essas questões, existem iniciativas para descentralizar o sistema de DNS e permitir que mais provedores de serviços atuem em um papel mais proeminente. Embora existam muitos provedores de serviços de DNS, a maioria do tráfego de DNS na internet é processado por um pequeno número de grandes provedores de serviços de DNS.

Neste trabalho, será abordado o tema da concentração do DNS na Internet. Para isso, será apresentado a ferramenta DNS Measuring, a qual foi a solução criada para realização de medições de DNS ao longo do tempo. Com os dados obtidos através dessas medições, foram respondidas algumas perguntas relevantes para a exploração do tema, os quais podem ser muito úteis para a resolução do problema em uma escala global.

DNS Measurement - Exploring and Investigating the DNS Scenario on the Internet

ABSTRACT

The Domain Name System (DNS) service is one of the pillars of the internet. This service allows users to access websites on the Internet through easy-to-remember domain names, rather than typing complex numeric IP addresses. DNS acts as a directory that translates the domain name that the user enters into the browser into a corresponding IP address, allowing communication between computers on different networks to take place. The DNS system is essential for internet browsing and for solving network connectivity problems.

The concentration of DNS service providers on the Internet is an issue that affects user security and privacy, as well as network accessibility. Reliance on a small number of large DNS providers can lead to risks of data breaches and disruption of service in the event of failures. To address these issues, there are initiatives to decentralize the DNS system and allow more service providers to play a more prominent role. Although there are many DNS service providers, the majority of DNS traffic on the internet is handled by a small number of large DNS service providers.

In this work, the issue of DNS concentration on the Internet will be approached. For this, the DNS Measuring tool will be presented, which was the solution created to perform DNS measurements over time. With the data obtained through these measurements, some questions relevant to the exploration of the theme were answered, which can be very useful for solving the problem on a global scale.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

DNS	Domain Name Service
DNSSEC	DNS Security Extensions
IP	Internet Protocol
NS RECORD	Registro DNS
HTTP	HyperText Transfer Protocol
S.O.	Sistema Operacional
IPv4	Internet Protocol versão 4
SCRIPT	Conjunto de instruções interpretados durante a execução
IPv6	Internet Protocol versão 6
ISP	Provedor de Internet
URL	Uniform Resource Locator
TLD	Top Level Domain
AS	Sistema Autônomo
CAIDA	Centro de Análise e Disseminação de Informação de Internet
DDOS	Ataques de Negação de Serviço
CACHE	Mecanismo de Armazenamento Temporário de Dados
QUERY	Consulta em Banco de Dados
REGISTRARS	Provedores de Serviços de Registro
REGISTRIES	Registradores de Nomes de Domínio

SUMÁRIO

1 INTRODUÇÃO

O serviço de DNS (Sistema de Nomes de Domínio - *Domain Name System*) é fundamental para a operação da Internet, fornecendo um mecanismo de nomeação globalmente hierárquico que permite a associação de servidores, redes e serviços em endereços IP (*Internet Protocol*). Desta maneira, torna-se possível que as pessoas acessem sites e endereços de páginas na Internet por meio de nomes fáceis de lembrar (como Google.com), em vez de números de endereços IP (*Internet Protocol*) difíceis de lembrar. Esses nomes fáceis de lembrar são chamados de domínios e, quando desejamos acessar algum site, digitamos esse nome de domínio em nosso navegador. Já os endereços IP são números nos quais contêm o "endereço" deste site na Internet, ou seja, contêm exatamente em qual servidor o conteúdo da página está disponível. Desta maneira, o DNS é um dos principais serviços da Internet.

1.1 Funcionamento do DNS

O serviço de DNS é o responsável por traduzir e associar um nome de domínio no seu endereço IP correspondente, para que seja possível acessá-lo sabendo somente o seu nome de domínio. Ao tentar resolver um nome de domínio (como por exemplo, *www.google.com*), o requisitante, primeiramente, envia uma consulta DNS para seu DNS resolvedor recursivo, que é um servidor DNS que tem autorização para resolver nomes na Internet pelo requisitante. Se o resolvedor não tiver o registro DNS (*NS Record*) armazenado, ele consultará o próximo servidor de DNS em sua hierarquia para obter uma resposta. Esses servidores DNS são os chamados autoritativos, sendo esses divididos em zonas. Desta forma, conhecem o conteúdo de uma zona DNS com base no conhecimento local, podendo assim responder a perguntas sobre essas zonas.

Por exemplo, um requisitante/cliente pode se conectar ao resolvedor de DNS público localizado no endereço IP 1.1.1.1 e realizar a solicitação do IP do domínio *www.google.com*. O resolvedor enviará consultas ao servidor de DNS autoritativo (caso não tenha os registros DNS armazenado) de *www.google.com* em nome do requisitante, que são *ns1.google.com* e *ns2.google.com*, e estes irão retornar com o endereço IP desejado ao requisitante, neste caso, o endereço IP de *www.google.com*.

1.2 Problema

Como muitas empresas não administram sua própria infraestrutura de DNS, mas sim dependem de provedores de DNS terceirizados, há indícios de que haja uma concentração entre os provedores de serviço de DNS na Internet. Isso vem se tornando uma preocupação crescente, uma vez que, caso realmente exista uma concentração, a maior parte dos registros de DNS de muitas empresas seriam gerenciados por um pequeno número de grandes provedores, podendo estar, desta maneira, compartilhando a mesma infraestrutura e *datacenter,* e este fato poderia levar a riscos significativos de segurança e privacidade, bem como a impossibilidade de resolução de nomes de domínio em caso de alguma interrupção ou falha de serviço em algum destes grandes provedores. Este fato já vem sendo explorado por atacantes pela Internet, sendo que diversos servidores de DNS autoritativos já foram vítimas de ataques de DDoS (Negação de Serviço) (VIXIE; SNEERINGER; SCHLEIFER, 2002; WEINBERG M., 2016).

1.3 Contribuições e perguntas de pesquisa

Neste trabalho, foi realizado um estudo para averiguar o cenário do serviço de DNS na Internet, visando validar se realmente existe uma concentração por parte de grandes provedores. Também visa entender o seu comportamento através da coleta de alguns dados e argumentar acerca dos possíveis riscos inerentes a este fato, caso comprovado. Para isso, foram realizadas medições de DNS em diversos domínios ativos na Internet, com o objetivo de verificar qual é o provedor responsável por manter os servidores DNS (*NS Records*) associados a este domínio, ou seja, qual o provedor responsável por manter a tradução entre o nome do domínio e o endereço IP correspondente a este domínio. As medições foram feitas através da ferramenta criada para este trabalho, denominada de *DNS Measurement*. No Capítulo 3, serão discutidos detalhes acerca da solução criada e também abordada sobre aspectos mais gerais e em alto nível. No Capítulo 4, são aprofundados os métodos e ferramentas utilizadas para o desenvolvimento da ferramenta.

Para auxiliar e guiar este estudo, foram elaboradas algumas perguntas de pesquisa acerca do tema, que serão respondidas ao longo do trabalho, com base nos dados coletados com as medições. São elas:

1) Qual foi o comportamento observado acerca do serviço de DNS na Internet? De fato, temos indícios que comprovam a existência da concentração deste serviço em poucos provedores, ou é um serviço bem distribuído?

2) Quais são os 10 provedores de serviços de DNS na Internet responsáveis pelo maior número de domínios? Eles foram os mesmos em todas as medições?

3) Os domínios dos maiores provedores de serviços de DNS estão hospedados em seus próprios serviços de DNS ou de terceiros?

4) Os domínios brasileiros costumam ter seus registros de DNS administrados por provedores brasileiros?

5) Os domínios russos e chineses costumam ter seus registros de DNS administrados por provedores de seus próprios países? Esta pergunta é uma curiosidade, pois se tratam de países historicamente rivais dos Estados Unidos.

Com os resultados e análises obtidos neste estudo, será possível ter uma visão mais clara sobre a possível concentração dos provedores de serviços de DNS na Internet e suas implicações. Espera-se que este trabalho possa contribuir para o debate sobre a importância da diversificação dos provedores de serviços de DNS, bem como para o desenvolvimento de soluções que possam garantir a segurança, a privacidade e a continuidade do serviço de DNS, essenciais para a operação da internet como um todo.

1.4 Próximas etapas

O restante deste trabalho está organizado em capítulos, sendo eles: Capítulo 2 - Contexto e Revisão de Literatura, onde será feita uma breve recapitulação de trabalhos relacionados a centralização do DNS; Capítulo 3 - Solução, onde é abordado a solução construída em alto nível; Capítulo 4 - Implementação, onde aprofunda-se em aspectos técnicos da solução; Capítulo 5 - Avaliação, onde serão avaliados os resultados e respondidas as perguntas de pesquisa abordadas neste capítulo; Capítulo 6 - Conclusão, onde será feito um fechamento de tudo o que foi discutido ao longo do trabalho, reflexões, aprendizados e próximos passos.

2 CONTEXTO E REVISÃO DA LITERATURA

Antes de começar a tratar sobre algumas obras relacionadas ao tema da centralização, é importante aprofundar um pouco mais nos conceitos e na história do DNS na Internet. As seções a seguir servirão para dar ao leitor mais detalhes sobre o funcionamento deste serviço, sua importância e um pouco de como e porque o mesmo surgiu.

2.1 O Sistema de Nomes na Internet - DNS

Dar nome para as coisas é uma característica típica humana, justamente pela sua facilidade de lembrar e distinguir. Não é diferente para os serviços/sites/servidores na Internet. Contudo, nem sempre tivemos um sistema robusto e desenvolvido como temos hoje com o DNS. Inicialmente, todos os sites conectados à rede doméstica tinham uma cópia de um arquivo chamado HOSTS.TXT. Este arquivo fornecia o mapeamento de nomes para endereços IP na rede (STEWART, 2019). No início de 1974, havia ainda menos de cinquenta servidores na ARPANET (STEWART, 2019), o que indicava que manter arquivos locais e sincronizados não parecia uma tarefa muito difícil. No entanto, com o tempo, percebeu-se que manter cópias separadas deste arquivo sincronizadas com uma rede que crescia rapidamente era impraticável. Neste momento, começaram a ser desenvolvidos estudos para entender qual era a melhor maneira de armazenar esses dados (DEUTSCH, 1973). No entanto, manter um banco de dados central era propenso a erros. Até que, no início de 1982, problemas com a transmissão de correspondência pela rede levaram ao início do conceito atual de nomes de domínio hierárquicos de forma estruturada (POSTEL, 1982). O primeiro conjunto de especificações do que conhecemos hoje como DNS data de 1983 (MOCKAPETRIS, 1983a; MOCKAPETRIS, 1983b). Em 1987, as especificações do DNS foram atualizadas, e isso resultou no protocolo básico que permanece em uso até hoje (MOCKAPETRIS, 1987a; MOCKAPETRIS, 1987b)

2.2 Hierarquia de DNS

A estruturação do DNS se dá como um banco de dados distribuído hierarquicamente (CHANDRAMOULI; ROSE, 2006). Ao realizar o acesso de serviços na Internet por meio de nomes de domínio amigáveis, em vez de endereços IP, os usuários precisam de um sistema de mapeamento entre os nomes de domínio para endereços IP. Essa tradução é a tarefa principal do DNS. A infraestrutura do DNS compreende entidades distribuídas geograficamente em todo o mundo. A Figura 2.1 mostra um pouco mais sobre a estrutura hierárquica do DNS:

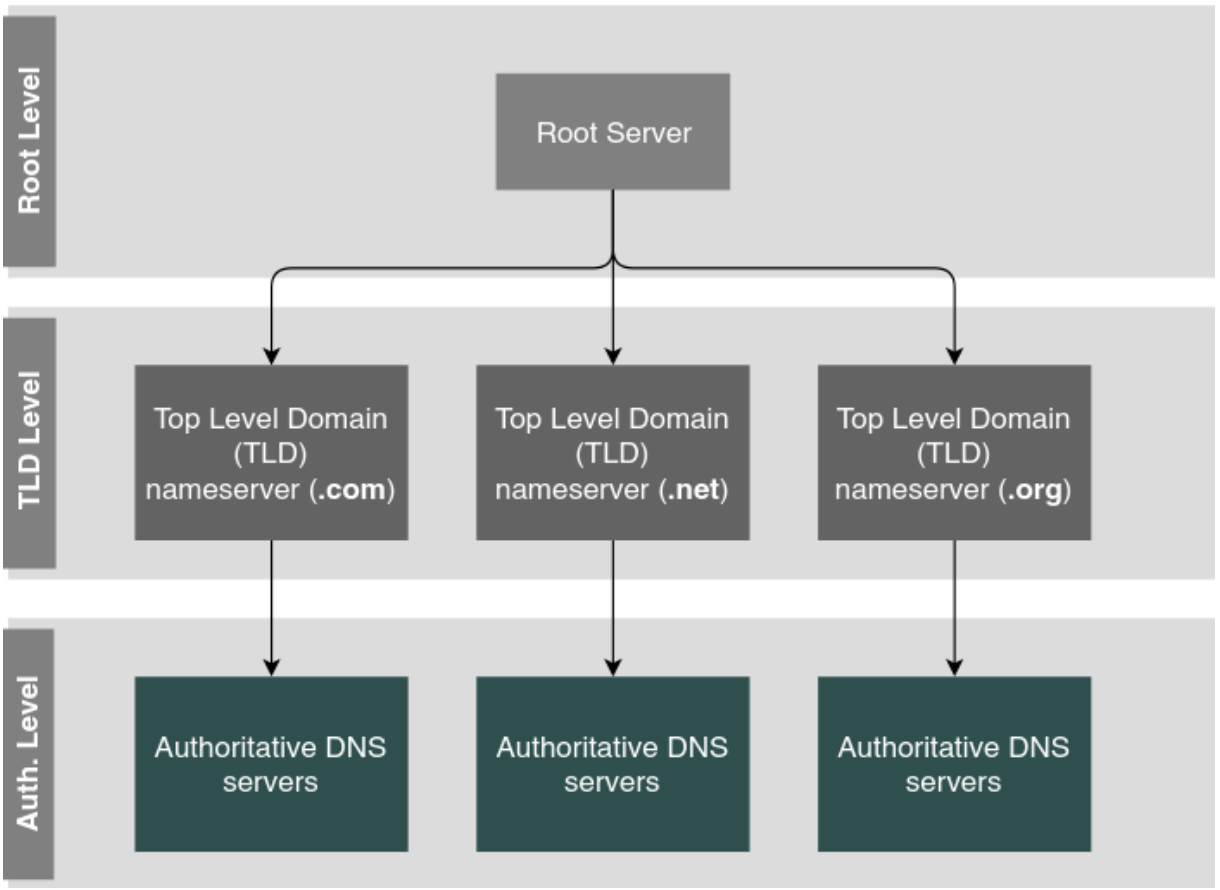

Figura 2.1 - Estrutura hierárquica do DNS

Servidores DNS Autoritativos - Os mais abaixo na imagem: pode ser hospedado pela própria infraestrutura ou por provedores terceirizados, como Cloudflare ou Amazon. Esses servidores DNS autoritativos são responsáveis por mapear e traduzir nomes de domínios individuais em endereços IP.

Servidores DNS raiz (Root) - No topo da hierarquia estão os servidores “raiz”. Estes servidores são responsáveis por armazenar os registros correspondentes ao próximo nível na hierarquia, que são os servidores de nomes de domínio (TLD).

Top Level Domains (TLD) - Os que estão no meio na imagem acima: são aqueles encontrados na extrema direita da URL (Uniform Resource Locator, em português - "localizador uniforme de recursos") como por exemplo os sufixos *.com, .net, .org* e também os identificadores de países como *.br* ou *.ar*. Cada servidor de nomes TLD é responsável por manter os registros correspondentes aos servidores DNS autoritativos dos domínios que se enquadram sob esse TLD. Por exemplo, os servidores DNS *.com* mantêm registros para os domínios terminados com esse TLD, como por exemplo o *facebook.com* ou *apple.com*.

2.3 Resolução de Nomes

Explicados alguns conceitos importantes sobre a estrutura de DNS, nesta seção será detalhado um pouco mais sobre como, de fato, partimos de um nome de domínio até o endereço IP correspondente a este domínio. Para isso, é necessário que seja utilizado o protocolo DNS.

Quando o usuário (também chamado de cliente, pois não necessariamente é uma pessoa) deseja resolver (acessar o seu endereço IP e, consequentemente, o seu conteúdo) o domínio *www.google.com*, por exemplo, é necessário, antes de realizar a chamada HTTP (*Hyper Text Transfer Protocol*), que seja realizada uma chamada de DNS para o resolvedor, que, conforme descrito na seção 1.1, é um servidor de DNS que tem a permissão para resolver nomes de domínio em nome do cliente. A Figura 2.2 a seguir ilustra os passos executados para a resolução de nomes de domínio no "pior caso", aquele onde não tem-se o endereço IP do domínio em cache local.

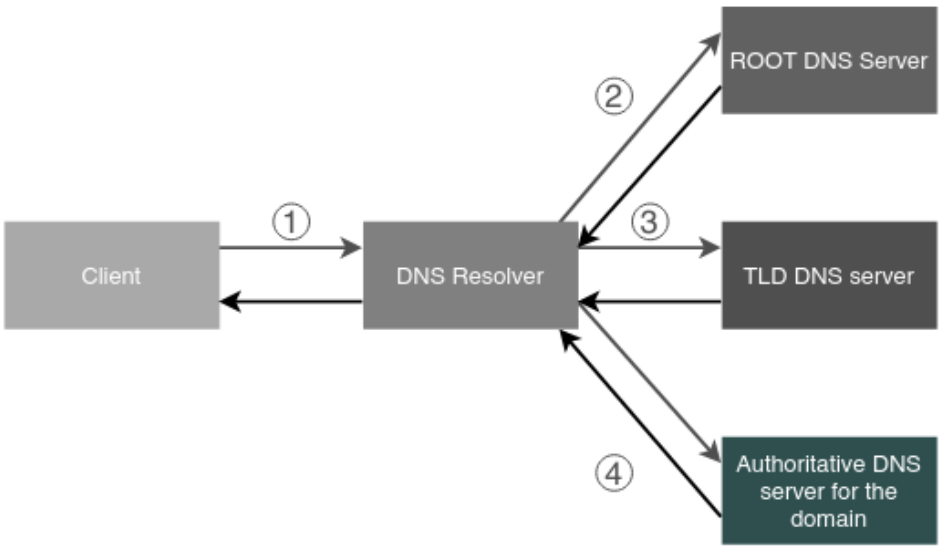

Figura 2.2 - Esquema do processo de resolução de nomes

Consulta a um resolvedor - No passo 1 da figura , o cliente envia uma consulta ao seu resolvedor, solicitando a resolução de algum nome. Esse resolvedor geralmente é o Provedor de Internet (ISP) do cliente.

Consulta a um servidor raiz - No passo seguinte, o resolvedor então envia uma consulta para um servidor raiz, perguntando sobre o domínio. O servidor raiz, por sua vez, não tem permissão para responder em nome do domínio. Contudo, o servidor raiz, conhecendo o endereço do TLD responsável pelo domínio, delega a atividade para o próximo nível hierárquico, que são os Top Level Domains (TLDs).

Consulta ao servidor TLD - No passo 3 da figura, o resolvedor do cliente consulta o servidor de TLD responsável pelo domínio. Se, por exemplo, a consulta fosse para resolver o domínio gov.br, seria consultado o TLD responsável pelo *.br*. Deste modo, o TLD já é capaz de responder em qual zona DNS está localizado o servidor autoritativo responsável pelo domínio, mas ainda não sabe exatamente onde está o endereço do domínio. Para isso, finalmente, é consultado o servidor de DNS autoritativo do domínio.

Consulta ao servidor autoritativo - No passo final, o resolvedor, agora em posse do endereço do servidor DNS autoritativo responsável pelo domínio desejado, consulta-o para, agora sim, obter o endereço IP onde está localizado o conteúdo do domínio.

2.4 Principais Preocupações

A preocupação sobre uma possível concentração/centralização de provedores de DNS na Internet, isto é, o fato de que um pequeno número de empresas e organizações controla a maioria dos servidores DNS na Internet, tem crescido com o passar dos anos e tem sido foco de estudo e observação de diversos pesquisadores, autores e instituições (WANG; XIE; YUAN, 2015). Isso se dá graças ao fato que a centralização de DNS pode levar a alguns sérios problemas, como os citados a seguir.

Em primeiro lugar, pode haver preocupações com privacidade e segurança de dados (KLEINWÄCHTER; KNIGHT, 2019). Quando os usuários fazem uma consulta de DNS, eles enviam informações sobre seus hábitos de navegação para o servidor DNS. Se um pequeno número de empresas/provedores controla a maioria dos servidores DNS, isso pode resultar em uma concentração de dados que pode ser alvo de violações de privacidade ou segurança.

Além disso, a concentração/centralização do DNS também pode levar a problemas de censura e controle de conteúdo (HUANG; KALBARCZYK; IYENGAR, 2018). Se poucas empresas/provedores controlam a maior parte dos servidores de DNS, ela pode ter o poder de

bloquear o acesso a determinados sites ou domínios, limitando assim a liberdade de expressão e acesso à informação.

Outro problema é a possibilidade de falhas e interrupções do serviço. Se um servidor DNS centralizado falha, pode afetar o acesso a sites e serviços em toda a internet. Isso pode ser problemático em situações de emergência ou quando há um grande número de usuários tentando acessar o mesmo servidor DNS.

Em resumo, a centralização do serviço de DNS na internet pode levar a preocupações bem relevantes com privacidade, segurança, liberdade de expressão e interrupções do serviço. Por isso, esse tema vem sendo amplamente discutido e estudado. É importante e prudente que sejam adotadas medidas para promover a descentralização do DNS e diversificar o número de provedores de serviços de DNS. Isso pode ajudar a reduzir os riscos associados à centralização do DNS e garantir que a internet continue sendo um espaço aberto, livre, resiliente e seguro.

2.5 Trabalhos Relacionados

Nesta seção, serão abordados e comentados sobre cinco estudos e artigos relacionados ao tema da centralização do DNS. Muitos autores e pesquisadores estão dedicando seus esforços a explorar melhor o tema, além de evidenciar os riscos que essa centralização pode causar. Repare que, na maioria dos estudos e artigos, alguns temas e preocupações se repetem, como por exemplo o tema de DDoS, o tema da privacidade, a questão governamental, entre outros.

2.5.1 *Centralization and Performance of the Domain Name System*

Este é um artigo de pesquisa (WANG; XIE; YUAN, 2015), onde é analisado como a centralização afeta o desempenho do DNS. Os autores descobriram que a centralização pode afetar negativamente o tempo de resposta do DNS em algumas regiões.

Este artigo analisa o impacto da centralização no desempenho do Sistema de Nomes de Domínio (DNS) na Internet. O estudo analisou a distribuição geográfica dos servidores raiz do DNS e os provedores de serviços de DNS em todo o mundo. Os resultados indicam que a concentração de servidores em alguns países pode afetar negativamente o tempo de resposta do DNS em outras regiões. O estudo sugere que a adoção de técnicas de distribuição geográfica pode melhorar o desempenho do DNS em nível global.

Além disso, o estudo também aborda a importância do desempenho do DNS para a qualidade da experiência do usuário na Internet. Ele destaca que a concentração de servidores DNS em algumas regiões do mundo pode levar a atrasos no tempo de resposta para usuários em outras partes do mundo. Seguem abaixo alguns trechos de destaque, sob a minha avaliação, do artigo citado:

- "A centralização do DNS é um problema crucial na Internet de hoje, pois pode levar a um único ponto de falha e vulnerabilidades de segurança."
- "Este estudo investiga os efeitos da centralização do DNS na performance do sistema. Descobrimos que a centralização não leva necessariamente a melhorias na performance do DNS, e que a descentralização pode melhorar a resiliência do sistema em relação a ataques de negação de serviço (DDoS)."
- "Observamos que a centralização do DNS é influenciada pelo tamanho do sistema, pela distribuição geográfica dos servidores DNS e pela concentração de domínios em um pequeno número de registradores."
- "Com base em nossas análises, sugerimos a adoção de políticas que promovam a descentralização do DNS, incluindo a distribuição geográfica dos servidores DNS, a promoção de registros de domínios em diferentes registradores e o incentivo à adoção de tecnologias como o DNS Security Extensions (DNSSEC) e o DNS sobre o protocolo Transport Layer Security (DNS-over-TLS)."

2.5.2 *Threats to the Decentralization of the Domain Name System*

Neste artigo (KLEINWÄCHTER; KNIGHT, 2019), são exploradas as ameaças à descentralização do DNS. Os autores argumentam que a concentração de poder em torno de algumas empresas de tecnologia pode comprometer a diversidade de provedores de DNS, resultando em menor concorrência e potencialmente colocando em risco a privacidade e segurança dos usuários da Internet.

O artigo sugere medidas para mitigar esses riscos, incluindo a promoção da diversidade de provedores de DNS. Alguns dos trechos que mais se destacam, sob a minha avaliação, em relação ao artigo, são os trechos destacados abaixo:

- "A centralização do DNS foi, em grande parte, impulsionada pela concentração do mercado de registradores de nomes de domínio (registries) e provedores de serviços de

registro (registrars), que resultou em um pequeno número de empresas que têm controle sobre um grande número de domínios."

- "A centralização do DNS apresenta vários riscos, incluindo a ameaça de ataques de negação de serviço (DDoS) direcionados a servidores DNS centrais, que poderiam interromper a resolução de nomes de domínio para grande parte da Internet."
- "A centralização do DNS também pode permitir que países ou empresas, individualmente ou em conjunto, exerçam controle sobre o sistema de nomes de domínio."
- "Embora o DNS continue a ser centralizado em muitos aspectos, existem esforços em andamento para promover a descentralização do sistema, incluindo a implementação do DNS Security Extensions (DNSSEC) e o desenvolvimento de sistemas de nome de domínio alternativos baseados em blockchain."

2.5.3 *Internet Security Challenges: DNS as a Case Study*

Neste artigo (HUANG; KALBARCZYK; IYENGAR, 2018), são discutidos os desafios de segurança relacionados à centralização do DNS. Os autores exploram como os ataques ao DNS podem comprometer a segurança da Internet em geral, e destacam a importância de medidas de segurança robustas para mitigar esses riscos.

O artigo explora diferentes tipos de ataques ao DNS, incluindo ataques de envenenamento de cache (mecanismo de armazenamento temporário de dados) DNS, ataques de negação de serviço distribuído (DDoS) e ataques de desvio de DNS. Os autores também destacam a importância de medidas de segurança, incluindo o uso de DNSSEC e firewalls (componente de segurança de rede) de DNS. Alguns dos trechos que mais se destacam, sob a minha avaliação, em relação ao artigo, são os trechos destacados abaixo:

- "Este artigo aborda a questão da centralização do DNS e seus possíveis impactos na privacidade dos usuários da Internet. A centralização do DNS pode permitir que uma única entidade controle e monitore grande parte do tráfego da Internet, incluindo informações sensíveis sobre os usuários."
- "Um dos principais desafios para abordar a centralização do DNS é a falta de transparência no mercado de registradores de domínios e provedores de serviços de DNS. As práticas de coleta e uso de dados dessas empresas são frequentemente opacas e mal compreendidas pelos usuários finais."

- "Embora a adoção de tecnologias como o DNSSEC possa melhorar a segurança do DNS, elas não resolvem o problema subjacente da centralização. É necessário um esforço conjunto de governos, organizações da sociedade civil e empresas para promover a descentralização do DNS e garantir a privacidade e a segurança dos usuários."
- "As possíveis soluções incluem a promoção de alternativas descentralizadas ao DNS, como o Namecoin e o Handshake, a adoção de padrões abertos e interoperáveis para o registro de nomes de domínio e a implementação de regulamentações mais rigorosas para empresas que controlam grandes partes do mercado de registradores de domínios e provedores de serviços de DNS."

2.5.4 *DNS: The Hidden Centralization*

Neste relatório de pesquisa (KARLIN; FORD; LIVSHITS; SHMATIKOV, 2019), é analisado como a centralização do DNS está oculta por trás da aparência de descentralização. O relatório destaca como as empresas de tecnologia concentram poder por meio da criação de infraestruturas centralizadas para gerenciar o DNS. Os autores também destacam a importância da governança transparente do DNS e medidas para promover a diversidade de provedores de DNS. Alguns dos trechos que mais se destacam, sob a minha avaliação, em relação ao artigo, são os trechos destacados abaixo:

- "Este artigo examina a centralização do sistema de DNS e seus impactos na segurança cibernética global. A centralização do DNS cria um ponto de falha único e pode tornar a Internet vulnerável a ataques de negação de serviço (DDoS) e outros tipos de ataques cibernéticos."
- "Uma das soluções propostas para a centralização do DNS é a descentralização do sistema através da implementação de tecnologias como o DNSSEC e o DNS-over-TLS, que ajudam a aumentar a segurança e a resiliência do sistema."
- "Além disso, a distribuição geográfica dos servidores DNS e a diversificação dos provedores de serviços de DNS também são essenciais para reduzir a centralização e melhorar a segurança do DNS."
- "É importante que os governos, organizações da sociedade civil e empresas trabalhem juntos para promover a descentralização do DNS e garantir a segurança cibernética global. Isso pode incluir o desenvolvimento de regulamentações mais rigorosas para

empresas que controlam grandes partes do mercado de registradores de domínios e provedores de serviços de DNS, bem como a promoção de alternativas descentralizadas ao DNS."

2.5.5 *Challenges in Internet Governance: The Case of DNS*

Neste artigo (COHN; GURBAXANI, 2020) são abordados alguns dos desafios na governança do DNS em um ambiente cada vez mais centralizado. Os autores exploram as tensões entre a necessidade de segurança e estabilidade do sistema e a necessidade de garantir a transparência e a participação na tomada de decisões.

O artigo também discute possíveis soluções para promover a governança transparente e participativa do DNS, incluindo a criação de fóruns de discussão e a promoção da participação da sociedade civil. Neste artigo, o autor argumenta que a comunidade tem um papel fundamental para ajudar na descentralização do DNS. Finalmente, seguem alguns dos trechos que, em minha avaliação, trazem uma ideia geral e abordam os pontos mais interessantes do artigo:

- "Este artigo discute os desafios associados à centralização do DNS e como a descentralização pode ajudar a resolver esses desafios. A centralização do DNS cria vulnerabilidades de segurança e privacidade, além de permitir que um único ator tenha um controle significativo sobre a Internet."
- "A descentralização do DNS pode ser alcançada através da adoção de alternativas descentralizadas ao DNS, como o Namecoin e o Handshake. Essas alternativas permitem que os usuários registrem nomes de domínio sem a necessidade de intermediários centralizados, reduzindo assim a centralização e melhorando a privacidade e a segurança dos usuários."
- "No entanto, a adoção dessas alternativas ainda é limitada e enfrenta desafios significativos. A interoperabilidade com o DNS existente e a falta de conscientização do público em geral são alguns dos principais obstáculos para a adoção generalizada de alternativas descentralizadas."
- "Para promover a descentralização do DNS, é necessário um esforço conjunto de governos, organizações da sociedade civil e empresas para incentivar a adoção de

alternativas descentralizadas e melhorar a conscientização do público em geral sobre os benefícios da descentralização."

2.6 Escolha do Tema

Dado a relevância do tema, os estudos relacionados e os impactos que os problemas que a concentração/centralização de DNS na Internet podem causar à sociedade, foi escolhido a realização do trabalho de conclusão abordando e aprofundando sobre este tópico. As principais contribuições esperadas são de reunir maiores informações acerca deste fenômeno através da criação de uma solução para medir/verificar o nível de concentração de DNS na Internet (*DNS Measurement*), além de promover maior debate acerca do tema na comunidade e estimular que novos estudos e esforços sejam direcionados ao tema.

No capítulo a seguir, será tratado sobre a ideia da solução e o que a ferramenta se propõe a fazer, abordando aspectos práticos de DNS na Internet.

3 SOLUÇÃO

Um dos objetivos deste estudo é averiguar a existência ou não de uma concentração acerca dos provedores de serviço DNS na Internet. Isto é, verificar se, de fato, poucos provedores de DNS são responsáveis pelo gerenciamento dos registros DNS de muitos domínios na Internet.

Para a coleta de informações que tornasse possível a investigação desse fenômeno, foi construída uma ferramenta específica para este trabalho, chamada de *DNS Measurement.* Neste capítulo do trabalho, será apresentado detalhes em alto nível desta solução, explicando a sua lógica e o seu funcionamento, bem como exemplificando quais foram os dados coletados e como foram coletados.

3.1 Idéia Geral

A ideia de solução foi pensada de forma com que fossem feitas medições periódicas sobre uma lista de domínios populares na Internet (detalhes dessa lista serão abordados posteriormente neste capítulo), sendo mapeados todos os registros DNS de cada domínio. Posteriormente, são identificados os endereços IPs destes registros DNS para que, por fim, seja feio o mapeamento para o provedor de DNS que os registros são gerenciados. Assim, ao final de uma medição, temos a "árvore" completa de DNS do domínio, no que tange o serviço de DNS, partindo do domínio até chegar ao provedor de DNS no qual é responsável pelo gerenciamento dos registros de DNS desse domínio. A imagem abaixo ilustra a ideia do que seria um esquema de uma árvore de DNS, desde o domínio até o mapeamento do provedor responsável pelo mesmo.

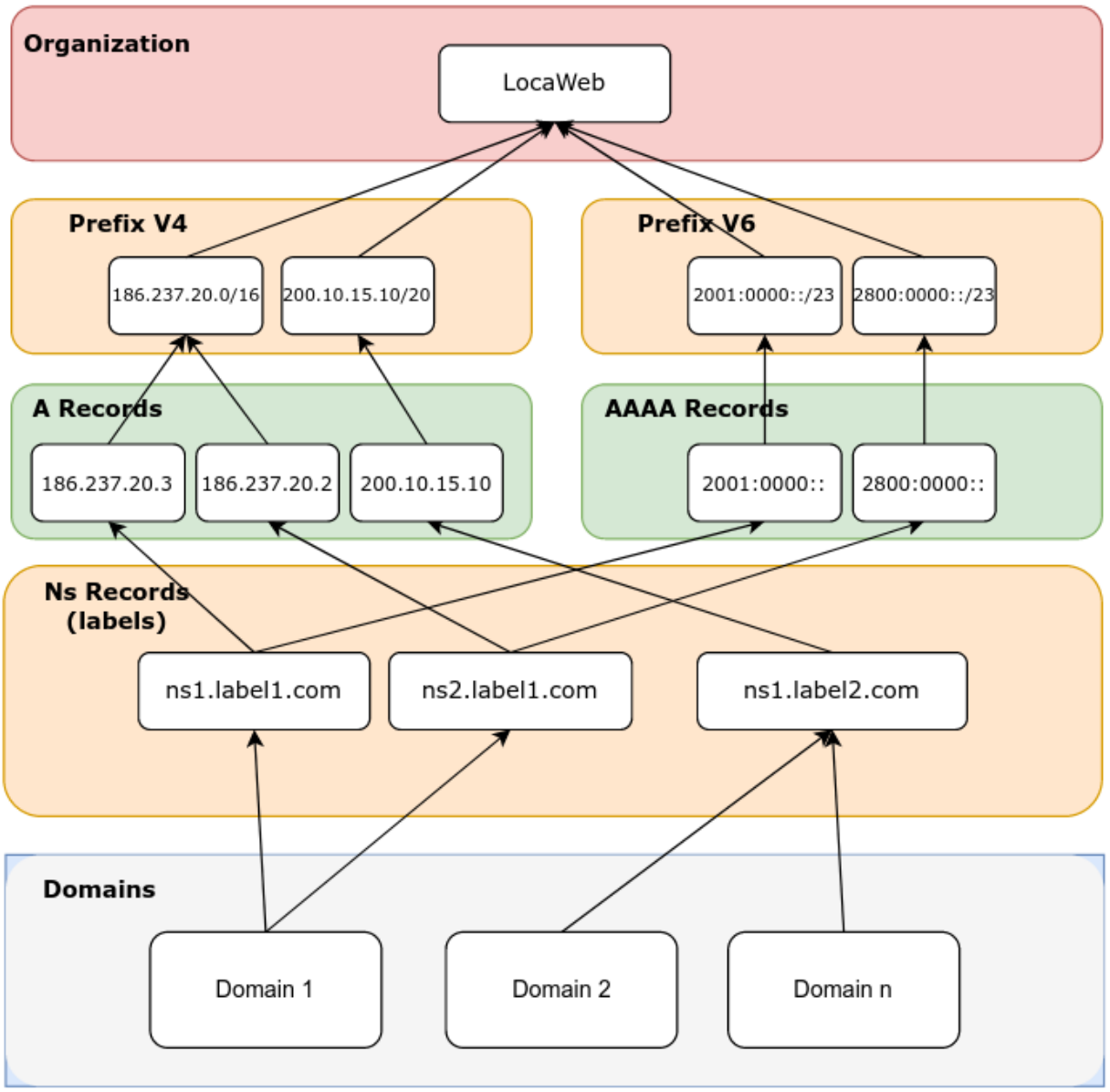

Figura 3.1 - Esquema de uma árvore de medição de provedor de DNS

3.2 Medições

A cada medição, foi construída a árvore de cada domínio pertencente à lista de domínios utilizada. As medições foram feitas com intervalos e afastamento entre si, dentro do período de dezembro de 2022 até março de 2023. Este intervalo foi utilizado para que pudéssemos observar um nível maior de diferenças entre uma medição e outra, ao contrário do que seria observado caso fossem realizadas medições todos os dias.

Abaixo, temos um recorte de uma saída de uma medição, onde pode-se observar as árvores obtidas para alguns domínios da lista. Nela, pode-se visualizar as informações de qual foi o domínio pesquisado, quais são os registros NS associados a estes domínios, os endereços IPs

(tanto em sua versão IPv4 quanto na sua versão IPv6) associados a esses registros NS, e qual é o provedor (ou lista de provedores) de DNS responsável pelo gerenciamento destes registros NS.

```
####################################################
dominio
google.com
lista de NS records
['ns3.google.com.', 'ns4.google.com.', 'ns1.google.com.', 'ns2.google.com.']
lista de A records
['216.239.36.10', '216.239.38.10', '216.239.32.10', '216.239.34.10']
lista de AAAA records
['2001:4860:4802:36::a', '2001:4860:4802:38::a', '2001:4860:4802:32::a', '2001:4860:4802:34::a']
lista de empresas
['GOOGLE']
####################################################
dominio
cloudflare.com
lista de NS records
['ns4.cloudflare.com.', 'ns5.cloudflare.com.', 'ns6.cloudflare.com.', 'ns7.cloudflare.com.', 'ns3.cloudflare.com.']
lista de A records
['162.159.1.33', '162.159.8.55', '162.159.2.9', '162.159.9.55', '162.159.5.6', '162.159.3.11', '162.159.4.8', '162.159.6.6', '162.159.7.226', '162.159.0.33']
lista de AAAA records
['2400:cb00:2049:1::a29f:837', '2400:cb00:2049:1::a29f:121', '2400:cb00:2049:1::a29f:209', '2400:cb00:2049:1::a29f:937', '2400:cb00:2049:1::a29f:506', '2400:cb00:2049:1::a29f:30b', '2400:cb00:2049:1::a29f:408',
'2400:cb00:2049:1::a29f:606', '2400:cb00:2049:1::a29f:21', '2400:cb00:2049:1::a29f:7e2']
lista de empresas
['CLOUDFLARENET']
####################################################
dominio
googletagmanager.com
lista de NS records
['ns4.google.com.', 'ns3.google.com.', 'ns1.google.com.', 'ns2.google.com.']
lista de A records
['216.239.38.10', '216.239.36.10', '216.239.32.10', '216.239.34.10']
lista de AAAA records
['2001:4860:4802:38::a', '2001:4860:4802:36::a', '2001:4860:4802:32::a', '2001:4860:4802:34::a']
lista de empresas
['GOOGLE']
```

Figura 3.2 - Recorte de uma saída de medição, com árvores de alguns domínios

Além da árvore completa do domínio, outra saída de cada medição bastante importante para este trabalho é a contagem de quantos domínios cada provedor de DNS é responsável pelo gerenciamento. Ao final de cada medição e da construção das árvores dos domínios, são contados quantos domínios cada provedor de DNS é responsável por gerenciar os registros NS, para que, assim, possa-se visualizar o número de domínios que cada provedor de DNS gerencia. Posteriormente, o número de domínios associados aos provedores é dividido pelo total de domínios medidos (este total de domínios e mais detalhes acerca da lista de domínios serão detalhados posteriormente), e, assim, ter-se a concentração de domínios gerenciados (no que tange o serviço de DNS) por provedor.

A imagem abaixo ilustra essa saída, onde tem-se, respectivamente e a cada linha, o nome do provedor de DNS, o número de domínios da lista medida nos quais tem seus registros NS gerenciados pelo provedor respectivo e, por fim, a concentração (em %) deste provedor, obtido através do resultado da divisão entre o número de domínios sob sua responsabilidade e do número total de domínios medidos (esses itens estão separados por “:”). Essa saída está organizada de tal forma com que os provedores de DNS responsáveis pelo maior número de domínios fiquem acima (ordem decrescente).

```
['CLOUDFLARENET']:207156:20.7156
['AMAZON-02']:72281:7.2281
['GODADDY-DNS']:30448:3.0448
['ALIBABA-CN-NET']:14211:1.4211
['GOOGLE']:9486:0.9486
['TIGGEE']:8138:0.8138
['MICROSOFT-CORP-MSN-AS-BLOCK']:7927:0.7927
['NSONE']:5660:0.566
['IONOS-AS']:4568:0.4568
['OVH']:3922:0.3922
['AS-REGRU']:3432:0.3432
['GANDI-LIVEDNS']:3299:0.3299
['NICAT']:3287:0.3287
['IONOS-AS', 'FASTHOSTS-INTERNET']:2999:0.2999
['ORACLE-BMC-31898']:2957:0.2957
['CLOUDNSNET']:2527:0.2527
['MYLOC-AS']:2153:0.2153
['ASEPL-AS-AP', 'ALIBABA-CN-NET']:2134:0.2134
['TimeWeb-AS']:1814:0.1814
['WYNETWORK']:1531:0.1531
['']:1518:0.1518
['ONECOM']:1390:0.139
['TRANSIP-AS']:1325:0.1325
['AUTOMATTIC']:1199:0.1199
['NUCDN']:1157:0.1157
['SELECTEL']:1154:0.1154
['TENCENT-NET-AP', 'CHINANET-SHANGHAI-MAN']:1140:0.114
['LIQUIDWEB']:1037:0.1037
['DIGITALOCEAN-ASN']:869:0.0869
['TUCOWS-3', 'TUCOWS', 'AS-TING-BACKBONE']:860:0.086
```

Figura 3.3 - Recorte de uma saída de medição, com a concentração por provedor

3.3 Lista de Domínios

Conforme dito anteriormente, chegou o momento de explorar a lista de domínios utilizada nas medições. Este é, com certeza, um fator importante para o seguimento do estudo, uma vez que a metodologia da solução utilizada neste trabalho foi através de medições partindo de uma lista de domínios, coletar seus dados acerca de seus provedores de serviço DNS e, posteriormente, tirar conclusões sobre o cenário de provedores de DNS na Internet. Portanto, faz bastante sentido este assunto ser abordado e explorado.

Existem algumas listas de domínios, elaboradas e ranqueadas por diferentes critérios e metodologias. Alguns exemplos dessas listas são:

- Alexa Top 1M (ALEXA, 2018)
- Majestic Million (MAJESTIC, 2009)
- SimilarWeb (SIMILARWEB, 2007)

A lista de domínios escolhida e utilizada para a realização das medições, entretanto, foi a lista da iniciativa “Tranco” (TRANCO, 2023). "Tranco" é um projeto que fornece uma lista regularmente atualizada dos top um milhão de sites na internet, com base em sua popularidade e tráfego. A lista é compilada usando várias fontes de dados, incluindo Alexa, SimilarWeb e Moz, e tem como objetivo fornecer informações para pesquisadores, desenvolvedores e outros interessados em compreender a popularidade e uso relativo de diferentes sites na internet (TRANCO, 2023). A meta do Tranco é fornecer uma classificação abrangente, transparente e confiável de sites, além de promover uma melhor compreensão de como a internet está sendo utilizada.

O artigo *"A Research-Oriented Top Sites Ranking Hardened Against Manipulation"* (TRANCO, 2019), da Tranco, descreve uma nova metodologia para criar uma lista de sites da web com o objetivo de ser mais resistente à manipulação e ao viés do que outras listas populares existentes, como Alexa, Quantcast, Majestic Million e SimilarWeb. Algumas das principais informações e tópicos abordados no artigo incluem:

1. **Metodologia da lista Tranco:** o artigo descreve a metodologia usada pela Tranco para compilar sua lista de sites da web. Isso inclui a coleta de dados de vários provedores de serviços DNS e a aplicação de técnicas estatísticas para identificar e remover domínios falsos e mortos.
2. **Diferenças em relação a outras listas populares:** o artigo compara a lista Tranco com outras quatro listas populares (Alexa, Quantcast, Majestic Million e SimilarWeb) em termos de tamanho, cobertura geográfica, grau de exclusão de domínios falsos e mortos e outros critérios.
3. **Resiliência contra manipulação:** um dos principais objetivos da lista Tranco é ser mais resistente à manipulação e ao viés do que outras listas populares. O artigo descreve várias técnicas que a Tranco usa para atingir esse objetivo, incluindo a inclusão de vários provedores de serviços DNS e a exclusão de domínios que mostram comportamento suspeito.

4. **Resultados experimentais:** o artigo apresenta os resultados de vários experimentos que mostram que a lista Tranco é menos suscetível a manipulação e viés do que outras listas populares. Esses experimentos incluem a adição de bots falsos aos dados de tráfego de sites e a comparação dos resultados da Tranco com outras listas populares.
5. **Uso da lista Tranco:** o artigo discute possíveis usos da lista Tranco, incluindo a avaliação da segurança e estabilidade da Internet, análise de tendências de tráfego na web e detecção de atividades maliciosas, como botnets e ataques de phishing.

O ranking utilizado foi criteriosamente escolhido para que o estudo deste trabalho fosse conduzido com os principais domínios da Internet, pois, quanto mais relevantes forem os domínios escolhidos para as medições, maior a relevância e importância de avaliarmos uma possível concentração de provedores de serviço DNS sobre eles.

A cada medição realizada, era realizado o *Download* da mais nova e atualizada lista da Tranco, de modo com que as pesquisas sempre eram feitas com o ranking dos um milhão de domínios principais na Internet para a data em questão, de acordo com a Tranco. Abaixo, segue imagens de recortes de 2 exemplos de listas de domínios da Tranco. A primeira lista, referente ao dia 16 de Dezembro de 2022, e a segunda referente ao dia 13 de Fevereiro de 2023.

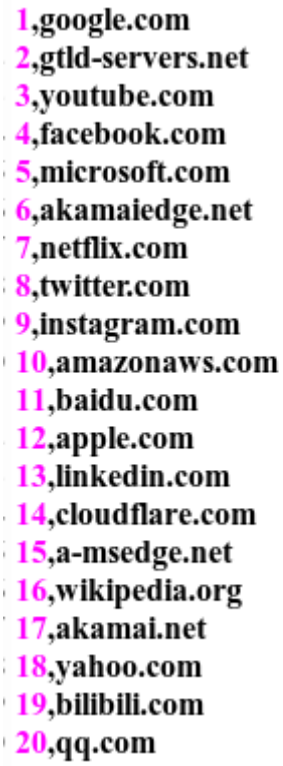

Figura 3.4 - Recorte do topo da Tranco-list, dia 16/12/2022

Figura 3.5 - Recorde do topo da Tranco-list, dia 13/02/2023

Como ficou evidenciado, em ambas as listas, tanto a do dia 16 de Dezembro de 2022 como a do dia 13 de Fevereiro de 2023, os primeiros 20 domínios ranqueados como principais são exatamente os mesmos. Avançando um pouco mais na lista e indo para o final, pode-se também verificar que este comportamento permanece, conforme imagens abaixo:

999980,broxus.com
999981,cactusmusic.ir
999982,sport4do.com
999983,manahosting.ca
999984,fishingcall.fit
999985,vuonthuocquy.vn
999986,roter-renner.de
999987,hew.com
999988,itechwebhosting.com
999989,hoborr.com
999990,monkeyjungle.com
999991,phlebotomyusa.com
999992,bergstrasse-odenwald.de
999993,msoms-anime.net
999994,wordstotime.com
999995,soekgai.com
999996,eurodownload.com
999997,thesoutheasternchannel.com
999998,heranet.info
999999,slot78vip.site
1000000,famemagazine.co.uk

Figura 3.6 - Recorte do final da Tranco-list, dia 16/12/2022

Figura 3.7 - Recorte do final da Tranco-list, dia 13/02/2023

Este fato chamou a atenção e, para a comparação entre as listas de domínios, foi construído um *script* para verificar se existiam diferenças entre os rankings da Tranco para os dois dias. Surpreendentemente, os rankings para as duas datas são exatamente iguais, tanto nos nomes de domínios, quanto para a ordem onde eles aparecem. Ou seja, não houve mudanças significativas, segundo os critérios da Tranco, para que a lista fosse alterada, mesmo em um intervalo de 02 meses.

Nas Figuras 3.8 e 3.9, estão ilustrados o *script* construído para a checagem e a saída do terminal. Como nenhum domínio foi impresso na tela, tem-se que o arquivo inteiro é exatamente igual. A metodologia utilizada foi a de ler as listas, linha por linha, e verificar se os domínios eram iguais para a linha em questão.

```
# Abrindo os arquivos de texto
arquivo1 = open("top-1m-16.csv", "r")
arquivo2 = open("top-1m-13.csv", "r")

# Lendo as linhas dos arquivos
linhas1 = arquivo1.readlines()
linhas2 = arquivo2.readlines()

# Comparando as linhas dos arquivos
for i in range(len(linhas1)):
    linha1 = linhas1[i].strip()
    linha2 = linhas2[i].strip()

    if linha1 != linha2:
        print (linha1, linha2)


# Fechando os arquivos de texto
arquivo1.close()
arquivo2.close()
```

Figura 3.8 - Script para comparação entre listas de domínios da Tranco

```
Demetrio@localhost:~/Documentos/TCC$ python3 comparadominios
Demetrio@localhost:~/Documentos/TCC$
```

Figura 3.9 - Retorno do terminal, indicando que não houve diferenças entre os arquivos

3.4 Mapeamento de provedores de DNS

Outra informação bastante relevante é, a partir destes domínios, extrair qual é o provedor de DNS responsável pelos registros NS dos domínios. Antes de aprofundar em como foi obtida esta informação, é importante ter em mente a definição do que é um AS (sistema autônomo).

Um Sistema Autônomo (AS), é uma rede de dispositivos de computação interconectados, que operam sob uma única política de roteamento e são administrados por uma única entidade, geralmente uma organização ou provedor de serviços de internet. O AS é identificado por um número de sistema autônomo (ASN) único atribuído pelo Registro de Números da Internet (IANA), e é usado para facilitar o roteamento de tráfego na Internet.

Pois bem, com o conceito de um AS já definido, pode-se aprofundar no método utilizado para, a partir de uma lista de domínios, mapear a qual provedor de DNS está responsável por manter os registros NS destes domínios. Para isso, foram utilizadas 2 tabelas fornecidas pela CAIDA (CAIDA, 2023).

CAIDA significa Centro de Análise e Disseminação de Informação de Internet (em inglês, Center for Applied Internet Data Analysis). É uma organização de pesquisa localizada na Universidade da Califórnia em San Diego, nos Estados Unidos. O objetivo principal do CAIDA é coletar e analisar dados sobre a Internet e seus comportamentos, bem como desenvolver ferramentas e metodologias para melhorar a compreensão e o desempenho da rede (CAIDA, 2023).

As tabelas utilizadas para a obtenção das informações desejadas foram as de mapeamento de prefixos de rede para AS (CAIDA-IP2AS, 2023), e também a tabela de mapeamento de AS para os provedores responsáveis por eles (CAIDA-AS2ORG, 2023).

As imagens abaixo ilustram as tabelas utilizadas para as medições referidas acima:

1.0.0.0	24	13335
1.0.4.0	22	38803
1.0.5.0	24	38803
1.0.16.0	24	2519
1.0.32.0	24	141748
1.0.64.0	18	18144
1.0.128.0	17	23969
1.0.128.0	18	23969
1.0.128.0	19	23969
1.0.128.0	24	23969
1.0.129.0	24	23969
1.0.130.0	23	23969
1.0.132.0	24	23969
1.0.133.0	24	23969
1.0.136.0	24	23969
1.0.137.0	24	23969
1.0.138.0	24	23969
1.0.139.0	24	23969
1.0.141.0	24	23969

Figura 3.10 - Recorte da tabela de Prefixos de Rede IPv4 para AS

Na Figura 3.10, temos, respectivamente e a cada linha, o prefixo da rede, a quantidade de bits (menor unidade de informação que um computador consegue armazenar e processar) que devem ser considerados para o prefixo da rede e o AS associado a esta rede. A Figura 3.11 é análoga a Figura 3.10, porém com o prefixo de rede no formato de IPv6.

2001:250:205::	48	24349
2001:250:206::	48	24351
2001:250:207::	48	24349
2001:250:208::	48	24349
2001:250:209::	48	24348
2001:250:20a::	48	24351
2001:250:20b::	48	24350
2001:250:20c::	48	24349
2001:250:20d::	48	24348
2001:250:20f::	48	24351
2001:250:210::	48	24349
2001:250:212::	48	24349
2001:250:214::	48	24351
2001:250:215::	48	24349
2001:250:217::	48	24348
2001:250:218::	48	24349
2001:250:219::	48	24348
2001:250:21a::	48	24348
2001:250:21b::	48	24348
2001:250:21d::	48	24348
2001:250:21e::	48	24349

Figura 3.11 - Recorte da tabela de Prefixos de Rede IPv6 para AS

42660||TRL-AS|ORG-TRLS1-RIPE||RIPE
42661||DETSKYMIR-AS|ORG-PDM1-RIPE||RIPE
42662||REGA-AS|ORG-SR20-RIPE||RIPE
42663||LATVIJASTALRUNIS|ORG-SLT3-RIPE||RIPE
42664||ipishnik-as|ORG-JR12-RIPE||RIPE
42665||SIRTI-AS|ORG-SORG1-RIPE||RIPE
42666||EMBOU|ORG-ENT1-RIPE||RIPE
42667||InterEvo-PL-AS|ORG-VA95-RIPE||RIPE
42668||NEVALINK-AS|ORG-NL486-RIPE||RIPE
42669||MEGAWEB_IT_BIELLA|ORG-CS76-RIPE||RIPE
42670||UNDERLAN-AS|ORG-PPEV1-RIPE||RIPE
42671||SANDYNET|ORG-ZPL2-RIPE||RIPE
42672||SPEDIMEX|ORG-SSZO81-RIPE||RIPE
42673||SKYWARE-AS|ORG-SSZO46-RIPE||RIPE
42674||BANKSOYUZ-AS|ORG-JSBS1-RIPE||RIPE

Figura 3.12 - Recorte da tabela de AS para organização/provedor

Na Figura 3.12, tem-se, de informação relevante para a medição, o número AS e a organização correspondendo a ele, logo à direita. Além disso, nesta mesma tabela, temos uma informação bastante importante e que será utilizada para ajudar a responder algumas perguntas de pesquisa posteriormente, que é o país e região de origem de alguns provedores de DNS. A Figura 3.13 ilustra um recorte dessa parte da tabela. Nele, tem-se exemplos de provedores de Taiwan e também da Indonésia.

Haosing Technology Co., Ltd.|TW|APNIC
GRONEXT CO., LTD.|TW|APNIC
FANTASY TECHNOLOGY CO., LTD.|TW|APNIC
WantEasy Info. Co, Ltd.|TW|APNIC
HTC Corp|TW|APNIC
Beyond Orbit Co., Ltd.|TW|APNIC
GUAN TING CO., LTD.|TW|APNIC
OpenFor Telecommunications Co., Ltd.|TW|APNIC
DQWLIT|TW|APNIC
Gangu Tech Co., Ltd.|TW|APNIC
Compal Electronics|TW|APNIC
InnoLux Display Corp.|TW|APNIC
PT. Net2Cyber Indonesia|ID|APNIC
Kementerian Agama (KEMENAG)|ID|APNIC
PT Jabikha Teknologi Indonesia|ID|APNIC
PT. KAPANLAGI.COM NETWORKS|ID|APNIC
PT Qeon Interactive|ID|APNIC
PT Samudera Indonesia Tbk.|ID|APNIC

Figura 3.13 - Recorte da tabela onde tem-se o provedor e seu país/região de origem

Assim, a cada medição, além das listas de domínios atualizadas, fornecida pela Tranco, também era realizado o *Download* das 3 tabelas da CAIDA atualizadas, indicadas acima. A justificativa para isso é justamente ter uma medição atualizada. Ou seja, caso o registro NS de algum domínio fosse alterado, ou o registro fosse mantido o mesmo, mas o provedor responsável por ele fosse alterado, isso seria evidenciado pela medição. Abaixo, segue uma imagem que ilustra o Download dos itens indicados a cada execução da medição.

```
Demetrio@localhost:~/Documentos/TCC$ python3 Medicao.py
Downloading Tranco List

100% [..............................................................] 9799670 / 9799670
Downloading Prefix2ASIPv4

100% [..............................................................] 3368337 / 3368337
Downloading Prefix2ASIPv6

100% [..............................................................] 658694 / 658694
Downloading AS2Organization

100% [..............................................................] 3655050 / 3655050
```

Figura 3.14 - Execução da medição e *Download* dos itens comentados

Após o *Download* de todos esses itens, já tem-se todas as informações necessárias para que a árvore de cada domínio seja montada, conforme mostrado no início deste capítulo. O fluxo de execução para obter-se a árvore de cada domínio se dá da seguinte maneira: à partir da lista de domínios, para cada domínio da lista, é realizada uma busca (detalhes da implementação dessa busca serão discutidos no capítulo de implementação) para mapear os registros NS desse domínio. Posteriormente, busca-se descobrir quais são os IP 's correspondentes a cada registro NS do domínio. Sabendo quais são os IP’s dos registros NS, busca-se na lista da CAIDA qual é o AS correspondente a este IP e, finalmente, na lista de mapeamento de AS para organização, também da CAIDA, qual é o provedor de DNS responsável pelo AS. Este processo é repetido para todos os um milhão de domínios da lista da Tranco, e, assim, ao final da execução, tem-se a árvore de todos os principais domínios da Internet para a data de medição.

4 IMPLEMENTAÇÃO

Neste capítulo, será mostrado e detalhado um pouco mais sobre os detalhes de implementação da ferramenta construída para este trabalho: o DNS Measurement. Esta ferramenta é a responsável pela realização das medições, construções das árvores de DNS e apuração da concentração de provedores de DNS nos domínios medidos, conceitos que foram apresentados nos capítulos anteriores do trabalho. Além de aprofundar detalhes acerca da solução propriamente dita, serão também abordados aspectos do armazenamento dos dados obtidos com as medições e da arquitetura do sistema utilizado para a execução da ferramenta,

Na seção 4.1, será detalhado os aspectos do DNS Measurement, como a linguagem de programação e as bibliotecas utilizadas. Na seção 4.2, será comentado sobre a arquitetura e *capacity* do sistema de execução do DNS Measurement e como foi estruturada a execução do mesmo. Na seção 4.3, será o momento para a abordagem sobre os dados obtidos e o seu armazenamento em banco de dados.

4.1 DNS Measurement

A seguir, serão abordados os principais aspectos técnicos acerca da construção da solução.

4.1.1 Linguagem de programação

A ferramenta utilizada neste trabalho foi construída de forma a executar alguns passos que são listados a seguir:

- Realizar o *download* das listas atualizadas da Tranco e da Caida (exploradas anteriormente),
- Iterar sobre a lista dos principais domínios da Tranco e construir a árvore de DNS correspondente a cada domínio.
- Por fim, medir a concentração por provedor de DNS.

Para a construção dos algoritmos responsáveis por essas tarefas a linguagem de programação escolhida foi o Python (PYTHON, 2023), em sua versão 3, mais especificamente a versão 3.9.2.

Os motivos para a escolha do uso desta linguagem de programação foram a simplicidade e familiaridade com a linguagem, o desejo de se fazer as medições com um *script* (conjunto de

instruções) executado diretamente pelo terminal do sistema operacional (S.O.) (sem a necessidade de geração de um programa executável), além de, principalmente, a existência de algumas bibliotecas bastante úteis para a realização das medições, como a biblioteca *dnspython* (PYPI, 2023) e a biblioteca *concurrent.futures* (PYTHON, 2023).

4.1.2 Bibliotecas

- *dnspython* - é uma biblioteca Python para consulta e manipulação de registros de DNS. O objetivo principal da biblioteca é permitir que os desenvolvedores escrevam *scripts* e aplicativos em Python que possam interagir com servidores DNS para obter informações sobre nomes de domínio, como endereços IP ou registros NS. Para a ferramenta DNS Measurement, o módulo da biblioteca utilizado foi o *dns.resolver*, no qual recebe um domínio e um parâmetro. Para o DNS Measurement, os parâmetros utilizados foram:
    1. NS - para consulta aos registros de DNS do domínio. No trecho abaixo, a variável "NS" receberá os registros DNS do domínio.

        NS = dns.resolver.resolve(dominio, 'NS')
    2. A - para consulta sobre os endereços IPv4 dos registros NS. No trecho abaixo, a variável “query” receberá o endereço IPv4 do registro de DNS solicitado.

        query = dns.resolver.resolve(nsrecord, 'A')
    3. AAAA - para consulta sobre os endereços IPv6 dos registros NS. No trecho abaixo, a variável “query” receberá o endereço IPv6 do registro de DNS solicitado.

        query = dns.resolver.resolve(nsrecord, 'AAAA’)

- *concurrent.futures* - A biblioteca concurrent.futures é uma biblioteca nativa do Python que fornece uma interface para execução concorrente de tarefas em paralelo usando processamento em thread ou em processo. O objetivo principal da biblioteca é facilitar a execução de tarefas assíncronas em Python. O funcionamento da biblioteca é relativamente simples: ela fornece duas classes principais, *ThreadPoolExecutor* e *ProcessPoolExecutor,* que permitem que o desenvolvedor execute funções em threads ou processos separados, respectivamente (PYTHON, 2023). Na implementação do DNS Measurement, foi utilizada a classe *ProcessPoolExecutor* para a execução paralela da função na qual gerava as árvores dos domínios. Esta foi fundamental, tendo em vista que a iteração é sobre um milhão de domínios

e, sem a utilização dela, o tempo de execução do DNS Measurement ultrapassa as 24h. A Figura 4.1 é o trecho do *script* onde é utilizado a biblioteca.

```
with concurrent.futures.ProcessPoolExecutor(max_workers=200) as executor:
    for result in executor.map(querydominio, Domainlist):
        pass
```

Figura 4.1 - Utilização da biblioteca *concurrent.futures* no DNS Measuring

Neste trecho, a classe *ProcessPoolExecutor* inicializa 200 processos paralelos. Para cada um desses processos, chamados de "executor", é realizada a chamada da função que constrói a árvore (que será detalhada posteriormente), passando a lista de domínios baixados da lista da Tranco como parâmetro.

4.1.3 Funções

- *listadominios* - Função utilizada para, a partir do arquivo de lista Tranco de domínios, criar uma lista Python (estrutura de dados nativa da linguagem Python) (PYTHONACADEMY, 2021) com esses domínios, normalizando o arquivo para salvar na lista Python somente os nomes de domínio, excluindo outras informações que possam atrapalhar, como números. A Figura 4.2 ilustra o funcionamento da função.

```
def listadominios():

    arquivo = glob.glob('*.csv')
    arquivo = str(arquivo)
    arquivo = arquivo[:-2]
    arquivo = arquivo[2:]


    with open (arquivo, 'r') as dominios:
        for dominio in dominios:
            dominio = ''.join([i for i in dominio if not i.isdigit()])
            dominio = dominio[1:]
            Domainlist.append(dominio)
```

Figura 4.2 - Função listadominios

- *criahash* - Função utilizada para criação de tabelas *hash* (estrutura de dados utilizada para armazenar e recuperar valores a partir de uma chave e de forma eficiente, com complexidade temporal) (UNICAMP, 2003), através dos dicionários da linguagem Python (PYTHONACADEMY, 2021), a fim de otimizar as consultas que serão realizadas às tabelas de tradução de prefixo de rede para AS e AS para provedor de DNS, da Caida (mais detalhes sobre estas tabelas estão descritas na seção 3.3), que são baixadas no começo da execução da ferramenta (mais detalhes sobre como elas são baixadas serão comentados na seção 4.1.4). A Figura 4.3 ilustra o funcionamento da função.

```
def criahash():

    arquivo = glob.glob('routeviews-rv2*')
    arquivo = str(arquivo)
    arquivo = arquivo[:-2]
    arquivo = arquivo[2:]


    with open (arquivo, "r") as ip:
        for line in ip:
            token = line.split('  ')
            rede = token[0]
            asn = token[2]
            iphash[rede] = asn

    #os.remove(arquivo)

    arquivo = glob.glob('routeviews-rv6*')
    arquivo = str(arquivo)
    arquivo = arquivo[:-2]
    arquivo = arquivo[2:]

    with open (arquivo, "r") as ip6:
        for line in ip6:
            token = line.split('  ')
            rede = token[0]
            asn = token[2]
            ip6hash[rede] = asn

    #os.remove(arquivo)

    arquivo = glob.glob('*.as-org2info.txt')
    arquivo = str(arquivo)
    arquivo = arquivo[:-2]
    arquivo = arquivo[2:]


    with open (arquivo, "r") as asn:
        for line in asn:
            token = line.split('|')
            try:
                asn = token[0]
                org = token[2]
                asnhash[asn] = org
            except:
                pass
    #os.remove(arquivo)
```

Figura 4.3 - Função criahash

- *getArecords* - Função criada para, dado um registro NS passado à função por parâmetro, retornar o endereço IPv4 associado a este registro. NS, através da biblioteca *dnspython*, utilizando o módulo *dns.resolver*. A Figura 4.4 ilustra o funcionamento da função.

```
def getArecords (nsrecord):
    nsrecord = nsrecord[:-1]
    Arecord = []
    try:
        query = dns.resolver.resolve(nsrecord, 'A')
        for item in query:
            Arecord.append(item.to_text())
        return Arecord

    except:
        pass
```

Figura 4.4 - Função getArecords

- *getAAAArecords* - Função criada para, dado um registro NS passado à função por parâmetro, retornar o endereço IPv6 associado a este registro NS, através da biblioteca *dnspython*, utilizando o módulo *dns.resolver*. A Figura 4.5 ilustra o funcionamento da função.

```
def getAAAArecords (nsrecord):
    nsrecord = nsrecord[:-1]
    AAAArecord = []
    try:
        query = dns.resolver.resolve(nsrecord, 'AAAA')
        for item in query:
            AAAArecord.append(item.to_text())
        return AAAArecord

    except:
        pass
```

Figura 4.5 - Função getAAAArecords

- *querydominio* - Função principal criada para a construção das árvores de domínio. Esta é a função na qual aparece na Figura 4.1, executada através de processos com paralelismo pela biblioteca *concurrent.futures*. A função recebe um domínio da lista de domínios da Tranco, realiza a consulta DNS para conhecer os registros DNS do domínio (através da biblioteca *dns.resolver*. Após isso, itera sobre os registros NS para conhecer seus respectivos endereços IPv4 e IPv6, utilizando as funções *getArecords* e *getAAAArecords*. Em posse dos endereços IPs de todos os registros NS do domínio, são consultadas as tabelas *hash* criadas na função *criahash* para verificar, assim, qual ou quais provedores de DNS são responsáveis pelo gerenciamento dos registros de DNS do domínio. Posteriormente, salva todas informações coletadas (registros de DNS, endereços IP destes registros e o provedor ou lista de provedores de DNS responsáveis pelo domínio) em um arquivo de texto denominado como "Medição dia xxxx", onde xxxx é a data na qual foi executada a medição. As figuras a seguir ilustram o funcionamento da função.

```
def querydominio (dominio):
    domainNSList = []
    Arecords = []
    AAAArecords = []
    dominio = dominio[:-1]
    listaempresas = []
    tempA = []
    tempAAAA = []

    try:
        NS = dns.resolver.resolve(dominio, 'NS')
        for answer in NS.response.answer:
            for item in answer.items:
                domainNSList.append(item.to_text())


        for nsrecord in domainNSList:
            tempA = getArecords(nsrecord)
            for ip in tempA:
                Arecords.append(ip)
            tempAAAA = getAAAArecords(nsrecord)
            for ip in tempAAAA:
                AAAArecords.append(ip)
```

Figura 4.6 - Função querydominio, primeira parte

```
for line in Arecords:
        token = line.split('.')
        token[3] = 0
        mascara = str(token[0]) + "." + str(token[1]) + "." + str(token[2]) + "." + str(token[3])
        try:
                empresa = asnhash[iphash[mascara][:-1]]
                if empresa not in listaempresas:
                        listaempresas.append(empresa)

        except:
                pass

for line in AAAArecords:
        token = line.split(':')
        mascara = str(token[0]) + ":" + str(token[1]) + ":" + str(token[2]) + "::"
        try:
                empresa = asnhash[ip6hash[mascara][:-1]]
                if empresa not in listaempresas:
                        listaempresas.append(empresa)
        except:
                pass


if len(listaempresas) != 0:
        try:
                print (dominio, domainNSList, Arecords, AAAArecords, listaempresas)

                listaempresas = str(listaempresas)
                with open ('Medição dia' + str(date.today()), 'a') as file:
                                file.write('#########################################################')
                                file.write("\n")
                                file.write("dominio \n")
                                file.write(str(dominio))
                                file.write("\nlista de NS records \n")
                                file.write(str(domainNSList))
                                file.write("\nlista de A records \n")
                                file.write(str(Arecords))
                                file.write("\nlista de AAAA records \n")
                                file.write(str(AAAArecords))
                                file.write("\nlista de empresas \n")
                                file.write(str(listaempresas))
                                file.write("\n")
                                file.close()

        except:
                pass
```

Figura 4.7 - Função querydominio, segunda parte

- *concentracao* - É a última função da ferramenta, utilizada para medir o número de domínios que cada provedor DNS é responsável, ou seja, a concentração de cada provedor de DNS. A Figura 4.8 ilustra o funcionamento da função.

```
def concentracao():

    concentracao = {}
    controle = 1

    with open ('Medição dia' + str(date.today()), 'r') as arquivo:
        for line in arquivo:
            line = str(line)
            if controle == 11:
                if line in concentracao.keys():
                    concentracao[line] = concentracao[line] + 1
                else:
                    concentracao[line] = 1
                controle = 1
            else:
                controle = controle + 1



    for item in sorted (concentracao, key = concentracao.get, reverse=True):
        with open ('Concentracao dia' + str(date.today()), 'a') as file:
            porcentagem = concentracao[item]/10000
            file.write(item.strip() + ':' + str(concentracao[item]).strip() + ':' + str(porcentagem) + '\n')
```

Figura 4.8 - Função concentracao

4.1.4 Downloader

Separado do script de coleta das informações da infraestrutura DNS, desenvolvemos também, um *script* em Python nomeado de *Downloader*. O mesmo é utilizado para realização do *Download* e descompactação das listas atualizadas da Tranco e das tabelas de tradução atualizadas da Caida. Este *script* é executado sempre na primeira etapa da ferramenta *DNS Measurement,* conforme mostra a Figura 3.11. A Figura 4.9 detalha o código fonte construído para esse *script*.

```
month = datetime.now().month
year = datetime.now().year

def listFD(url, ext=''):
    page = requests.get(url).text
    soup = BeautifulSoup(page, 'html.parser')
    return [url + '/' + node.get('href') for node in soup.find_all('a') if node.get('href').endswith(ext)]

def gunzip_shutil(source_filepath, dest_filepath, block_size=65536):
    with gzip.open(source_filepath, 'rb') as s_file, \
            open(dest_filepath, 'wb') as d_file:
        shutil.copyfileobj(s_file, d_file, block_size)

url_prefix2as_ipv4 = f'https://publicdata.caida.org/datasets/routing/routeviews-prefix2as/{year}/0{month}/'
ext = 'gz'

url_prefix2as_ipv6 = f'https://publicdata.caida.org/datasets/routing/routeviews6-prefix2as/{year}/0{month}/'
ext = 'gz'

result = listFD(url_prefix2as_ipv4, ext)
prefix2as_ipv4 = result[-1]

result = listFD(url_prefix2as_ipv6, ext)
prefix2as_ipv6 = result[-1]

url_as2organization = f'https://publicdata.caida.org/datasets/as-organizations/'
ext_as2organization = 'txt.gz'

result = listFD(url_as2organization, ext_as2organization)
as2organization = result[-1]

tranco = 'https://tranco-list.s3.amazonaws.com/tranco_VXYGN-1m.csv.zip'

print('Downloading Tranco List\n')
wget.download(tranco)
print('\nDownloading Prefix2ASIPv4\n')
wget.download(prefix2as_ipv4)
print('\nDownloading Prefix2ASIPv6\n')
wget.download(prefix2as_ipv6)
print('\nDownloading AS2Organization\n')
wget.download(as2organization)


print('Unzipping Tranco List')
with zipfile.ZipFile("tranco_VXYGN-1m.csv.zip","r") as zip_ref:
    zip_ref.extractall(".")

os.remove("tranco_VXYGN-1m.csv.zip")

files_gz = glob.glob('*.gz')
```

Figura 4.9 - *Script Downloader*

4.2 Arquitetura Sistêmica e Ambiente

Em um primeiro momento, foi pensando que o ideal para a execução do DNS Measurement fosse um servidor do próprio instituto de informática da UFRGS (UFRGS, 2023), por motivos de capacidade de máquina (ou seja, melhor hardware em relação a minha máquina pessoal), e também por questões de disponibilidade e agendamento automático da execução dos scripts, uma vez que os servidores estão sempre ligados. O primeiro servidor da UFRGS disponibilizado para a execução da ferramenta foi o ico.inf.ufrgs.br. Os arquivos do DNS Measurement foram todos transferidos para o servidor, através da realização de uma conexão via SSH (Secure Shell, protocolo de rede), a partir de minha máquina pessoal até o servidor. O sistema operacional do servidor era um Linux, na distribuição RedHat (REDHAT, 2023). Foram realizadas algumas medições a partir deste servidor. Entretanto, na maioria das vezes, a execução era finalizada sem sucesso, devido ao número excessivo de requisições de DNS enviadas em um curto intervalo de tempo. Isso ocasionou na interrupção da execução da solução pelos administradores da rede do instituto de informática da UFRGS.

Houve, ainda, uma segunda tentativa de execução do DNS Measurement dentro da rede do instituto de informática da UFRGS. Este se deu através do servidor dnsmeasuring-pc.inf.ufrgs.br, criado especificamente para este fim. Contudo, infelizmente, o mesmo problema ocorrido com o servidor anterior, ocorreu neste segundo. Assim sendo, foi desistido da ideia de execução das medições através da rede da universidade, sendo optado pelo uso de um computador pessoal.

O dispositivo utilizado, então, para realizar as medições e execuções da ferramenta DNS Measurement foi um LapTop com sistema operacional Linux, na distribuição Debian, em sua versão 11 (DEBIAN, 2021), com processador AMD Ryzen 5-5500U, com 6 núcleos de CPU, 3MB de memória cache L2 e 12 threads (AMD, 2023) e 8GB de memória. A execução da medição utilizando o DNS Measurement, utilizando este LapTop, leva, em média, 6 horas para construir a árvore de DNS dos 1 milhão de domínios da Tranco List. Esse tempo varia em função da velocidade da rede na qual o LapTop está conectado, uma vez que as medições precisam de conexão com a Internet para serem executadas.

4.3 Banco de Dados

Para o armazenamento e para eventuais consultas dos dados obtidos com as medições, tanto as árvores de DNS quanto os arquivos de concentração (respectivamente ilustrados nas Figuras 3.2 e 3.3) foram salvos em bancos de dados. A tecnologia de SGBD escolhida foi a SQLite, mais especificamente a API (Interface de Programação de Aplicações) sqlite3 (PYTHON, 2023). Essa é uma interface de programação para bancos de dados SQLite, bastante simples e com fácil integração com a linguagem Python, além de ser uma tecnologia de banco de dados extremamente simples e bastante eficiente, o que atende perfeitamente as necessidades do escopo deste estudo.

Para o armazenamento destes dados, foram criados 2 bancos de dados:

- O banco *medicoes.db* é o banco de dados onde estão armazenados os dados do arquivo de concentração (Figura 3.3). Para cada medição, é criado uma tabela dentro do banco *medicoes.db*, chamada de 'medicao_xxxx', onde 'xxxx' é a data onde foi executada a medição. As tabelas possuem 3 colunas: um campo texto para a empresa/provedor de DNS, um campo de número inteiro para a quantidade de domínios associados a este provedor, e um campo de número real para a porcentagem de domínios sob gerência deste provedor. A Figura 4.10 ilustra a criação deste banco e de uma destas tabelas, via *script* em Python. Neste caso, para a data do dia 16 de Dezembro de 2022.

```
import sqlite3

# Abre a conexão com o banco de dados ou cria um novo se não existir
conn = sqlite3.connect('/home/Demetrio/Documentos/TCC/medicoes.db')

# Cria a tabela se não existir
conn.execute('CREATE TABLE IF NOT EXISTS medicao_22_16_12 (empresa TEXT, quantidade INTEGER, porcentagem REAL)')

# Lê o arquivo de medições de concentração
with open('Concentracao dia 2022-12-16', 'r') as f:
    lines = f.readlines()

# Insere cada medição na tabela
for line in lines:
    empresa, quantidade, porcentagem = line.strip().split(':')
    quantidade = int(quantidade)
    porcentagem = float(porcentagem)
    conn.execute('INSERT INTO medicao_22_16_12 (empresa, quantidade, porcentagem) VALUES (?, ?, ?)', (empresa, quantidade,
porcentagem))

# Salva as mudanças e fecha a conexão com o banco de dados
conn.commit()
conn.close()
```

Figura 4.10 - Banco *medicoes.db*, o qual armazena os dados de concentração

- O banco *arvores.db* é o banco de dados onde estão armazenados os dados do arquivo de árvores de DNS (Figura 3.2). Analogamente ao banco *medicoes.db,* para cada medição, é criado uma tabela dentro do banco arvores.db, chamada de 'arvore_xxxx', onde 'xxxx' é a data onde foi realizada a medição. As tabelas possuem 5 colunas: um campo para o domínio em questão, um campo para os registros NS deste domínio, um campo para os registros IPv4 deste domínio, um campo para os registros IPv6 deste domínio, e finalmente um campo para o provedor de DNS correspondente. A Figura 4.11 ilustra a criação deste banco e de uma destas tabelas, via script em Python. Neste caso, para a data do dia 16 de Dezembro de 2022.

```
import sqlite3
import json
# Abre a conexão com o banco de dados ou cria um novo se não existir
conn = sqlite3.connect('/home/Demetrio/Documentos/TCC/arvores.db')

# Cria a tabela se não existir
conn.execute('CREATE TABLE IF NOT EXISTS arvore_22_16_12 (dominio TEXT, NSrecords TEXT, Arecords TEXT , AAAArecords
TEXT, empresas TEXT)')

# Lê o arquivo de arvores de DNS

with open('Medição dia2022-12-16', 'r') as file:
    lines = file.readlines()
    for i, line in enumerate(lines):
        if 'dominio' in line:
            domain = lines[i+1].strip()
        elif 'lista de NS records' in line:
            ns_records = lines[i+1].strip()
            ns_records = json.dumps(ns_records)
        elif 'lista de A records' in line:
            a_records = lines[i+1].strip()
            a_records = json.dumps(a_records)
        elif 'lista de AAAA records' in line:
            aaaa_records = lines[i+1].strip()
            aaaa_records = json.dumps(aaaa_records)
        elif 'lista de empresas' in line:
            companies = lines[i+1].strip()
            companies = json.dumps(companies)
            conn.execute('INSERT INTO arvore_22_16_12 (dominio, NSrecords, Arecords, AAAArecords, empresas) VALUES (?, ?, ?, ?, ?)',
(domain, ns_records, a_records, aaaa_records, companies))

# Salva as mudanças e fecha a conexão com o banco de dados
conn.commit()
conn.close()
```

Figura 4.11 - Banco *arvores.db*, o qual armazena os dados de árvores de DNS

5 ANÁLISE E RESULTADOS

Neste capítulo, serão analisados os resultados obtidos com as medições realizadas através da ferramenta DNS Measurement. Para guiar esta análise, serão novamente revistadas as perguntas de pesquisa, nas quais foram apresentadas no Capítulo 1 - Introdução. As próximas seções serão divididas para cada pergunta de pesquisa, explorando e tentando respondê-las através dos dados gerados.

5.1 Primeira Pergunta de Pesquisa

Relembrando, a primeira pergunta de pesquisa elaborada é:

- Qual foi o comportamento observado acerca do serviço de DNS na Internet? De fato, temos indícios que comprovam a existência da concentração deste serviço em poucos provedores, ou é um serviço bem distribuído?

Para respondê-la, serão utilizados os dados de concentração por provedor de DNS. O gráfico a seguir, mostra a porcentagem de domínios (dos 1 milhão de domínios da lista Tranco) que tem seus registros de DNS hospedados pelos 10 provedores de DNS que aparecem no topo da lista de concentração (ou seja, os 10 maiores provedores de DNS em quantidade de domínios) somados, ao longo das medições realizadas.

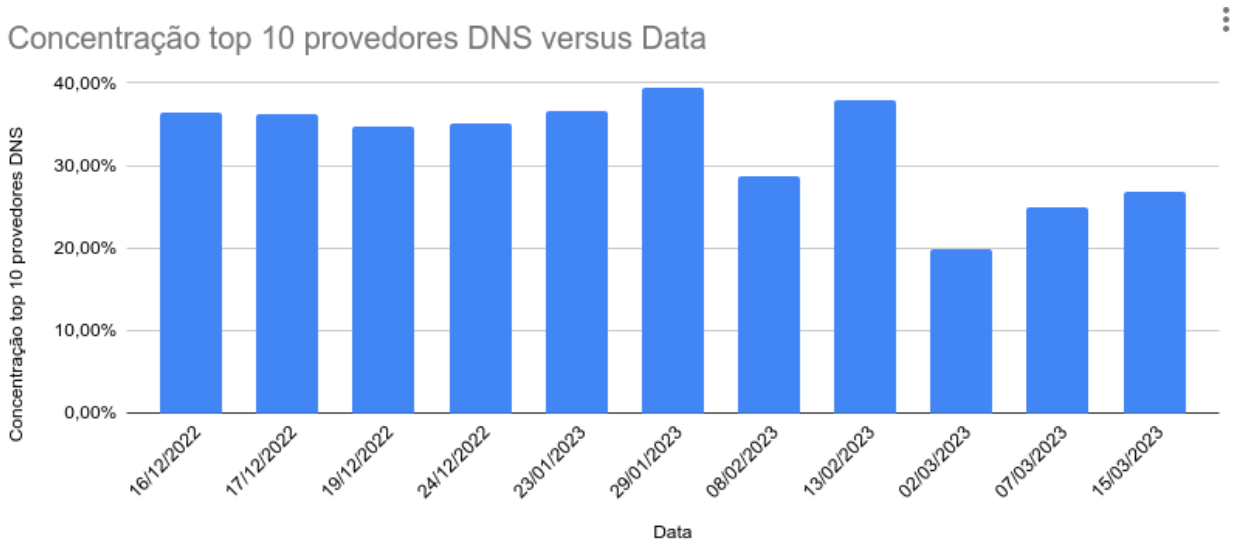

Figura 5.1 - Concentração de domínios dos Top 10 provedores de DNS

No gráfico, pode-se observar que, neste intervalo de 03 meses de medições (entre 16 de Dezembro de 2022 e 15 de Março de 2023), a média ficou em torno de 30% de concentração para os domínios medidos. Em outras palavras, no período onde foram realizadas as medições, na média, 30% dos domínios da lista da Tranco de um milhão de domínios (ou seja, trezentos mil domínios), tinham seus registros de DNS hospedados nos 10 maiores provedores de DNS. Observando a lista completa de concentração de provedores, tem-se que o número total de provedores de DNS responsáveis por gerenciar estes um milhão de domínios foi de pouco mais de 3000 provedores, conforme mostra a Figura 5.2.

```
3033 ['BPC-76-AS']:1:0.0001
3034 ['INITLAB']:1:0.0001
3035 ['XANTARO']:1:0.0001
3036 ['FASSBERG', 'GOTANET']:1:0.0001
3037 ['PACKET', 'DIGITALOCEAN-ASN']:1:0.0001
3038 ['MDCLOUD']:1:0.0001
3039 ['ACWEB-AS']:1:0.0001
3040 ['PORTAL-TO-WEB', 'PRGMR']:1:0.0001
3041 ['WEDOS2']:1:0.0001
3042 ['SENTENTIA-DC1-AU-AP']:1:0.0001
3043 ['LIMESTONENETWORKS']:1:0.0001
3044 ['serverdiscounter', 'de-Metaliance-DUS']:1:0.0001
3045 ['IRF-AS']:1:0.0001
3046 ['HOSTPOINT-AS', 'NICAT']:1:0.0001
3047 ['ASJZTKFT']:1:0.0001
3048 ['belcloud']:1:0.0001
```

Figura 5.2 - Final do arquivo de concentração para o dia 15/03/2023

Ou seja, observaram-se mais de 3000 provedores de DNS ao todo, porém, os 10 maiores, somados, são responsáveis por, em média, 30% dos domínios. Com isto, é bastante confortável de se afirmar que, de fato, o serviço de DNS na Internet está bastante concentrado em poucos provedores.

5.2 Segunda Pergunta de Pesquisa

Relembrando, a segunda pergunta de pesquisa elaborada é:

- Quais são os 10 provedores de serviços de DNS na Internet responsáveis pelo maior número de domínios? Eles foram os mesmos em todas as medições?

Para essa pergunta, também serão utilizados os dados coletados de concentração de provedores de DNS. A Tabela 5.1 traz a informação dos 10 provedores de DNS responsáveis pelo maior número de domínios ao longo das medições.

Posição	16, 17, 19, 24 de Dezembro 23, 29 de Janeiro 13 de Fevereiro	08 de Fevereiro 15 de Março	02 de Março	07 de Março
1	CLOUDFLARENET	CLOUDFLARENET	CLOUDFLARENET	CLOUDFLARENET
2	AMAZON-02	AMAZON-02	AMAZON-02	AMAZON-02
3	GODADDY-DNS	GODADDY-DNS	GODADDY-DNS	GODADDY-DNS
4	ALIBABA-CN-NET	ALIBABA-CN-NET	ALIBABA-CN-NET	ALIBABA-CN-NET
5	GOOGLE	GOOGLE	GOOGLE	GOOGLE
6	TIGGEE	MICROSOFT-CORP-MSN-AS-BLOCK	MICROSOFT-CORP-MSN-AS-BLOCK	MICROSOFT-CORP-MSN-AS-BLOCK
7	MICROSOFT-CORP-MSN-AS-BLOCK	TIGGEE	TIGGEE	TIGGEE
8	NSONE	NSONE	AKAMAI-ASN2	NSONE
9	IONOS-AS	IONOS-AS	NSONE	OVH
10	OVH	OVH	IONOS-AS	IONOS-AS

Tabela 5.1 - Top 10 provedores de DNS por medição

Conforme mostra a Tabela 5.1, o ranking dos 10 maiores provedores de DNS por número de domínios manteve-se bastante estável ao longo do tempo. Houveram algumas mudanças, como a mudança de ranking entre o provedor Tiggee e Microsoft e a mudança entre o provedor Ionos e Ovh. Fora isso, também houve a entrada do provedor de DNS Akamai no ranking do dia 02 de Março, ficando na oitava posição. Entretanto, esse provedor já não aparece nos rankings do dia 07 e 15 de Março.

Os provedores não foram exatamente os mesmos durante as medições. Entretanto, houveram poucas mudanças, e pode-se concluir que, no período analisado, o ranking manteve-se relativamente estável.

5.3 Terceira Pergunta de Pesquisa

Relembrando, a terceira pergunta de pesquisa elaborada é:

- Os domínios das maiores provedoras de serviços de DNS estão hospedados em seus próprios serviços de DNS ou de terceiros?

Para responder essa pergunta, serão utilizados os dados da árvore de DNS, executando uma consulta no banco de dados *arvores.db* (descrito na seção 4.3 do documento) para visualizar quem são os provedores de DNS responsáveis por manter os registros de DNS dos maiores provedores. Para os maiores provedores, utilizou-se todos os que aparecem na seção anterior. A Tabela 5.2 mostra quais são os provedores de DNS nos quais são responsáveis pelos registros de DNS dos maiores provedores de DNS.

Domínio	Provedor DNS
cloudflare.com	CLOUDFLARENET
amazon.com	ORACLE-BMC-31898
godaddy.com	GODADDY-DNS, AKAMAI-ASN2
alibaba.com	ALIBABA-CN-NET
google.com	GOOGLE
tiggee.com	TIGGEE
microsoft.com	MICROSOFT-CORP-MSN-AS-BLOCK
ns1.com	NSONE
ionos.com	IONOS-AS
ovh.com	OVH
akamai.net	AKAMAI-ASN2

Tabela 5.2 - Quais são os provedores de DNS dos maiores provedores de DNS

Pode-se notar que nem todos utilizam o seu próprio serviço de DNS, como é o caso da Amazon, segundo maior provedor de DNS segundo a Tabela 5.1, que utiliza os serviços de DNS da Oracle, ou então a Godaddy, que utiliza o seu próprio serviço de DNS mas também utiliza o serviço de DNS da Akamai. Todos os outros grandes provedores utilizam o seu próprio serviço de DNS. É interessante o fato de que, mesmo estando entre os maiores provedores de DNS da Internet, o domínio da *tiggee.com* não está entre os um milhão de principais domínios, segundo a Tranco.

5.4 Quarta Pergunta de Pesquisa

Relembrando, a quarta pergunta de pesquisa é:

- Os domínios brasileiros costumam ter seus registros de DNS administrados por provedores brasileiros?

Para responder a esta pergunta, serão utilizados os dados obtidos com as árvores de DNS, através do banco de dados *arvores.db*. Para domínios brasileiros, neste caso foi feito uma consulta no banco de dados procurando por domínios nos quais continham o ‘.br’ em seu nome, conforme a query a seguir:

cur.execute("SELECT dominio, empresas FROM arvore_22_17_12 WHERE dominio LIKE '%.br%'")

A tabela utilizada foi a medição do dia 17 de Dezembro de 2022, escolhida, para esta pergunta, aleatoriamente. A Figura 5.3 e 5.4 ilustram um recorte do resultado da consulta.

```
('tripadvisor.com.br', '"[\'NSONE\']"')
('climatempo.com.br', '"[\'GOOGLE\']"')
('kiwify.com.br', '"[\'CLOUDFLARENET\']"')
('letras.mus.br', '"[\'AMAZON-02\']"')
('submarino.com.br', '"[\'AMAZON-02\']"')
('espn.com.br', '"[\'AMAZON-02\']"')
('colaboraread.com.br', '"[\'MICROSOFT-CORP-MSN-AS-BLOCK\']"')
('pelando.com.br', '"[\'CLOUDFLARENET\']"')
('estacio.br', '"[\'AMAZON-02\']"')
('buscape.com.br', '"[\'AMAZON-02\']"')
('senac.br', '"[\'MICROSOFT-CORP-MSN-AS-BLOCK\']"')
('webmotors.com.br', '"[\'AMAZON-02\']"')
('redtube.com.br', '"[\'NSONE\']"')
('meliuz.com.br', '"[\'AMAZON-02\']"')
('olhardigital.com.br', '"[\'AMAZON-02\']"')
('rdstation.com.br', '"[\'AMAZON-02\']"')
('alura.com.br', '"[\'CLOUDFLARENET\']"')
```

Figura 5.3 - Recorte do resultado da consulta acerca de domínios brasileiros

```
('vteximg.com.br', '"[\'AMAZON-02\']"')
('sumicity.net.br', '"[\'GOOGLE\']"')
('bling.com.br', '"[\'AMAZON-02\']"')
('monetizze.com.br', '"[\'AMAZON-02\']"')
('spoleto.com.br', '"[\'MICROSOFT-CORP-MSN-AS-BLOCK\']"')
('pucrs.br', '"[\'CLOUDFLARENET\']"')
('brasildefato.com.br', '"[\'CLOUDFLARENET\']"')
('totvs.com.br', '"[\'GOOGLE\']"')
('vtex.com.br', '"[\'AMAZON-02\']"')
('mercadoshops.com.br', '"[\'AMAZON-02\']"')
('drogaraia.com.br', '"[\'AMAZON-02\']"')
('gastecnologia.com.br', '"[\'AMAZON-02\']"')
('asn.net.br', '"[\'CLOUDFLARENET\']"')
('suamusica.com.br', '"[\'AMAZON-02\']"')
('xpi.com.br', '"[\'AMAZON-02\']"')
('criticalhits.com.br', '"[\'CLOUDFLARENET\']"')
('adrenaline.com.br', '"[\'CLOUDFLARENET\']"')
('ahnegao.com.br', '"[\'CLOUDFLARENET\']"')
('estantevirtual.com.br', '"[\'AMAZON-02\']"')
```

Figura 5.4 - Recorte do resultado da consulta acerca de domínios brasileiros

Por curiosidade, também foi realizada a mesma consulta para a tabela de medição do dia 15 de Março de 2022.

```
('gestaofranquias.com.br', '"[\'CLOUDFLARENET\']"')
('diariodosertao.com.br', '"[\'CLOUDFLARENET\']"')
('naosalvo.com.br', '"[\'CLOUDFLARENET\']"')
('soportugues.com.br', '"[\'CLOUDFLARENET\']"')
('calculareconverter.com.br', '"[\'CLOUDFLARENET\']"')
('genesisweb.com.br', '"[\'CLOUDFLARENET\']"')
('florestal.gov.br', '"[\'HWCLOUDS-AS-AP\', \'HWCSNET\']"')
('bbbaterias.com.br', '"[\'AMAZON-02\']"')
('revistamenu.com.br', '"[\'GOOGLE\']"')
```

Figura 5.5 - Recorte do resultado da consulta acerca de domínios brasileiros

A Figura 5.6 mostra a quantidade total de vezes que os provedores são utilizados para os domínios brasileiros (que contenham o TLD ‘.br’) da Tranco para o dia 17 de Dezembro de 2022. A lista está ordenada do mais utilizado para o menos e pode-se observar que, quando levados apenas domínios brasileiros em consideração, a lista dos maiores provedores é alterada (comparar com a Figura 5.1).

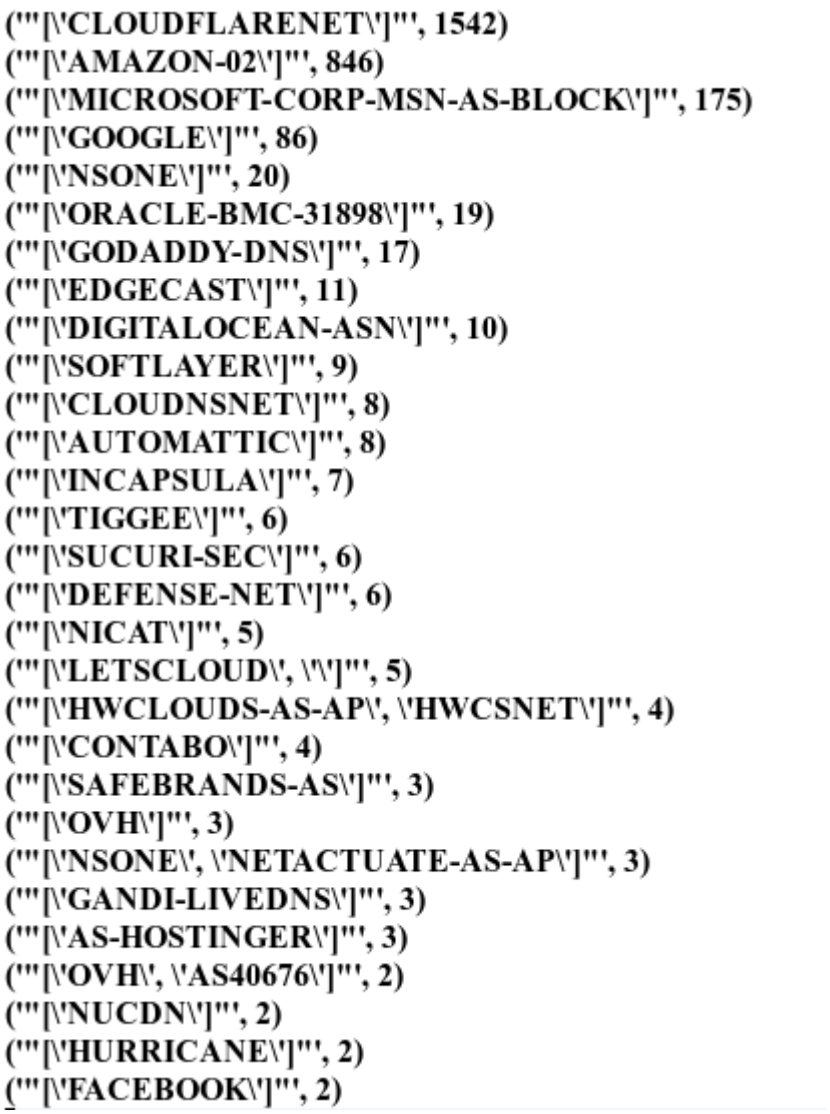

Figura 5.6 - Maiores provedores de DNS para domínios brasileiros

A Tabela 5.3 indica o país de origem dos provedores de DNS dos domínios brasileiros presentes na lista da Tranco (Figura 5.6). As informações de país de origem estão disponíveis na tabela da Caida, descrita na seção 3.3 e apresentadas na Figura 3.13.

Provedor	País de Origem
NSONE	EUA
GOOGLE	EUA
CLOUDFLARENET	EUA
AMAZON-02	EUA
MICROSOFT-CORP-MSN-AS-BLOCK	EUA
TIGGEE	EUA
ORACLE-BMC	EUA
HWCSNET	CHINA
HWCLOUDS-AS-AP	CHINA
GODADDY-DNS	ALEMANHA
EDGECAST	EUA
DIGITALOCEAN-ASN	EUA
SOFTLAYER	EUA
CLOUDNSNET	BULGÁRIA
AUTOMATTIC	EUA
INCAPSULA	EUA
SUCURI-SEC	EUA
DEFENSE-NET	EUA
NICAT	ÁUSTRIA
LETSCLOUD	EUA
CONTABO	ALEMANHA
SAFEBRANDS-AS	FRANÇA
OVH	FRANÇA
NETACTUATE	EUA
GANDI-LIVEDNS	FRANÇA
AS-HOSTINGER	CHIPRE
AS40676	EUA
NUCDN	EUA
HURRICANE	EUA
FACEBOOK	EUA

Tabela 5.3 - País de origem dos provedores de DNS dos domínios brasileiros

Claramente, a resposta para essa pergunta é que não, os domínios brasileiros não têm seus registros de DNS administrados por provedores brasileiros. Existe uma concentração bastante grande em provedores norte-americanos (Figura 5.6), apesar de alguns domínios estarem em provedores de DNS europeus e asiáticos.

5.5 Quinta Pergunta de Pesquisa

Relembrando, a quinta e última pergunta de pesquisa elaborada, 'é:

- Os domínios russos e chineses costumam ter seus registros de DNS administrados por provedores de seus próprios países?

Novamente, foi utilizado o banco de dados *arvores.db* para responder a essa pergunta. Para os domínios russos, foi utilizado o TLD (Top Level Domain) *.ru,* e para o caso da china, foi utilizado o TLD *.cn.* As consultas executadas foram, respectivamente:

cur.execute("SELECT dominio, empresas FROM arvore_23_15_03 WHERE dominio LIKE '%.ru%'")

e

cur.execute("SELECT dominio, empresas FROM arvore_23_15_03 WHERE dominio LIKE '%.cn%'")

A Figura 5.7 mostra um recorte do resultado da pesquisa para os domínios chineses.

```
('hcnu.edu.cn', '"[\'CNGI-BJ-IX2-AS-AP\', \'ERX-CERNET-BKB\']"')
('cqdx.gov.cn', '"[\'CHINANET-Henan-Luoyang-IDC\', \'CSTNET-AS-AP\', \'WYNETWORK\', \'CHINA169-Backbone\']"')
('cyapi.cn', '"[\'ASEPL-AS-AP\', \'ALIBABA-CN-NET\']"')
('gxgg.gov.cn', '"[\'TENCENT-NET-AP\', \'UNICOM-GuangZhou-IDC\', \'CHINANET-SHANGHAI-MAN\', \'CHINAMOBILE-CN\']"')
('china-cer.com.cn', '"[\'ALIBABA-CN-NET\']"')
('join-tsinghua.edu.cn', '"[\'ERX-CERNET-BKB\']"')
('sxpi.edu.cn', '"[\'ERX-CERNET-BKB\', \'CNGI-XA-IX-AS-AP\']"')
('nmgjjjc.gov.cn', '"[\'ALIBABA-CN-NET\']"')
('jxust.cn', '"[\'ALIBABA-CN-NET\']"')
('hexinfo.cn', '"[\'ALIBABA-CN-NET\', \'ASEPL-AS-AP\']"')
('jsrailway.com.cn', '"[\'ALIBABA-CN-NET\']"')
('cggc.cn', '"[\'ALIBABA-CN-NET\', \'ASEPL-AS-AP\']"')
('googlenav.cn', '"[\'WYNETWORK\']"')
('gmdzkj.cn', '"[\'Baidu\']"')
```

Figura 5.7 - Recorte do resultado da consulta acerca de domínios chineses

A Figura 5.8 mostra a quantidade total de vezes que os provedores são utilizados para os domínios chineses (que contenham o TLD '.cn') da Tranco para o dia 17 de Dezembro de 2022.

```
('"[\'ALIBABA-CN-NET\']"', 4026)
('"[\'ASEPL-AS-AP\', \'ALIBABA-CN-NET\']"', 544)
('"[\'TENCENT-NET-AP\', \'CHINANET-SHANGHAI-MAN\']"', 237)
('"[\'TENCENT-NET-AP\', \'UNICOM-GuangZhou-IDC\', \'CHINAMOBILE-CN\', \'CHINANET-SHANGHAI-MAN\']"', 230)
('"[\'CLOUDFLARENET\']"', 186)
('"[\'CHINANET-SCIDC-AS-AP\']"', 169)
('"[\'CHINAMOBILE-CN\', \'CHINANET-SHANGHAI-MAN\', \'TENCENT-NET-AP\', \'UNICOM-GuangZhou-IDC\']"', 139)
('"[\'ALIBABA-CN-NET\', \'ASEPL-AS-AP\']"', 139)
('"[\'TENCENT-NET-AP\', \'CHINAMOBILE-CN\', \'CHINANET-SHANGHAI-MAN\', \'UNICOM-GuangZhou-IDC\']"', 133)
('"[\'TENCENT-NET-AP\', \'UNICOM-GuangZhou-IDC\', \'CHINANET-SHANGHAI-MAN\', \'CHINAMOBILE-CN\']"', 123)
('"[\'CHINANET-SHANGHAI-MAN\', \'TENCENT-NET-AP\', \'CHINAMOBILE-CN\', \'UNICOM-GuangZhou-IDC\']"', 118)
('"[\'CSTNET-AS-AP\', \'CHINA169-Backbone\', \'CHINANET-Henan-Luoyang-IDC\']"', 112)
('"[\'CNGI-BJ-IX2-AS-AP\', \'ERX-CERNET-BKB\']"', 112)
('"[\'CHINANET-Henan-Luoyang-IDC\', \'CSTNET-AS-AP\', \'CHINA169-Backbone\']"', 74)
('"[\'CT-DongGuan-IDC\']"', 65)
('"[\'CHINANET-SHANGHAI-MAN\', \'TENCENT-NET-AP\']"', 65)
('"[\'AMAZON-02\']"', 51)
```

Figura 5.8 - Maiores provedores de DNS para domínios chineses

A Tabela 5.4 indica o país de origem dos provedores de DNS dos domínios chineses presentes na lista da Tranco (Figura 5.8). As informações de país de origem estão disponíveis na tabela da Caida, descrita na seção 3.3 e apresentadas na Figura 3.13.

Provedor	País de Origem
CNGI-BJ-IX2-AS-AP	CHINA
ERX-CERNET-BKB	CHINA
CHINANET-HENAN-LUOYANG-IDC	CHINA
CSTNET-AS-AP	CHINA
WYNETWORK	CHINA
CHINA169-BACKBONE	CHINA
CHINAMOBILE-CN	CHINA
ALIBABA-CN-NET	CHINA
ASEPL-AS-AP	SINGAPURA
TENCENT-NET-AP	CHINA
UNICOM-GUANGZHOU-IDC	CHINA
CHINANET-SHANGHAI-MAN	CHINA
BAIDU	CHINA
CLOUDFLARENET	EUA
CHINANET-SCIDC-AS-AP	CHINA
CT-DONGGUAN-IDC	CHINA
AMAZON-02	EUA

Tabela 5.4 - País de origem dos provedores de DNS dos domínios chineses

Nota-se que, para os domínios chineses, quase exclusivamente são utilizados provedores de DNS de companhias chinesas. Cloudflarenet e Amazon (os maiores provedores de DNS) também aparecem na lista. Como pode-se perceber, boa parte dos domínios chineses possuem mais de um provedor DNS associado, o que é ótimo para evitar quedas de disponibilidade na resolução de seus nomes de domínio. Além disso, outro fato curioso é o de que somente a Alibaba está na lista dos 10 maiores provedores de DNS (Tabela 5.1), muito embora apareçam várias outras gigantes chinesas, como a Tencent e a Baidu.

```
('bshlv.ru', '"[\'AS-REGRU\']"')
('marjani-mechet.ru', '"[\'CLOUDFLARENET\']"')
('incambodia.ru', '"[\'RU-CENTER-AS\', \'SWEB-AS\']"')
('goloskubani.ru', '"[\'CLOUDFLARENET\']"')
('pokerdomwins.ru', '"[\'CLOUDFLARENET\']"')
('arlecin.ru', '"[\'HURRICANE\']"')
('fuxtec.ru', '"[\'CLOUDFLARENET\']"')
('sitebill.ru', '"[\'AS-REGRU\']"')
('iceshow-perm.ru', '"[\'RU-JSCIOT\']"')
('sntrans.ru', '"[\'MASTERHOST-AS\']"')
('profarmy.ru', '"[\'MASTERHOST-AS\', \'SELECTEL\', \'UNITEDNET\']"')
('bus.ru', '"[\'SELECTEL\']"')
('uoedu.ru', '"[\'SELECTEL\']"')
```

Figura 5.9 - Recorte do resultado da consulta acerca de domínios russos

A Figura 5.10 mostra a quantidade total de vezes que os provedores são utilizados para os domínios russos (que contenham o TLD '.ru') da Tranco para o dia 17 de Dezembro de 2022.

```
('"[\'CLOUDFLARENET\']"', 8909)
('"[\'AS-REGRU\']"', 2719)
('"[\'TimeWeb-AS\']"', 1696)
('"[\'SELECTEL\']"', 976)
('"[\'MASTERHOST-AS\']"', 795)
('"[\'RU-JSCIOT\']"', 419)
('"[\'AMAZON-02\']"', 320)
('"[\'OOOVPS-AS\', \'GD-EMEA-DC-SXB1\']"', 318)
('"[\'GD-EMEA-DC-SXB1\', \'OOOVPS-AS\']"', 309)
('"[\'RU-CENTER-AS\', \'SWEB-AS\']"', 288)
('"[\'SWEB-AS\', \'RU-CENTER-AS\']"', 241)
('"[\'TIGGEE\']"', 149)
('"[\'MYLOC-AS\']"', 114)
('"[\'SELECTEL\', \'STORMWALL-RUS\', \'\', \'EUROBYTE\']"', 102)
('"[\'CLOUDNSNET\']"', 99)
```

Figura 5.10 - Maiores provedores de DNS para os domínios russos

A Tabela 5.5 indica o país de origem dos provedores de DNS dos domínios russos presentes na lista da Tranco (Figura 5.10). As informações de país de origem estão disponíveis na tabela da Caida, descrita na seção 3.3 e apresentadas na Figura 3.13.

Provedor	País de Origem
AS-REGRU	RÚSSIA
CLOUDFLARENET	EUA
RU-CENTER-AS	RÚSSIA
SWEB-AS	RÚSSIA
TIMEWEB-AS	RÚSSIA
RU-JSCIOT	RÚSSIA
MASTERHOST-AS	RÚSSIA
SELECTEL	RÚSSIA
CLOUDNSNET	RÚSSIA
AMAZON-02	EUA
OOOVPS-AS	RÚSSIA
GD-EMEA-DC-SXB1	ALEMANHA
TIGGEE	EUA
MYLOC-AS	ALEMANHA
STORMWALL-RUS	RÚSSIA
EUROBYTE	RÚSSIA

Tabela 5.5 - País de origem dos provedores de DNS dos domínios russos

Já para o caso dos domínios russos, tem-se a incidência de provedores de DNS norte-americanos, como é o caso dos líderes em número de domínios administrados (Cloudflarenet, Amazon e Tiggee). Também há a incidência de provedores alemães. Entretanto, a grande maior parte dos provedores para os domínios russos são de origem na própria Rússia.

Portanto, a resposta para essa última pergunta de pesquisa, é que, predominantemente, os domínios russos e chineses utilizam provedores de DNS locais, isto é, provedores pertencentes ao seu próprio país de origem. Este fato é positivo pelo fato de diminuir a concentração de grandes provedores norte-americanos e também diluir a concentração geográfica.

5.6 Fechamento do capítulo

Com isso, conclui-se as perguntas de pesquisa, sendo todas elas respondidas satisfatoriamente, utilizando-se dos dados coletados pelas medições através da ferramenta *DNS Measurement.* Com a primeira pergunta de pesquisa respondida, tem-se a informação central que guiou e motivou este estudo, que é a confirmação da existência de uma centralização do serviço de DNS na Internet. Pelos dados obtidos na pesquisa, não foi difícil perceber que há, de fato, uma centralização e concentração bastante clara no que tange os serviços de DNS na Internet (ver Figura 5.1), onde apenas 10 provedores de DNS ficaram, em média ao longo das medições, responsáveis por 30% dos domínios da TrancoList. Na medição realizada no dia 29 de Janeiro de

2023, inclusive, estes provedores beiravam a concentração de 40% sobre os domínios investigados.

Dito isso, é bastante prudente e desejável, visto todos os problemas citados e estudos realizados sobre o tema (ver seções 1.2, 2.5 e 2.6), de que haja alguma ação para mitigar essa concentração de serviços de DNS sob a responsabilidade de poucos provedores. Este estudo, os dados coletados e esta ferramenta construída para realização das medições (*DNS Measurement)*, podem ser bastante úteis para que essas medições possam ser realizadas periodicamente, a fim de consultar o cenário atual de DNS na Internet, ou até mesmo auxiliar em uma possível solução futura para apoiar na descentralização destes serviços.

6 CONCLUSÃO

Finalmente, chega-se ao capítulo final deste trabalho. Nele, serão retomados alguns aspectos abordados nos capítulos iniciais, os aprendizados e descobertas e também os próximos passos para a continuidade deste estudo tão relevante para a saúde do ecossistema da Internet.

Logo no início do texto, houve uma contextualização sobre o tema, a fim de conscientizar e instigar a curiosidade do leitor acerca do problema da centralização, onde foram apresentados dados e estudos relevantes na área. Para guiar o estudo, então, foram propostas algumas perguntas de pesquisa relevantes para que, além da testagem sobre uma possível concentração entre os provedores de DNS na Internet (no qual é um dos principais objetivos deste trabalho), também fossem descobertas alguns outros aspectos sobre o cenário de DNS no Brasil e no mundo. Pode-se conhecer por exemplo, através da segunda pergunta, quais são os maiores provedores de DNS ao longo do tempo, sua posição em um ranking e a variação entre eles. Como era esperado, este ranking é dominado por grandes empresas de tecnologia ao redor do mundo, mas majoritariamente empresas norte-americanas. Além disso, observou-se que, apesar de haver pouca variação entre os nomes, em um curto período de 3 meses de medições, já foram observadas variações relevantes, tanto do ponto de vista de quantidade de domínios administradas pelos 10 maiores provedores, o que sugere que este tema é um assunto bastante vivo e sujeito a mudanças.

Outra descoberta interessante foi a confirmação de que não temos provedores de DNS brasileiros muito relevantes, nem mesmo para domínios brasileiros. Isso pode ser, inclusive, parte de uma futura solução para este problema, tendo em vista que o problema da concentração de DNS na Internet é também um problema geográfico, e que o risco geográfico não pode ser ignorado. A expansão do fator geográfico pode ser muito útil em casos de desastres naturais ou até mesmo conflitos humanos, por exemplo. Em contrapartida, observou-se que, para os casos da China e Rússia, que são historicamente países rivais aos Estados Unidos, seus domínios utilizam muitos provedores de DNS nacionais, principalmente para o caso chines. Isso é bastante positivo do ponto de vista da descentralização, inclusive geográfica, pois são países que ficam bastante distantes dos americanos e estão em grande expansão tecnológica, podendo ser aliados no combate ao problema central deste estudo. Outro aspecto positivo dessa expansão seria a questão da privacidade e legislação. Quanto menos concentrado e mais distribuído estão os registros de DNS na Internet, menor é o risco destes pontos (claro que, desde que os países responsáveis tenham políticas claras e transparentes acerca do tratamento desses dados).

No geral, a realização deste trabalho foi bastante positiva, tanto para o meu ponto de vista pessoal, onde aprendi, pesquisei e me interessei sobre o tema, mas também para toda a comunidade, podendo agora ser possível medir a situação de concentração de DNS na Internet. Este tópico não é um tópico abordado com frequência, mesmo no meio da computação, e os fatos mostram que a sua importância é bastante significativa para o bom funcionamento da Internet. Portanto, contribuições e aprofundamentos como esse são bastante positivos.

Para o futuro e próximos passos, a ferramenta DNS Measurement poderia ser aperfeiçoada, a fim de possibilitar uma melhor experiência para usuários, podendo ser adicionada uma interface gráfica e consultas mais avançadas nos bancos de dados. Creio que ela tenha potencial para ajudar a comunidade a encontrar soluções para a descentralização do DNS, como por exemplo, encontrar os países onde não há provedores de DNS relevantes, ou ainda encontrar nichos específicos de domínios (domínios de sites de comércio, por exemplo), que tem seus registros de DNS todos gerenciados por um mesmo provedor. Isso pode ocasionar, em caso de falha do provedor, que um determinado setor ou até mesmo país tenha seus recursos comprometidos, o que com certeza seria muito negativo e prejudicial.

7 REFERÊNCIAS

MOCKAPETRIS, P. Domain names: Concepts and facilities. [S.l.], 1983

MOCKAPETRIS, P. Domain names: Implementation specification. [S.l.], 1983

MOCKAPETRIS, P. Domain names - concepts and facilities. [S.l.], 1987

MOCKAPETRIS, P. Domain names - implementation and specification. [S.l.], 1987

STEWART, W. Living Internet. 2019

DEUTSCH, L. Host names on-line. [S.l.], 1973

POSTEL, J. Computer mail meeting notes. [S.l.], 1982.

IONOS. Managing DNS Services. Disponível em https://www.ionos.com/help/domains/general-information-about-dns-settings/managing-dns-services/

NIC. Disponível em https://www.nic.ru/en/catalog/for-domain-use/dns-hosting/

VIXIE, P.; SNEERINGER, G.; SCHLEIFER, M. Events of 21-Oct-2002. 2002. Disponível em: http://www.vix.com/mirror/events.021021.html. Acesso em: 17 fev. 2023.

WEINBERG, M. W. D. Review and analysis of attack traffic against A-root and J-root on November 30 and December 1, 2015. 2015.

WANG, Gang; XIE, Geoffrey; YUAN, Shun. Centralization and Performance of the Domain Name System. ACM Transactions on the Web, v. 14, n. 4, p. 1-24, 2020.

KLEINWÄCHTER, Wolfgang; KNIGHT, Brian. Threats to the Decentralization of the Domain Name System. Internet Governance Project Research Paper Series, n. 7, p. 1-35, 2020

ALEXA. Top 1M sites. https://www.alexa.com/topsites. 2018

CAIDA-AS2ORG. Inferred AS to Organization Mapping Dataset. https://www.caida.org/data/as-organizations/, Acessado em: 01/11/2022

CAIDA-IP2AS. Routeviews Prefix to AS mappings Dataset for IPv4 and IPv6. https://publicdata.caida.org/datasets/routing/routeviews-prefix2as/, Acessado em: 01/11/2022

PYTHON. SQLite 3. Disponível em https://docs.python.org/3/library/sqlite3.html

TRANCO. Tranco-list https://tranco-list.eu/#aboutus

TRANCO. A Research-Oriented Top Sites Ranking Hardened Against Manipulation https://tranco-list.eu/assets/tranco-ndss19.pdf.

MAJESTIC. Majestic Million. Disponível em: https://majestic.com/reports/majestic-million. Acesso em: 27 mar. 2023.

UNICAMP. Tabelas Hash. Disponível em https://www.dca.fee.unicamp.br/cursos/EA876/apostila/HTML/node26.html

SIMILARWEB LTD. SimilarWeb Top Websites. Disponível em https://www.similarweb.com/top-websites/. Acesso em: 27 mar. 2023.

HUANG, Dijiang; KALBARCZYK, Zbigniew T.; IYENGAR, Arun. Internet Security Challenges: DNS as a Case Study. IEEE Security & Privacy, v. 18, n. 5, p. 54-62, 2020.

CAIDA. Disponível em https://www.caida.org/. Acesso em 28/03/2023

KARLIN, Anna; FORD, Bryan; LIVSHITS, Ben; SHMATIKOV, Vitaly. DNS: The Hidden Centralization. Proceedings of the 2020 Conference on Computer Communications Security (CCS '20), p. 1-16, 2020.

PYTHON. Disponível em https://www.python.org/

AMD Disponível em https://www.amd.com/pt/products/apu/amd-ryzen-5-5500u

PYTHON, Concurrent Futures. Disponível em
https://docs.python.org/3/library/concurrent.futures.html

PYTHONACADEMY, Dicionários. Disponível em
https://pythonacademy.com.br/blog/dicts-ou-dicionarios-no-python

COHN, Gidon; GURBAXANI, Vijay. Challenges in Internet Governance: The Case of DNS. Journal of Management Information Systems, v. 36, n. 1, p. 1-35, 2019.

DEBIAN. Instalar GNU/Linux 11. Disponível em
https://www.debian.org/releases/stable/i386/pr01.pt.html

UFRGS. Instituto de informática. Disponível em https://www.inf.ufrgs.br/site/