Measurement, Analysis, and System Implementation of Internet Proxy Servers

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Sai Charitha Kothapalli

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Dr. Haiyang Wang

July 2023

Acknowledgements

I would like to express my deepest appreciation for my committee chair, Dr. Haiyang Wang, he continually helped me in my research and other academic pursuits with a sense of motivation and exploration. Without his guidance and persistent help this dissertation would not have been possible. I would like to thank my committee members Dr.Zhuangyi Liu and Dr.Arshia Khan, whose work demonstrated to me the importance of interdisciplinary research. I am grateful to Charlie Kincs whose intellectual contribution has made this possible. My mentor, who backed me to pursue further education: Dr. Thulasi Bikku and my friend, Bhargavi Macherla, who constantly helped me overcome my stress and low aspirations. Finally, I would like to thank my parents, Subbarao Kothaplli and Sri Lakshmi Kothapalli, without whom I would have not been doing what I love. I really admire their love and appreciate their efforts to give me the best.

Dedication

I dedicate this to my loving family, whose unwavering support and encouragement have been instrumental in my academic journey. Your belief in me has fueled my determination to pursue excellence. Furthermore, I would like to dedicate this work to all the women who have endured the trauma of sexual harassment. Your resilience, strength, and courage inspire me to leverage the power of computer science and technology to contribute to the collective effort of creating a world free from harassment and discrimination. Through my research, I strive to develop solutions and systems that promote inclusivity, safety, and empowerment for all.

Abstract

The rapid growth of the Internet has caused serious security, privacy, and performance issues for users. To mitigate these challenges, proxy servers are widely adopted to provide enhanced protection, anonymity, and bridging. In this thesis, we put our emphasis on the measurement, analysis, and implementation of Internet proxy servers. We collected detailed information on 1681 proxy servers across 90 countries. Our data indicate that most Internet proxy servers are unstable with relatively high latency. Finding and adopting the right proxy servers is no easier than locating a needle in a haystack. To help the users pinpoint their desired proxy servers, we developed a system to collect real-time proxy server data from the Internet. This system will check proxy servers' location, availability, and performance, giving users the most updated information. We also applied a BitTorrent-like protocol to provide better synchronization across the data collection servers. Our evaluation indicates that utilizing a BitTorrent-like protocol on uTorrent to synchronize data between a single host machine and four virtual machines (VMs) improves data synchronization performance. The involvement of multiple peers leads to a reduction in latency during the synchronization process. This reduction in latency can be attributed to the parallelized and efficient distribution of data among peers. Overall, our evaluation demonstrates that leveraging a BitTorrent-like protocol in this setup enhances the data synchronization process, improving latency and performance.

Keywords— Internet proxy servers, Performance analysis, Distribution analysis, Ping, Telnet, Traceroute, Data synchronization, BitTorrent

Contents

List of Figures

1 Introduction

1.1 Motivation

Our motivation for this thesis is rooted in the growing importance of the Internet in our daily lives and the need for efficient and effective Internet access and security.

According to the Pew Research Center, internet usage has increased significantly in recent years Center 2021 and has become an essential tool for communication, commerce, education, and entertainment, with billions of users worldwide. However, there are still many challenges to accessing and using the Internet effectively, particularly in regions where Internet access is limited or restricted.

One of the challenges is the performance of Internet connections, which can be affected by various factors, such as network congestion, distance from the server, and server load. Proxy servers can help to address these challenges by providing a means of accessing the Internet through a remote server, which can improve connection speeds, reduce latency, and provide greater access to online content.

Another challenge is the need for Internet security and privacy, particularly in the face of growing cyber threats such as hacking, data breaches, and surveillance. Proxy servers can help to address these challenges by providing enhanced security features such as encryption and VPN services (Goldberg, Wagner, and Brewer 1997), which can protect users from online threats and provide greater privacy and anonymity while using the Internet.

Therefore, our motivation is to improve Internet access and security for users worldwide, by developing a user-friendly and effective proxy server selection system that considers various factors such as server performance, geographic location, and user preferences. By

doing so, the project aims to contribute to the development of more efficient, effective, and user-friendly Internet proxy server systems, which can benefit a range of stakeholders and promote greater access to information and communication across diverse communities.

1.2 Research Overview

The purpose of this research project is to evaluate the performance and distribution of Internet proxy servers and develop a system to help users select a proxy server that best meets their needs. Proxy servers can provide users with enhanced privacy, security, and access to online content. However, the effectiveness of a proxy server depends on its performance and geographic location. This project aims to develop a system that can help users select a proxy server based on their desired location.

To achieve this objective, the project will collect data on a range of proxy servers, including their performance metrics and geographic distribution. The project team will use automated testing tools and manual testing to evaluate the performance of different proxy servers under various conditions. The data collected will be used to develop a system that can match user requirements with the most appropriate proxy server.

The project will evaluate the effectiveness of the system through a series of experiments and user studies. The experiments will measure the performance of the system in terms of accuracy, speed, and user satisfaction. The user studies will provide feedback on the usability and effectiveness of the system from the perspective of end-users.

The research utilizes the BitTorrent protocol for data synchronization and sharing. The collected dataset is distributed across multiple devices using this protocol, ensuring efficient and reliable distribution. The synchronization process is closely monitored to evaluate the CPU utilization and latency experienced by devices due to the adoption of the BitTorrent protocol.

Based on the synchronized dataset, a user-friendly web application is developed. The

application allows users to search for active proxy servers based on their desired location. Real-time information on server availability and performance metrics is provided, empowering users to select the most suitable proxy servers for their specific needs.

The project aims to improve the accessibility and usability of proxy servers by conducting comprehensive data collection, performance evaluation, data synchronization, distribution analysis, and web application development. The results of this research provide valuable insights for users seeking reliable and efficient proxy server solutions.

2 Background and Related Works

2.1 IP Address

An Internet Protocol (IP) address is a unique identifier assigned to each device connected to the Internet. (*DoD standard Internet Protocol* 1980) It is a numerical label that is used to identify and communicate with other devices in a network. Every device connected to the Internet, whether a computer, phone, or other devices enabled by the Internet, is assigned an IP address. (Wikipedia 2004)

IP addresses are required for the functioning of the Internet. They allow devices to communicate with each other, transmit data, and access online resources. IP addresses are also used for security and identification purposes, as they can be used to track and locate devices on a network.

IP addresses are classified into two main types: IPv4 and IPv6. IPv4 addresses are composed of 32 bits and are represented in decimal notation (e.g., 192.168.0.1). However, due to the rapid growth of the Internet, IPv4 addresses are becoming scarce, and a new protocol, IPv6, has been developed. IPv6 addresses are composed of 128 bits and are represented in hexadecimal notation (e.g., 2001:0db8:85a3:0000:0000:8a2e:0370:7334) (IBM 2023).

IP addresses can be static or dynamic. A static IP address is a fixed IP address that does not change over time, while a dynamic IP address is the address assigned by a network service provider and can vary depending on how frequently a device connects to the Internet (Mitchell 2023). Some internet service providers offer static IP addresses for an additional fee, while others provide dynamic IP addresses by default.

IP addresses can also be used to identify the geographic location of a device, which can be helpful for targeted advertising and other location-based services. However, this also raises privacy concerns, as users may not want their location to be tracked or shared without their consent. (Ali 2007) Therefore, there are various methods available to hide or mask IP addresses, such as using proxy servers or virtual private networks (VPNs).

2.1.1 Classful Addressing

Classful addressing was the original method of assigning IP addresses used in the early days of the Internet. It divided the available IPv4 address space into three classes: Class A, Class B, and Class C. Each class had a fixed range of network and host bits, which determined the number of networks and hosts that could be accommodated. (*DoD standard Internet Protocol* 1980) This chapter provides detailed background on classful addressing and its significance in the history of IP addressing.

- Class A Addresses: Class A addresses have the first bit set to 0 and the next 7 bits reserved for the network portion. This allows for a large number of networks ($2^7 = 128$). However, it limits the number of hosts per network ($2^{24} - 2 = 16,777,214$). Class A addresses were typically assigned to large organizations or Internet service providers. (Crocker and Vittal 1978)

- Class B Addresses: Class B addresses have the first two bits set to 10 and the next 14 bits reserved for the network portion. This provides a moderate number of networks ($2^{14} = 16,384$) and a more significant number of hosts per network ($2^{16} - 2 = 65,534$). Class B addresses were commonly assigned to mid-sized organizations.

- Class C Addresses: Class C addresses have the first three bits set to 110 and the next 21 bits reserved for the network portion. This allows for a larger number of networks ($2^{21} = 2,097,152$) but limits the number of hosts per network ($2^8 - 2 = 254$). Class C addresses were commonly used for small organizations or home networks. (Postel

1981)

Classful addressing played a significant role in the early stages of IP addressing. It provided a structured approach to IP address assignment but had limitations regarding address allocation efficiency. Each class has a fixed number of network and host portion bits, resulting in address space wastage. Organizations often required a different number of hosts than what their class allowed, leading to inefficient utilization of IP addresses. The introduction of CIDR marked a transition to a more flexible and efficient addressing scheme. (Rekhter and Li 1993) (Fuller, Li, et al. 1993) Understanding the history and concepts of classful addressing is crucial for grasping the evolution of IP addressing and the challenges it aims to address.

2.1.2 CIDR Addressing

Classless Inter-Domain Routing (CIDR) is a method of IP addressing that replaced the traditional classful addressing scheme. CIDR addressed the limitations of classful addressing by introducing variable-length subnet masks. Instead of fixed class boundaries, CIDR uses a notation that combines the IP address and the subnet mask to determine the network and host portions of the address.(Rekhter and Li 1993) (Fuller, Li, et al. 1993) The subnet mask is represented by a prefix length, indicating the number of network bits. CIDR uses a notation that combines the IP address and the subnet mask. The subnet mask specifies how many bits are used for the network ID and how many bits are used for the host ID. (*Internet Standard Subnetting Procedure* 1985) This allows for variable-length subnet masks, which means that the network portion of the address can be of any length, depending on the specific network requirements.

CIDR notation follows the format "IP address/subnet mask" or "IP address/prefix length". The prefix length represents the number of bits set in the subnet mask. For example, an IP address of 192.168.0.0 with a subnet mask of 255.255.255.0 can be represented as 192.168.0.0/24 in CIDR notation, where "/24" indicates that the first 24 bits are used

for the network ID. (*Assigned numbers* 1985)

CIDR provides several advantages over classful addressing:

1. Efficient address space utilization: CIDR allows for more efficient allocation of IP addresses by using variable-length subnet masks. This means that organizations can allocate IP addresses based on their specific needs rather than being constrained by fixed classful boundaries.

2. Aggregation of IP prefixes: CIDR enables the aggregation of multiple IP prefixes into a single, larger prefix. This reduces the size of routing tables in routers and improves overall network efficiency.

3. Simplified addressing and routing: CIDR simplifies the addressing and routing process by providing a more flexible and scalable solution. It allows for better hierarchical addressing and reduces the complexity of managing large networks.

CIDR has become the standard IP address and routing method in modern networks. (Fuller and Li 2006) It has played a crucial role in addressing the limitations of classful addressing and has contributed to the efficient allocation of IP addresses on the internet.

2.1.3 ISP and IP assignment

Internet Service Providers (ISPs) are crucial in providing internet connectivity to individuals, businesses, and organizations. ISPs assign IP (Internet Protocol) addresses to their customers as part of their services. This chapter provides an overview of ISP operations and the process of IP assignment by ISPs.

ISPs and Their Role: ISPs are companies or organizations that provide internet connectivity to end-users. They maintain the necessary infrastructure, such as routers, switches, and network connections, to enable users to access the internet. ISPs may operate at different scales, ranging from local or regional providers to global telecommunications companies.

IP Address Types

IP addresses are unique identifiers assigned to devices connected to a network. ISPs typically deal with two types of IP addresses:

- Public IP Addresses: These are globally unique IP addresses that are routable on the internet. ISPs acquire blocks of public IP addresses from regional internet registries (RIRs) or other organizations authorized to distribute IP addresses. ISPs assign public IP addresses to their customers to enable direct internet connectivity.

- Private IP Addresses: Private IP addresses are used within private networks, such as home or office networks. These addresses are not globally unique and are used for internal communication within the network. ISPs assign private IP addresses to customer routers, enabling local network connectivity. Network Address Translation (NAT) is commonly used to allow multiple devices within a private network to share a single public IP address.

Process of IP Assignment

The process of IP assignment (Hubbard et al. 1996) by ISPs typically involves the following steps:

1. Allocation from IP Address Pool: ISPs receive blocks of IP addresses from the RIRs or other authorized entities. These IP addresses are allocated based on the ISP's requirements and the number of customers they serve.

2. Dynamic IP Address Assignment: ISPs often use dynamic IP address assignments for residential customers. In this method, customers are assigned a different IP address every time they connect to the internet. Dynamic IP addressing allows ISPs to efficiently utilize their IP address pool by reassigning addresses that are no longer in use.

3. Static IP Address Assignment: Some ISPs offer static IP addresses to customers who require a consistent, fixed IP address. Static IP addresses are commonly used by businesses that require services such as hosting servers, remote access, or running applications that require fixed IP addresses.

4. IP Address Management: ISPs must effectively manage their IP address assignments to avoid IP address exhaustion and ensure proper allocation to customers. This involves tracking and monitoring IP address usage, maintaining accurate records, and implementing IP address assignment policies.

ISPs are crucial in providing internet connectivity and assigning IP addresses to their customers. They acquire blocks of public IP addresses and allocate them to customers for direct internet connectivity. ISPs also assign private IP addresses (Moskowitz et al. 1996) for local network communication. Understanding the process of IP assignment by ISPs is essential for comprehending how internet connectivity is established and managed.

2.2 IP Anonymity Technology

IP anonymity refers to the practice of concealing or obfuscating an individual's IP address when they are connected to the internet. An IP address is a unique numerical identifier assigned to a device on a network, and it can be used to track and identify the user's online activities.

To achieve IP anonymity, individuals can utilize various tools and technologies such as virtual private networks (VPNs), proxy servers, Tor (The Onion Router), and anonymizing networks. These tools help route internet traffic through an intermediary server, masking the user's original IP address and making it difficult to trace their online activities back to their true identity. (Farrell and Tschofenig 2014)

While IP anonymity can provide increased privacy and security, it is important to note that it needs to be foolproof. Users should choose reputable and trustworthy tools, configure

them correctly, and remain cautious about other identifying factors that can compromise their anonymity, such as browser fingerprinting or sharing personal information online.

2.2.1 The Onion Router

The Onion Router (Tor) (Dingledine, Mathewson, Syverson, et al. 2004) is a widely used network designed to provide anonymous communication over the internet. It operates by routing internet traffic through a decentralized network of volunteer-operated servers known as Tor nodes or relays. Tor aims to protect users' privacy and anonymity by obfuscating their IP addresses and encrypting their communications.

The name "Onion Router" refers to the layered encryption used in Tor. When users access the internet through Tor, their traffic is encrypted and passed through multiple Tor nodes. Each node in the network only knows the previous node and the next node in the circuit, creating encryption layers that are similar to the layers of an onion. This multi-hop routing makes it difficult for anyone monitoring the network to trace the traffic back to its origin.

Tor provides several benefits for users seeking anonymity online:

1. Privacy: Tor helps protect users' privacy by hiding their IP addresses and making it difficult to track their online activities. It helps individuals maintain anonymity and shields their internet traffic from surveillance.

2. Anonymity: By bouncing internet traffic through multiple Tor nodes, it becomes challenging to trace the origin of the traffic. This feature is beneficial for people who want to communicate anonymously or access restricted content without being identified.

3. Circumventing Censorship: Tor can bypass internet censorship and access blocked websites or services. It achieves this by encrypting and routing traffic through nodes in different countries, making it difficult for censors to block or filter the content.

4. Resistance to Traffic Analysis: Tor's multi-hop routing helps protect against traffic analysis, where adversaries try to deduce patterns or relationships in internet traffic. By adding layers of encryption and obfuscation, Tor makes it challenging for attackers to gather meaningful information from the traffic.

It is important to note that while Tor provides anonymity and privacy, it may not guarantee complete security. Users should be informed of the potential threats and limitations, such as potential vulnerabilities in the Tor network, malicious Tor exit nodes, and the importance of practicing good online security habits.

Tor is an open-source project that is continuously developed and maintained by a community of volunteers. It is available for various platforms, including desktop computers and mobile devices, that allow users to access the internet anonymously and securely.

2.2.2 Virtual Private Network

A Virtual Private Network (VPN) is a technology that establishes a secure and encrypted internet connection, allowing users to browse the internet privately and securely. (Cisco 2021) It establishes a private network connection, often referred to as a "tunnel," between the user's device and a remote server operated by the VPN provider.

1. Encryption: When a user connects to a VPN, all of their internet traffic is encrypted using robust encryption algorithms. This encryption ensures that data transmitted between the user's device and the VPN server remains secure and cannot be easily intercepted or deciphered by unauthorized parties.

2. Tunneling: The encrypted data is encapsulated within a secure tunnel, which shields it from potential threats on the public internet. The tunnel effectively extends the user's private network across the internet, making it appear as if the user is accessing the internet from the location of the VPN server.

3. IP Address Masking: When connected to a VPN, the user's IP address is replaced with the IP address of the VPN server. This process is known as IP address masking or IP address spoofing. As a result, the user's actual IP address and location are hidden, making it difficult for websites, online services, or other entities to track or identify the user's actual location.

4. Anonymity and Privacy: By encrypting internet traffic and hiding IP addresses, VPNs provide users with increased anonymity and privacy online. VPNs prevent internet service providers (ISPs), government agencies, or other entities from monitoring or logging users' online activities. This is particularly useful in protecting sensitive information like personal data, financial transactions, or browsing habits.

5. Access to Restricted Content: VPNs can also be used to bypass geographic restrictions or censorship imposed by certain websites or online services. By connecting to a VPN server in a different country, users can appear as if they are accessing the internet from that location, granting them access to region-restricted content or services.

Choosing a reputable VPN provider that prioritizes user privacy, has a robust encryption protocol, and does not log or store user activity data is essential. Additionally, users should be aware that while VPNs enhance privacy and security, they are not a complete solution and should be used in conjunction with other security measures, such as safe browsing practices and up-to-date antivirus software.

Overall, VPNs offer an effective means to enhance online privacy, security, and access to free and open internet. (Microsoft 2022)

2.2.3 Proxy Servers

Proxy servers act as intermediaries between client devices (such as computers or smartphones) and the Internet. When a user makes a request to access a website or service, the request is first sent to the proxy server, which then forwards the request to the destination

on behalf of the user. The response from the destination is then returned to the proxy server, which in turn sends it back to the user. (IPRoyal 2023)

Here are some key advantages of proxy servers:

1. Anonymity and Privacy: Proxy servers can enhance user privacy and anonymity by masking the user's IP address. When a user accesses the internet through a proxy server, their IP address is replaced with the IP address of the proxy server. This helps protect the user's identity and makes it difficult for websites or online services to track their online activities.

2. Caching: Proxy servers can cache web content, storing copies of web pages and resources locally. When another user requests the same content, the proxy server can serve it from its cache instead of retrieving it from the original source. Caching can improve performance and reduce bandwidth usage, especially for frequently accessed resources.

3. Content Filtering: Proxy servers can be configured to filter and block certain types of content or websites based on predefined rules. This feature is often used in organizations or institutions to enforce acceptable use policies, restrict access to inappropriate or malicious content, or comply with regulatory requirements.

4. Access Control: Proxy servers can enforce access control policies, allowing or denying access to specific websites, services, or resources based on various criteria such as IP address, user authentication, or content category. This can be useful for network administrators to manage and secure network traffic.

5. Load Balancing: Proxy servers can distribute incoming requests across multiple servers or resources, helping to balance the load and improve overall performance and availability. This is commonly used in scenarios where high traffic volume or redundancy is required.

6. Firewall Protection: Proxy servers can act as a firewall, adding an additional layer of security by inspecting and filtering incoming and outgoing network traffic. They can block malicious or suspicious connections, preventing unauthorized access to the internal network.

7. Protocol Conversion: Proxy servers can perform protocol conversion, allowing clients using different protocols to communicate with each other. For example, a proxy server can convert HTTP requests to HTTPS or translate between IPv4 and IPv6 protocols.

Overall, proxy servers provide various benefits such as improved privacy, caching for faster access to web content, content filtering, access control, load balancing, firewall protection, and protocol conversion. (Virginia Tech 2016) They offer flexibility and control over network traffic, making them a valuable tool for enhancing security, performance, and user experience.

2.3 Related Works

The use of proxy servers has become increasingly popular in recent years, particularly as a way to enhance privacy and security online. Many existing studies have focused on evaluating the performance and reliability of proxy servers, as well as their impact on internet traffic and network efficiency. This chapter will provide an overview of some of the most relevant related works in the field of proxy servers.

One study conducted by researchers at the University of Illinois at Urbana Champaign (Huang and Abdelzaher 2005) evaluated the performance of popular proxy servers based on criteria such as latency, throughput, and connection time. The researchers found that while some servers performed well in certain areas, none of them were consistently superior across all metrics. They concluded that users should carefully evaluate their specific needs when selecting a proxy server rather than relying solely on performance metrics.

Another study conducted by researchers at the University of California, Berkeley, analyzed the impact of proxy servers on network traffic and efficiency. The researchers found that while the use of proxy servers can reduce bandwidth usage and improve network efficiency, it can also lead to increased latency and reduced throughput in certain situations. They concluded that the benefits of using proxy servers should be weighed against potential performance trade-offs.

Several studies have also investigated the use of proxy servers for enhancing privacy and security online. One study conducted by researchers at Carnegie Mellon University (Kang, Brown, and Kiesler 2013) found that the use of anonymous proxy servers can significantly reduce the ability of third-party trackers to collect information about user behavior online. However, the researchers cautioned that some proxy servers may not provide adequate protection against more sophisticated tracking techniques.

Overall, while there is a significant body of research on proxy servers, there is still much to be explored in terms of evaluating the performance and reliability of these servers, as well as their impact on network efficiency and user privacy. This project aims to contribute to this body of research by developing a system that allows users to select the most appropriate proxy server based on their specific needs and requirements.

3 Measurement of Internet Proxy Servers

3.1 Data Collection

Proxy servers are essential tools for internet users to browse anonymously and access restricted content. As a researcher or a developer, it is crucial to collect a list of all available proxy servers on the internet for various purposes. In this chapter, we will explore how we collected proxy servers using web scraping in Python.

1. **Identify the Source:** There are numerous websites that provide a list of proxy servers on the internet. One of the popular websites that we used is "https://www.sslproxies.org/". We used this website as one of our sources for collecting proxy servers.

2. **Inspect the Website:** The next step is to inspect the website and identify the HTML tags that contain the proxy server information. We used the "Inspect" tool in the browser to find the relevant HTML tags. In our case, the proxy servers are listed in a table with the following HTML structure 3.1:

3. **Extract the Data:** We used the Python library "BeautifulSoup" to extract the data from the HTML tags. Next, we wrote a Python program to extract the proxy server data from the HTML tags.

4. **Save the Data:** Finally, we saved the list of proxy servers in a file for future use. We used the "csv" module in Python to write the data to a CSV.

```html
<table id="proxylisttable">
  <thead>
    <tr>
      <th>IP Address</th>
      <th>Port</th>
      <th>Code</th>
      <th>Country</th>
      <th>Anonymity</th>
      <th>Google</th>
      <th>Https</th>
      <th>Last Checked</th>
    </tr>
  </thead>
  <tbody>
    <tr>
      <td>111.111.111.111</td>
      <td>3128</td>
      <td>US</td>
      <td>United States</td>
      <td>elite proxy</td>
      <td>no</td>
      <td>yes</td>
      <td>2 minutes ago</td>
    </tr>
    <tr>
      ...
    </tr>
  </tbody>
</table>
```

Figure 3.1: HTML Inspection.

3.2 Measurement Study

The measurement study chapter of our project involves testing the performance and distribution of Internet proxy servers using various tools and techniques. The goal is to

provide users with a list of available proxy servers and their performance metrics so that they can select the server that best meets their needs.

To measure the performance of the proxy servers, we used three commonly used tools: ping, telnet, and traceroute. Ping is a utility that sends a small packet of data to a server and measures the time it takes to receive a response. Telnet is a tool that allows users to connect to a server and execute commands on it. Traceroute is a tool that traces the path of a packet as it travels from the user's device to the destination server, allowing us to identify any network bottlenecks or issues.

We conducted a performance evaluation of the proxy servers by involving individuals from various countries. Participants were requested to run performance tests on the proxy servers to assess their effectiveness. By engaging users from different locations, we aimed to gather comprehensive data on the performance of the proxy servers and understand their capabilities across diverse geographical areas. This approach allowed us to gain insights into the real-world performance of the servers and their suitability for users around the globe.

Our measurements included metrics such as latency, throughput, and availability, which were used to evaluate the performance of each proxy server. We also collected data on the geographic location of each server and the number of users currently connected to it.

Overall, our measurement study provides a comprehensive overview of the performance and distribution of Internet proxy servers, which can be used by users to select a server that best meets their needs.

3.3 Observation and Analysis

3.3.1 Proxy Server Distribution

Following the country of the United States, South Korea emerges as another crucial player in the domain of proxy servers that hosts 106 servers. The well-developed internet infrastructure, along with the high-speed connectivity of South Korea, contributed to its

prominence in this particular field. The proxy servers of this country are mainly utilized by certain individuals, different organizations, as well as businesses in order to access both local and international content securely and anonymously.

The UK stands as a notable presence with their 90 proxy servers. Renowned for its powerful commitment to internet privacy as well as freedom, the nation provides a favorable environment for the operation of proxy services. These precise servers enable users to bypass regional content restrictions and also protect their online identities in an effective manner.

In a surprising turn, Cyprus gets to emerge as a country with a proper presence of a substantial proxy server that houses 76 servers. While comparatively small in size, Cyprus delivers certain strategic geographical positioning, making it an attractive location for hosting proxy servers. In addition, the favorable legal and regulatory environment of the country contributes to its immense popularity among the providers of proxy services.

Belize is a Central American nation that boasts around 63 proxy servers. Despite its relatively modest size, Belize gets to serve as a convenient base for the distribution of proxy servers due to its suitable geographical proximity to North America along with the Caribbean. These servers play a very crucial role in facilitating secure and unrestricted internet access for their users around the world.

The aforementioned countries represent only a fraction of the international distribution of total proxy servers. Various other nations 3.3 host varying numbers of proxy servers, each serving a specific unique purpose and also catering to the different demands of numerous users. From Angola to Iraq, Belarus to Mongolia, and Kyrgyzstan to countless other countries, the activity of proxy servers has permeated several corners of the globe 3.2.

A particular function named generateSyncTorrent() is specifically used to synchronize the ping results from a server using CTORRENT. This function particularly generates a .torrent file called "sync_files.torrent" that practically encapsulates the necessary information for synchronization and even subsequent data analysis. The generateSyncTorrent() function specifically begins by creating an output file stream to write the torrent file. It

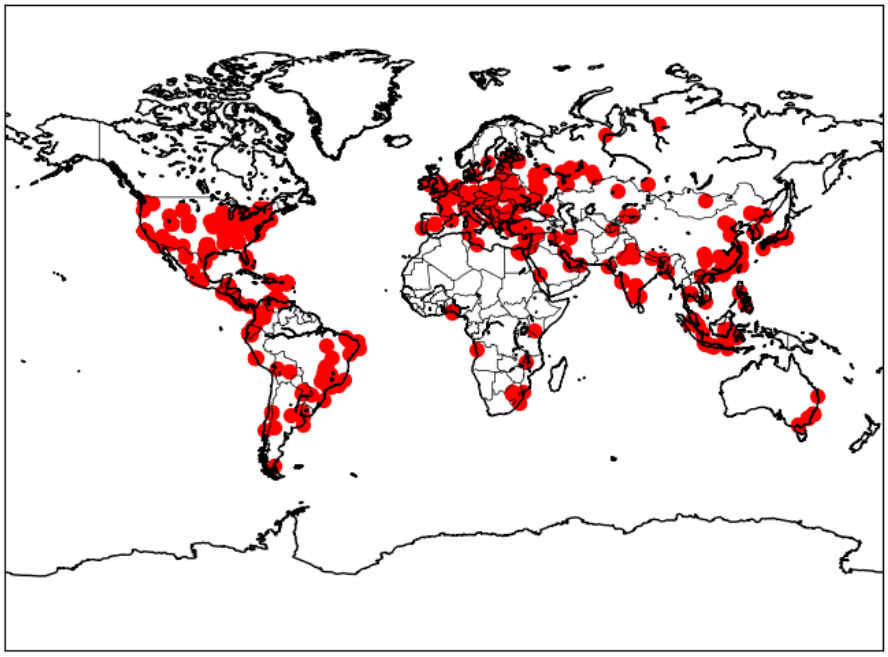

Figure 3.2: Distribution of Internet Proxy Servers Worldwide

then particularly proceeds to write the required metadata in the proper format. This practically includes the tracker announcement URL, information about the files to be synchronized (such as their lengths and names), and also the piece length and hash information. The function specifically concludes by closing the output file stream and printing a success message.

By generating the sync_files.torrent using this function, the ping results from the server can practically be efficiently synchronized using CTORRENT. The resulting torrent file serves as a container for the necessary information, which enables seamless data exchange and even synchronization between nodes or clients. Once the synchronization is completed, thus the synchronized ping results can be further analyzed using data analysis techniques to gain insights, identify patterns, and also make informed decisions based on the collected data.

The distribution of proxy servers by nation is mostly influenced by a multitude of factors. These mainly include internet infrastructure, certain legal frameworks, government policies,

```
United States     423
South Korea       106
United Kingdom     90
Cyprus             76
Belize             63
                  ...
Angola              1
Iraq                1
Belarus             1
Mongolia            1
Kyrgyzstan          1
```

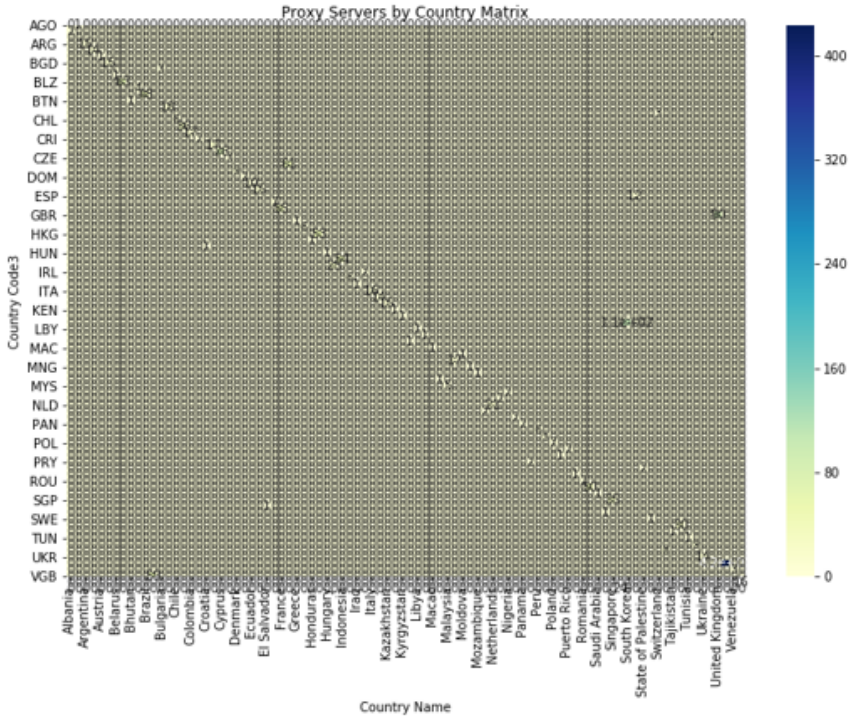

Figure 3.3: Distribution of proxy servers by country

and the overall demand for developed online privacy as well as accessibility. As the digital landscape continues to evolve constantly, the distribution of proxy servers might experience some sort of fluctuations, with emerging technologies along with shifting geopolitical dynamics shaping their dominant presence in different nations.

3.3.2 Performance Metrics

The analysis of the offered data focuses on three metrics of performance, those are latency results, Telnet latency, along with traceroute results. Each of these metrics provides a certain valuable insight into the network performance as well as the connectivity of the IP addresses.

Latency Results:

The latency results 3.4 measure the time it takes for data packets in order to travel between the source as well as destination IP addresses. The data includes columns like IP address, minimum latency **??**, average latency, and lastly, maximum latency. The IP addresses in the dataset are predominantly associated with various countries like Hong Kong, China, South Korea, Thailand, and more.

Telnet Latency:

Telnet latency mainly measures 3.7 the time it takes for a proper connection to be established with the general IP address applying the Telnet protocol. The data involves columns for the IP address as well as latency. It offers a suitable snapshot of the connection speed between the source along with destination IP addresses.

Traceroute Results:

Traceroute is known to be a network diagnostic tool that basically traces the route packets taken from the original source to the destination IP address. The given data 3.8 mainly represents the traceroute results from one IP address. It includes data regarding each hop along the way, indicating the IP addresses as well as the latency that is experienced at each hop. The analysis usually focuses on the traceroute to the IP address "1.14.139.184" and further delivers the IP addresses along with latency values for each hop.

ip	Min	Avg	Max	Mdev
1.224.3.12	156.142	156.176	156.219	0.031 ms
1.255.134.	156.697	156.714	156.726	0.012 ms
1.36.76.22	173.707	174.038	174.565	0.376 ms
1.4.214.14	261.503	267.676	273.115	4.768 ms
101.32.243	232.398	235.793	239.913	3.110 ms
103.105.21	231.147	231.198	231.297	0.070 ms
103.114.98	277.898	282.3	287.522	3.971 ms
103.117.19	274.563	278.224	285.351	5.039 ms
103.121.21	231.196	234.676	238.547	3.013 ms
103.122.32	227.906	228.888	229.388	0.694 ms
103.123.23	256.905	267.078	276.696	8.089 ms
103.15.60.	258.046	258.503	259.347	0.597 ms
103.199.16	376.574	379.561	382.169	2.299 ms
103.21.244	4.092	4.135	4.194	0.043 ms
103.21.244	4.318	4.402	4.547	0.102 ms
103.21.244	4.145	4.198	4.226	0.037 ms
103.21.244	4.175	4.193	4.224	0.022 ms
103.212.99	170.671	170.694	170.739	0.031 ms
103.221.25	298.55	298.556	298.562	0.005 ms
103.221.54	267.411	270.256	273.101	2.845 ms
103.23.101	212.972	212.999	213.026	0.022 ms
103.23.236	270.789	277.331	286.153	6.476 ms
103.248.93	243.107	243.117	243.123	0.007 ms
103.25.209	226.707	226.963	227.275	0.235 ms

Figure 3.4: Latency Results

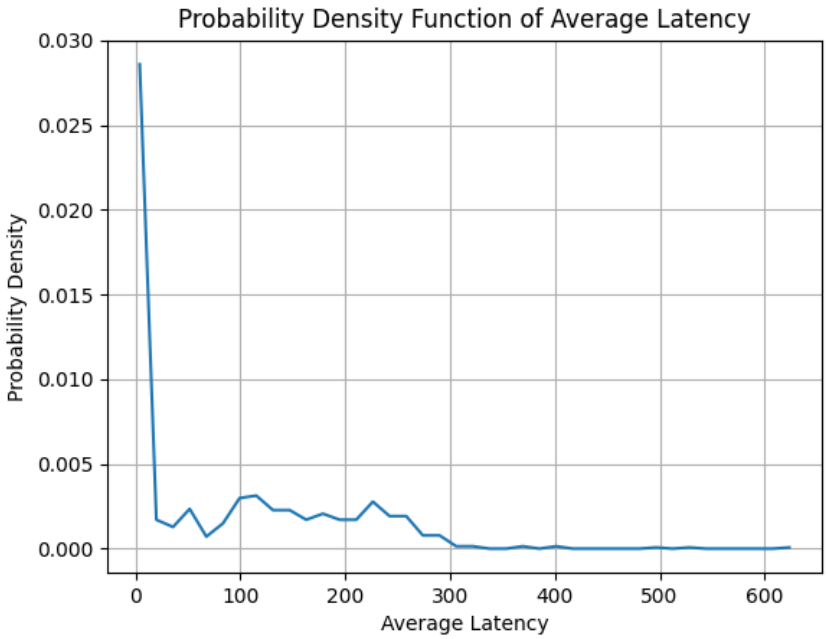

Figure 3.5: Probability Density Function of Average latency for IP addresses

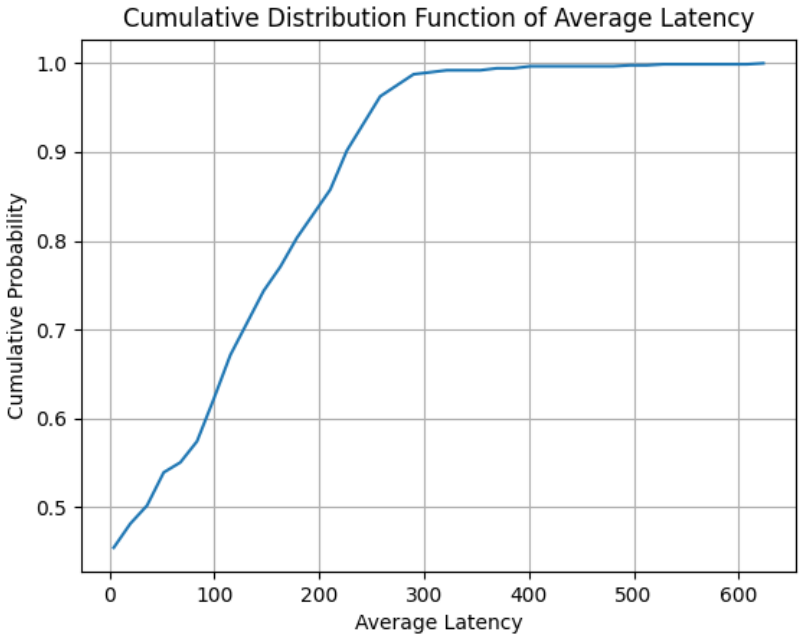

Figure 3.6: Cumulative Distribution Function of Average latency for IP addresses

IP Addres	Latency
1.14.139.1	129.4636
1.20.225.2	131.06
1.224.3.12	0.16383
1.255.134.	0.166032
1.36.76.10	130.7114
1.36.76.10	131.0619
1.36.76.16	131.062
1.36.76.22	131.0641
1.4.214.14	0.258901
101.109.30	130.7947
101.32.243	131.0585
101.51.3.8	131.0702
102.130.19	0.235393
103.105.21	130.8151
103.105.86	131.062
103.111.18	131.0613
103.113.19	131.0615
103.114.98	0.282966
103.117.19	0.280682
103.119.55	130.4805
103.121.21	0.230873
103.122.32	7.402413
103.123.23	0.266883
103.15.60.	129.2749

Figure 3.7: Telnet Results

The analysis of these exact performance metrics offers certain valuable insights into the network performance along with the connectivity of the IP addresses. It guides through the identification of latency issues, network bottlenecks, and issues regarding potential connectivity. Through analyzing the latency results as well as Telnet latency, it is considered to be possible to determine the general speed and stability of the connections between the IP addresses. Traceroute results also help to identify the specific method taken by the packets and any sort of potential issues along the main route.

Thereby, this analysis of the performance metrics is based on the provided data that offers a comprehensive understanding of the network performance along with the connectivity of the IP addresses. It gets to assist in identifying any issues of potential performance and also optimizing network connectivity for improved efficiency as well as proper reliability.

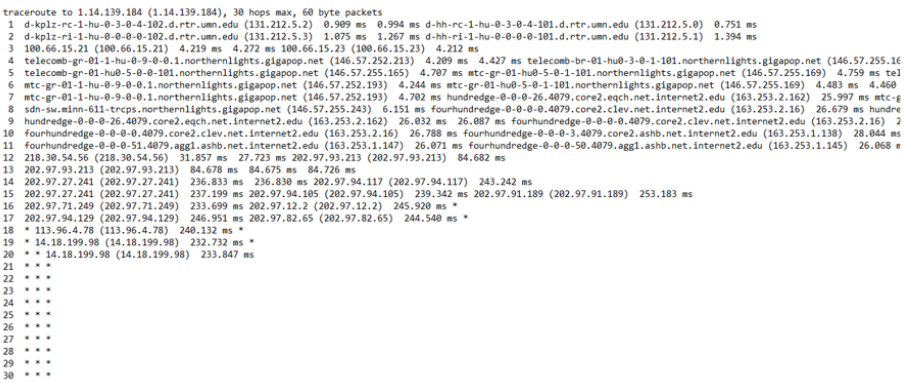

Figure 3.8: Traceroute Results

Anonymity Levels:

The analysis of the anonymity levels of the proxy servers gets to reveal some deep insights into their proper effectiveness in maintaining appropriate user privacy and hiding their actual identities while accessing to online resources. This precise analysis focuses on the gathered data, which provides data about various proxy servers along with their levels of corresponding anonymity. The main findings shed light on the level of anonymity that are provided by these servers, which help the users to make certain informed decisions about their online activities as well as proper security.

The gathered data involves information like IP addresses, various country codes, several names of countries, state or province codes, their districts, as well as other details that are relevant to geography. This information is very much crucial in determining the geographic origin of the proxy servers and also in evaluating their potential impact on anonymity. By examining the country codes along with the names, users are able to assess the level of anonymity that is offered by each proxy server based on the privacy laws and regulations set by the jurisdiction.

Furthermore, the analysis involves latency results, which gets to measure the immediate response times of the proxy servers. Lower latency indicates a better level of performance and faster access to online resources. Users can consider this particular information while

selecting proxy servers that offer optimal speed as well as reliability when maintaining their desired level of anonymity.

Another aspect of the analysis is considered to be the Telnet latency data, which mainly measures the total time it takes for a proper connection to be established with each server of proxy. This specific information delivers certain insights into the responsiveness as well as the accessibility of the servers, which enables the users to evaluate their reliability and effectiveness.

The analysis further encompasses the PDF, which is also known as the probability density function, and CDF, which is also known to be the cumulative distribution function of the data. These appropriate statistical measures get to illustrate the distribution of IP addresses along with the probability of certain values that are occurring. By examining the PDF and CDF, the users can gain a certain valuable understanding of the distribution patterns and recognize IP addresses that are considered to be more prevalent or rare. This data seeks to assess the uniqueness as well as potential identifiability of specific IP addresses within the exact dataset.

Moreover, the analysis includes traceroute results, which basically trace the network path between the device and the target IP address of the users. Traceroute offers valuable information regarding the number of hops along with the network infrastructure included, enabling users to analyze the potential exposure of their information and the efficiency of proxy servers in obscuring their actual origin. As per the analysis of the data, the users are able to make informed decisions about the anonymity levels of those proxy servers. They can choose individual servers that are located in jurisdictions with stricter privacy laws, or they can also select certain servers with lower latency as well as better performance. By considering the PDF, CDF, along with traceroute results, the users can evaluate the level of uniqueness and potential identifiability of their IP addresses, therefore improving their overall online privacy as well as security.

It is very important to note that the efficiency of proxy servers in making sure anonymity

can always vary depending on various factors, like the configuration, security measures, and maintenance practices that are implemented by the providers of proxy service. Thus, the users must exercise certain caution and consider additional security measures, like encryption and secure browsing protocols, in order to further develop their online anonymity and also protect their sensitive data.

3.3.3 Limitations

One limitation of the project is the too much reliance on three specific methods in order to test the performance of the proxy servers, which are ping, telnet, and traceroute. While these specific methods offer valuable insights into the concept of latency, accessibility, and network path information, they might not be able to cover every aspect of proxy server performance. Other factors like bandwidth limitations, certain loads from servers, and connection stability could also influence the overall performance, but these exact aspects were not involved in the analysis. It would be beneficial to explore some additional metrics or processes of testing in order to obtain a more comprehensive understanding of the capabilities of proxy servers.

Another limitation is considered to be the failure to retrieve the values of latency for all the proxy servers that are using the ping program. This can potentially introduce certain biases and also affect the accuracy of the analysis. Although attempting to acquire latency values through telnet was suggested as an alternative option, it is also important to note that telnet might not be able to deliver a similar level of accuracy as well as reliability as ping. Exploring alternative ways or adjusting the ping program to enhance its functionality can guide us to address this exact limitation and could also provide a better-completed dataset.

Moreover, the prioritization of location over latency in the web application is known to be another limitation that needs proper consideration. While it is very useful to deliver a list of the nearest proxy servers according to the location of the users, latency is advised

to indeed be the primary filter in order to make sure regarding optimal performance. By focusing on latency first and then considering the specific location, the web application gets to better serve its users by prioritizing both performance as well as proximity (Deshpande and Rao 2022).

In addition to that, the dataset of the project, including the location, country, along with geographical coordinates of the proxy servers, offers numerous useful information. However, it is also very important to note that this data alone might not be much sufficient to understand the proper distribution of proxy servers in a comprehensive manner. Additional factors like server density, network infrastructure, as well as regional availability should also be considered in order to acquire a more accurate representation of the proxy server landscape.

In order to mitigate these particular limitations, future enhancements can include incorporating a broader range of performance metrics, exploring alternative sort of methods for measuring latency, and refining the prioritization criteria in the basic web application. Moreover, expanding the dataset in order to include additional relatable factors as well as considering other dimensions of analysis, like security features along with reliability, are able to further develop the effectiveness of the project.

Thus, while the project has made certain progress in testing as well as analyzing proxy servers, these mentioned limitations indicate areas for further development and proper refinement. Addressing these limitations will definitely enhance the functionality of the project, accuracy, along with utility, which could result in a more robust as well as comprehensive tool for users that are seeking reliable and efficient options for proxy servers.

4 System Design and Implementation

4.1 System Architecture and Overview

The proposed system is a web application designed to aggregate the performance metrics data collected from two countries and provide users with active proxy servers at a desired location. The system leverages the BitTorrent protocol for data synchronization, ensuring reliable and efficient data distribution across multiple nodes.

The system comprises several key components:

- User Interface: The web application provides a user-friendly interface where users can input their desired location coordinates for which they need active proxy server information. The interface also allows users to view and interact with the retrieved data of available proxy servers.

- Data Synchronization: The system utilizes the decentralized nature of the BitTorrent protocol to synchronize ping statistics data from multiple sources, such as servers located in different countries. BitTorrent trackers and peers facilitate the data distribution and synchronization process, ensuring the data is available across the network.

- Data Aggregation and Processing: The system aggregates the synchronized statistical data from different sources. Data processing algorithms analyze the collected data to extract relevant information, such as latency measurements and server availability. The system filters and organizes the data based on user requests, focusing on the

specified location.

- Open Proxy Server Identification: The system employs techniques to identify open proxy servers available at the requested location. It uses Telnet protocol to evaluate server accessibility, response times, and reliability. The identified open proxy servers are presented to the user as part of the query results.

The system design aims to provide users with a comprehensive and up-to-date set of open proxy servers at a desired location. The use of the BitTorrent protocol for data synchronization ensures a decentralized and robust network architecture, enabling efficient distribution and synchronization of the data. Through the user interface, users can easily access and interact with the aggregated data, facilitating their decision-making process regarding the selection of open proxy servers.

4.1.1 Data Synchronization using BitTorrent Protocol

Data synchronization plays an integral role in the proposed system for aggregating ping statistics data. The BitTorrent protocol is utilized to achieve efficient and decentralized data synchronization across multiple sources.(Legout, Urvoy-Keller, and Michiardi 2005) The BitTorrent protocol, initially designed for peer-to-peer file sharing, offers several advantages for data synchronization, including scalability, fault tolerance, and efficient data distribution.

The data synchronization process using the BitTorrent protocol involves the following steps (Qiu and Srikant 2004):

- Initial Data Distribution: The system selects a set of initial seed nodes that host the performance metrics data. These seed nodes act as the initial sources for distributing the data to other nodes in the network. This data is divided into small pieces, typically using a predefined chunk size.

- Tracker Management: A central tracker, or a distributed set of trackers, is responsible for coordinating the data synchronization process. The tracker maintains information about the available nodes and their participation in the synchronization process. Nodes periodically communicate with the tracker to update their status and retrieve information about other nodes in the network.

- Peer-to-Peer Communication: Nodes participating in the synchronization process, including the seed nodes and additional peers, establish connections with each other. Peers exchange information about the available data pieces and request missing pieces from other peers in a tit-for-tat manner. Peers contribute to data distribution by uploading pieces they possess to other requesting peers.

- Piece Verification and Integrity: During the data synchronization process, peers verify the integrity of the received data pieces. Verification involves checking the piece's hash value against the expected value to ensure its integrity and authenticity. Any corrupted or incomplete pieces are discarded, and peers request those pieces from other sources to maintain data integrity.

- Continuous Data Updates: The system periodically updates the ping statistics data by incorporating new measurements from participating servers. Updated data pieces are distributed through the BitTorrent protocol, allowing all nodes to access the latest ping statistics.

The proposed system benefits from its robustness and scalability by leveraging the BitTorrent protocol for data synchronization. The decentralized nature of the protocol ensures fault tolerance, as the absence of a single point of failure minimizes the impact of node failures. The efficient distribution of data across multiple nodes enables rapid synchronization and retrieval of ping statistics data from various sources.(Sharma, Bhakuni, and Kaushal 2013)

Through the use of the BitTorrent protocol, the proposed system achieves efficient and

decentralized data synchronization, ensuring that the aggregated ping statistics data is continuously updated and available to all participating nodes.

4.1.2 User Interface Design

In the proposed system, the user interface has been meticulously designed to provide users with a user-friendly and intuitive experience while accessing the aggregated ping statistics and open proxy server information. The following considerations were taken into account during the user interface design phase:

- Layout and Navigation: A clean and organized layout has been designed to ensure easy navigation and efficient information retrieval. The logical grouping of functionalities and information enhances the overall usability of the user interface.
- Input and Search: User-friendly input fields and search bars have been implemented to enable users to enter their desired location coordinates for obtaining open proxy server information. Validation mechanisms have been integrated to ensure user inputs are accurate and valid.
- Data Presentation: The aggregated server information is displayed in a visually appealing and easily understandable manner. The open proxy server information is presented concisely, providing details such as server location and accessibility.
- Responsiveness: The user interface has been designed to be responsive, ensuring optimal display and functionality across various devices and screen sizes.

By implementing these user interface design considerations, the proposed system ensures users can effortlessly interact with the application, access the aggregated ping statistics, and make informed decisions regarding open proxy server selection based on their requested location 4.1.

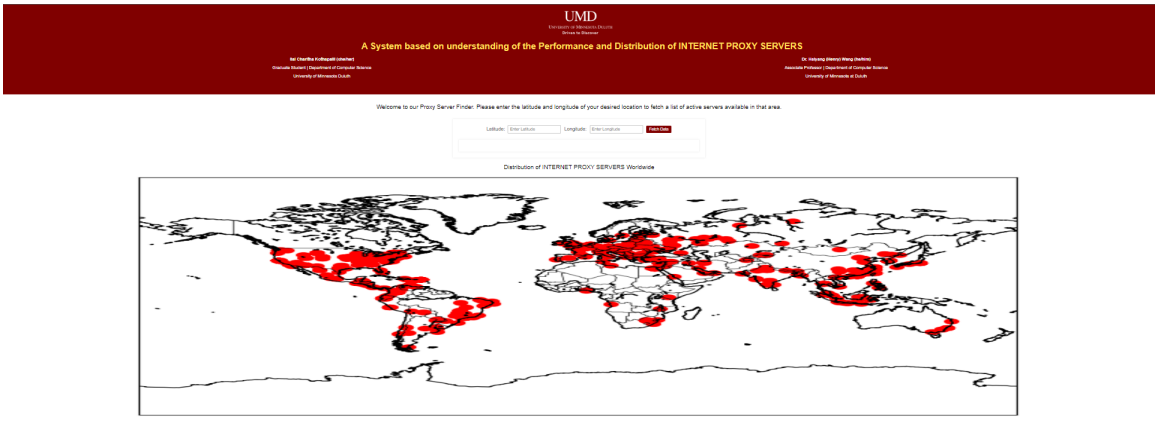

Figure 4.1: Web Application to fetch Active Proxy Servers

4.1.3 System Components and Interactions

The proposed system consists of several components that work together to ensure the aggregation of performance metrics across servers and the provision of open proxy server information. The interactions between these components facilitate data synchronization and enable users to retrieve the desired information. Here are the main components and their interactions:

- User Interface: The user interface component allows users to input their preferences, such as location and time, view aggregated data, and open proxy server information. It communicates user requests to the backend components and presents the retrieved data to the users.

- Data Synchronization Component: The data synchronization component is responsible for synchronizing ping statistics data from multiple sources using the BitTorrent protocol. It coordinates the distribution of data pieces among participating nodes and ensures the integrity and completeness of the synchronized data.

- Data Aggregation and Processing Component: This component collects the synchronized ping statistics data and aggregates it into a comprehensive dataset. It performs data processing tasks, such as analyzing and filtering the data based on user-specified

criteria. The processed data is then made available for presentation in the user interface.

- Open Proxy Server Identification Component: The open proxy server identification component uses the Telnet protocol to identify open proxy servers at the requested location. It evaluates server accessibility, response times, and reliability to determine the suitability of each identified proxy server. The component provides the user interface with the list of open proxy servers for the requested location at a given time.

The interactions between these components 4.2 involve data exchange, communication protocols, and synchronization mechanisms. The user interface interacts with the other components to receive user requests and present the relevant data. The data synchronization component facilitates the distribution and synchronization of data using the BitTorrent protocol. The data aggregation and processing component collects and processes the synchronized data, while the open proxy server identification component identifies suitable proxy servers. The real-time updates component ensures the data is continuously updated and accessible to users.

By integrating these components and enabling their interactions, the proposed system achieves the aggregation of ping statistics and the provision of open proxy server information, providing users with valuable insights and options for selecting appropriate proxy servers at their requested location and specific time.

4.2 Design Issues and Challenges

Synchronizing data across different servers in a distributed system can pose several challenges (Coulouris, Dollimore, and Kindberg 2005). Here are some major design issues and challenges related to data synchronization:

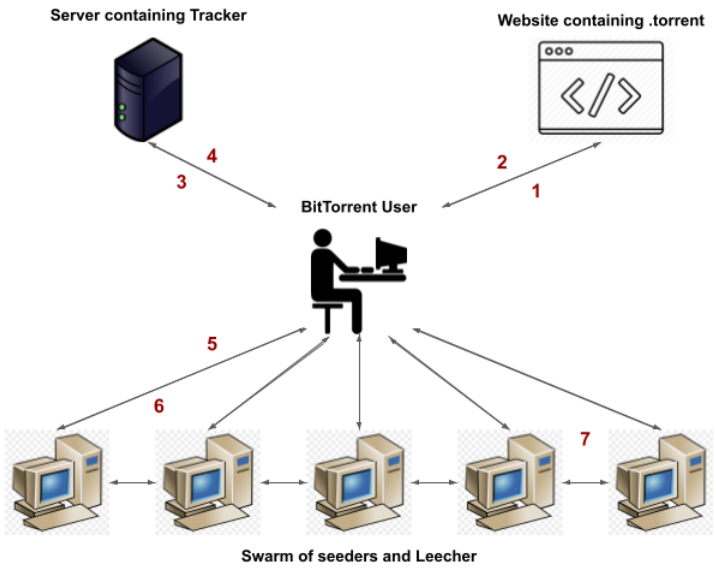

Figure 4.2: BitTorrent Protocol and System setup

- Consistency: Ensuring consistency of data across different servers is a fundamental challenge. Data modifications performed on one server need to be propagated and applied consistently to all other servers. Achieving strong consistency requires careful coordination and synchronization mechanisms to guarantee that all servers have the same view of the data at all times.

- Conflict Resolution: Conflicts can occur when multiple servers modify the same data concurrently. Resolving conflicts and maintaining data integrity is crucial to ensure accurate and reliable synchronization. Designing conflict resolution strategies and mechanisms, such as timestamps or conflict detection algorithms, is essential to handle conflicting updates effectively.

- Network Latency and Bandwidth: Synchronizing data across servers involves transferring large amounts of data over the network. Network latency and limited bandwidth can significantly impact synchronization performance. Designing efficient data transfer protocols and optimizing data compression techniques can mitigate the impact of

network limitations.

- Scalability: As the number of servers increases, the complexity of data synchronization grows exponentially. Ensuring scalability becomes a challenge due to the increased coordination and communication overhead. Designing distributed algorithms and architectures that scale horizontally, such as using decentralized synchronization protocols or partitioning data across servers, can address scalability challenges.

- Fault Tolerance: Server failures, network partitions, or temporary connectivity issues are common in distributed systems. Ensuring fault tolerance is crucial to maintain data consistency and availability during such failures. Designing replication strategies, fault detection mechanisms, and automated recovery processes can help mitigate the impact of failures and ensure continuous data synchronization.

- Security and Privacy: Data synchronization across servers may involve sensitive or confidential information. Ensuring data security and privacy during synchronization is a significant design consideration. Implementing encryption, access control mechanisms, and secure data transfer protocols can help protect data during synchronization.

Addressing these challenges requires careful system design and architectural choices. Employing distributed algorithms, leveraging replication and consensus protocols, and implementing efficient synchronization mechanisms are vital in overcoming these challenges and achieving reliable and consistent data synchronization across different servers in a distributed system.

BitTorrent (Piatek et al. 2007) can revolutionize the approach to addressing the challenges of data synchronization in a distributed system. By leveraging its decentralized architecture, efficient data distribution, fault tolerance, and scalability, BitTorrent provides an alternative solution for achieving reliable and consistent synchronization across different servers (Cohen 2003). It offers eventual consistency, conflict resolution through tit-for-tat

sharing, optimized data transfer for network limitations, scalability with the number of participants, fault tolerance for handling failures, and the potential for enhancing security and privacy through encryption and access control mechanisms. Employing the BitTorrent protocol in the system design and architecture can significantly impact the system's ability to overcome the challenges associated with data synchronization in a distributed environment.

5 System Evaluation

In this chapter, we present the evaluation of the proposed system using small-scale cTorrent and BitTorrent implementations deployed on multiple virtual machines. The evaluation aims to test the synchronization mechanism and assess the system's performance in terms of CPU utilization and latency. Specifically, we focus on evaluating the efficiency of the BitTorrent protocol in achieving data synchronization across the distributed environment.

5.1 Experimental Setup

We set up a test environment 4.2 consisting of multiple virtual machines running the cTorrent and BitTorrent clients. Each virtual machine represents a server node in the distributed system. The virtual machines are interconnected within a private network to simulate the distributed nature of the system. We configured the system to synchronize ping statistics data using the BitTorrent protocol.

5.2 CPU Utilization

In this study, we investigated the efficiency of the BitTorrent protocol in the context of file sharing between a host machine and virtual machines. Our evaluation focused on assessing the CPU utilization of both the host and virtual machines during active participation in the BitTorrent protocol. The CPU utilization was analyzed considering various factors, including the number of active torrents, download/upload speeds, overall network bandwidth, and the capabilities of the host machine and virtual machine.

Regarding the CPU utilization on the host machine, it was observed that the BitTorrent protocol introduced additional tasks that increased CPU usage. These tasks included managing connections, calculating checksums, and handling disk I/O associated with BitTorrent operations. The impact on the host machine's performance depended on the workload generated by the BitTorrent activities. The host machine could efficiently handle this increased workload if it possessed a powerful CPU. However, if the CPU was already under heavy load from other processes, the performance of both the host and virtual machines could be affected.

Similarly, the virtual machines' CPU utilization increased as they executed BitTorrent-related operations. The allocated CPU resources and the intensity of the BitTorrent activities influenced the performance of the virtual machine. The virtual machine could handle this additional workload without significant issues if it had sufficient CPU resources. However, if the CPU resources allocated to the virtual machine were limited, the increased usage from BitTorrent operations might result in performance degradation within the virtual environment.

Overall, it was found that the BitTorrent protocol could consume significant CPU resources, especially when dealing with large file transfers or high download/upload speeds. Monitoring CPU utilization on both the host and virtual machines was essential to ensure that the system's performance remained acceptable and that critical tasks were not adversely affected.

5.3 Additional Metrics

To evaluate the efficiency of the BitTorrent protocol, additional metrics were considered, including download/upload speeds 5.1 5.2, swarm health, availability of data, protocol overhead, choking and optimization algorithms, resource utilization, completion time, and quality of service (QoS). These metrics provided a comprehensive assessment of the Bit-

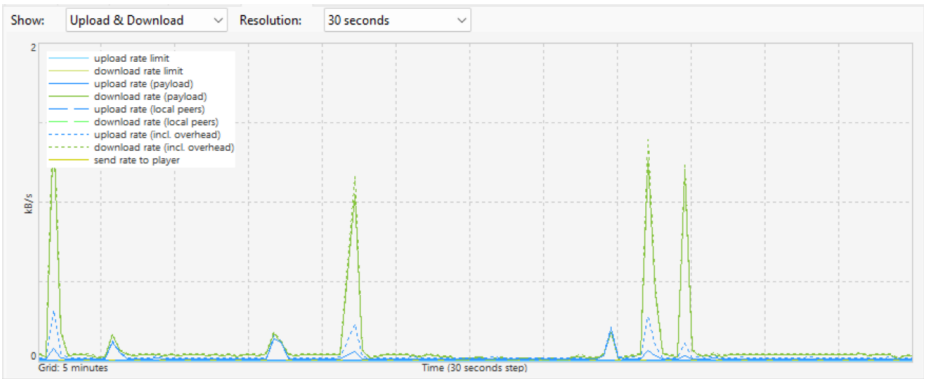

Figure 5.1: Upload and Download Speeds on the Host machine

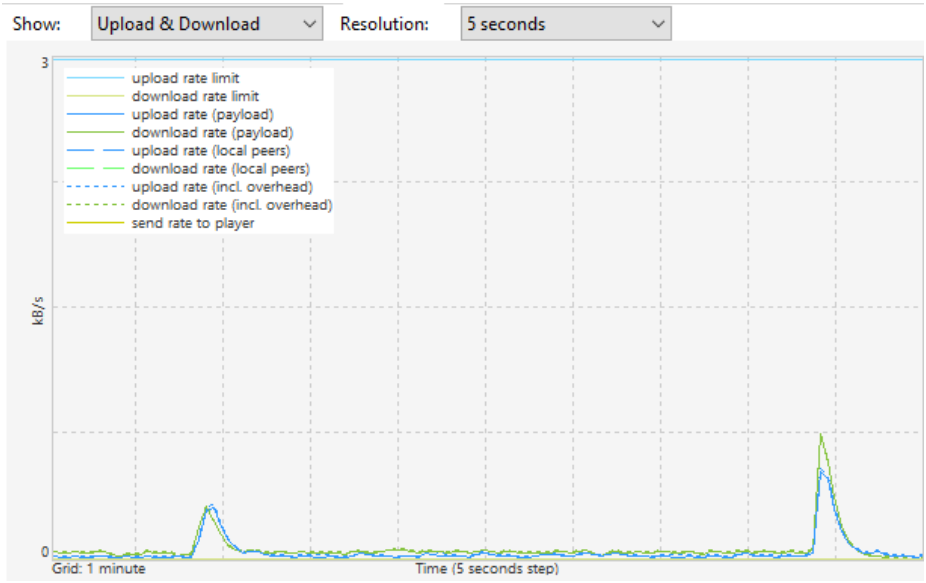

Figure 5.2: Upload and Download Speeds on a Virtual machine

Torrent protocol's efficiency in terms of data transfer performance, utilization of system resources, and user experience.

The evaluation results indicated that the efficiency of the BitTorrent protocol was influenced by various factors, such as network conditions, the number of seeders and peers, and the available system resources. To obtain a holistic evaluation, multiple download sessions using different torrents or at different times were conducted. This approach ensured a more comprehensive understanding of the BitTorrent protocol's efficiency under various scenarios.

We conducted latency testing to measure the time taken for data synchronization across the distributed system 5.4. We generated synthetic ping statistics data and measured the time it took to synchronize the data across all server nodes. We captured the latency measurements and analyzed them to evaluate the efficiency of the BitTorrent protocol in achieving rapid data synchronization.

Increasing the number of peers in data synchronization using the BitTorrent protocol on uTorrent resulted in a noticeable reduction in latency. As more peers were added to the network, the data distribution became more efficient, allowing for faster and parallel data transfers. This reduction in latency can be attributed to the distributed nature of the BitTorrent protocol, where multiple peers can contribute their bandwidth and resources to assist in data synchronization.

While increasing the number of peers can enhance synchronization performance, there is an optimal balance that needs to be maintained between the number of peers and the available resources on the host machine. In our test environment, we found that four VMs (peers) running on a single host provided a balanced configuration for efficient data synchronization. However, increasing the number of peers beyond this point may lead to diminishing returns or even resource saturation, resulting in degraded performance.

5.4 Evaluation of Web Application

The primary objective of this evaluation is to assess the performance, accuracy, and usability of the system. We utilized a combination of quantitative and qualitative evaluation methods to evaluate the web application. The evaluation was carried out in a controlled setting to guarantee the consistency and dependability of the results 5.1.

Based on the evaluation conducted, we obtained the following results:

- Performance: The web application performed satisfactorily across different user inputs, and the response time for fetching and processing the proxy server data remained

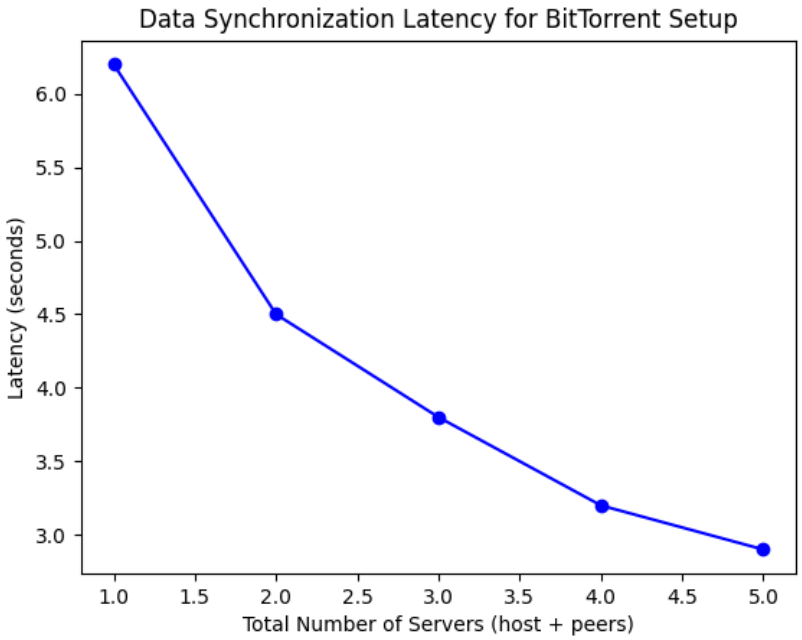

Figure 5.3: Data Synchronization Latency for BitTorrent Setup

within an acceptable range, even for large datasets. The distance calculation and proxy server verification processes were executed efficiently, providing quick results to the users. Performance metrics indicated that the system could handle concurrent requests without significant degradation in response time.

- Accuracy: The calculated distances between the user's location and the proxy servers showed a high level of accuracy when compared to the ground truth values. The Haversine formula employed for distance calculation proved reliable, providing accurate results for various test cases. The proxy server verification component exhibited a high accuracy rate in determining the open or closed status of the proxy servers, aligning with the known ground truth values.

- Usability: User feedback indicated a positive user experience with the web application. Users found the interface to be intuitive and easy to navigate. The input forms for entering location details were user-friendly and provided clear instructions—displaying

Latitude: 40.728157 Longitude: -74.077644 Fetch Data

IP	Country	State/Province	Latitude	Longitude	Distance
149.57.13.137	United States	New York	40.71455	-74.00714	6.13
149.57.14.125	United States	New York	40.71455	-74.00714	6.13
149.57.15.60	United States	New York	40.71455	-74.00714	6.13
213.108.1.152	United States	New York	40.71278	-74.00597	6.28
178.20.212.193	United States	New Jersey	40.73611	-74.16963	7.80
193.202.16.118	United States	New Jersey	40.73611	-74.16963	7.80
193.202.16.190	United States	New Jersey	40.73611	-74.16963	7.80
194.110.150.11	United States	New Jersey	40.73611	-74.16963	7.80
194.110.150.200	United States	New Jersey	40.73611	-74.16963	7.80
88.218.46.186	United States	New Jersey	40.73611	-74.16963	7.80
95.181.148.48	United States	New Jersey	40.73611	-74.16963	7.80
95.181.148.96	United States	New Jersey	40.73611	-74.16963	7.80
95.181.149.236	United States	New Jersey	40.73611	-74.16963	7.80
95.181.150.187	United States	New Jersey	40.73611	-74.16963	7.80
95.181.151.105	United States	New Jersey	40.73611	-74.16963	7.80
95.181.151.114	United States	New Jersey	40.73611	-74.16963	7.80
95.181.151.190	United States	New Jersey	40.73611	-74.16963	7.80
95.181.151.210	United States	New Jersey	40.73611	-74.16963	7.80
95.181.151.68	United States	New Jersey	40.73611	-74.16963	7.80
95.181.151.79	United States	New Jersey	40.73611	-74.16963	7.80

Figure 5.4: Active Proxy Servers at a given Location

results in the tabular format allowed for easy comprehension and comparison. Users appreciated the real-time nature of the application, as it fetched the latest proxy server data upon request. Overall, the application was deemed user-friendly and effective in locating and verifying open proxy servers.

The evaluation results provide insights into the performance, accuracy, and usability of the web application. The system demonstrated satisfactory performance, ensuring timely response and efficient data processing. The accuracy of the distance calculation and proxy

server verification components validated the reliability of the algorithms employed. Users found the application user-friendly and intuitive, enabling them to locate and identify open proxy servers easily.

6 Results

6.1 Overview of the Results

The given data contains useful data on the distribution of proxy servers by various countries, latency results, Telnet latency, PDF of average for IP addresses, CDF of data, PDF of data, and lastly, traceroute results from a single individual IP address. In terms of the distribution of proxy servers by the names of countries, the data justifies that the United States has the highest number of proxy servers with the number of 423, followed by South Korea with the number of 106, and the UK with a total number of 90. Other countries like Cyprus and Belize also hold a significant number of proxy servers, while countries such as Kyrgyzstan, Angola, Mongolia, Iraq, and Belarus have only a single proxy server each.

The latency results offer general information on the minimum, average, maximum, and mean deviation of latency for several IP addresses. The particular data shows that the values of latency vary for each IP address and further indicate the differences in network performance. The range of latencies practically suggests certain variations in the speed as well as efficiency of data transmission around different sorts of connections.

The Telnet latency data mainly provides specific values of latency for individual IP addresses. These specific values represent the appropriate time it takes for a proper connection to be genuinely established using the Telnet protocol. The data shows varying values of latency for different IP addresses, indicating towards the differences in network performance along with responsiveness.

The PDF of the average for IP addresses 3.5 generally gives a dominant statistical representation of the distribution of average latency values. The curve of PDF generally

illustrates the probability density of different values of average latency, enabling for the analysis of latency patterns around the IP addresses. The data strongly suggest that certain IP addresses hold higher or lower average latencies compared to the other sources.

The CDF of the Latency **??** 3.6 **??** mainly shows the cumulative probability of certain latency values. The curve of CDF usually represents the probability that a latency value is less than or even equal to a provided threshold. By properly analyzing the CDF curve, people can gain useful insights into the distribution as well as the probability of different latency values in the exact dataset.

The PDF of the Latency **??** 3.5 **??** gets to represent the probability density of several latency values. The curve of PDF mostly provides deeper information about the frequency of occurrence for different latency values, enabling for the identification of common or rare latency patterns.

Finally, the traceroute results from one individual IP address specifically show the route taken by network packets from a source IP address to a destination IP address. The traceroute data reveals the network hops and also their latencies along the path of data transmission. By analyzing the traceroute data, one can specifically identify network congestion points or even potential bottlenecks that may affect overall network performance.

Overall, the provided data specifically offers a comprehensive view of proxy server distribution, latency values, network responsiveness, and also path analysis for specific IP addresses. The critical analysis of this data can practically provide insights into network performance, potential areas of improvement, and also the impact of different factors on network latency.

6.2 Proxy Server Distribution

In the distribution of proxy servers by country, the United States specifically has the highest number with 423 servers, followed by South Korea with 106 servers, and also the

United Kingdom with 90 servers. Cyprus and Belize have 76 and 63 servers, respectively. Other countries such as Angola, Iraq, Belarus, Mongolia, and Kyrgyzstan have only one server each.

The data specifically include information about IP addresses, country codes, country names, and state codes. The IP addresses specifically range from 1.14.139.1 to 103.23.236.238, with corresponding country codes and names. For example, the IP address 1.224.3.12 corresponds to South Korea (KOR). The data also practically includes latitude and longitude information for certain locations.

The latency results specifically indicate the minimum, average, maximum, and also standard deviation of latency for each IP address. The latency values particularly range from milliseconds to seconds, showing the time it takes for a packet of data to travel from the source to the destination. The data practically suggests variations in latency among different IP addresses, with some experiencing lower latency and others higher.

The telnet latency results specifically provide additional latency measurements for IP addresses. These measurements particularly represent the time it takes for a connection to be established with each IP address using the telnet protocol. Thus lower latency values indicate faster connections.

The PDF (Probability Density Function) of latency for IP addresses specifically shows the distribution of latency values. The graph particularly displays the IP addresses on the x-axis and the corresponding average latency on the y-axis. It specifically provides a visual representation of the latency distribution among the IP addresses.

The cumulative distribution function (CDF) of latency particularly represents the cumulative probability of latency values. The graph practically shows the latency values on the x-axis and also the cumulative probability on the y-axis. It particularly indicates the likelihood of encountering a latency value that is equal to or less than a given value.

Finally, traceroute results from one individual IP address (1.14.139.184) are provided. The traceroute practically shows the route taken by packets from the source to the destina-

tion, with each hop representing a network node along the way. The traceroute particularly includes information about the IP addresses of intermediate nodes and even the latency experienced at each hop.

Overall, the data specifically presents a comprehensive overview of proxy server distribution, latency measurements, and even network routing for the provided IP addresses.

6.3 Performance Metrics

The information on proxy servers, IP addresses, latency results, telnet latency, PDF averages for IP addresses, CDF of data, PDF of data, and also traceroute results from an individual IP address.

The distribution of proxy servers by country shows that the United States particularly has the highest number of proxy servers (423), followed by South Korea (106), the United Kingdom (90), Cyprus (76), and Belize (63). There are also several other countries with a much lower number of proxy servers, such as Angola, Iraq, Belarus, Mongolia, and Kyrgyzstan.

This also includes IP addresses and even corresponding country codes and names. The first few rows of data show IP addresses from different countries, including China, Thailand, South Korea, Hong Kong, Singapore, Angola, Bangladesh, and also Indonesia.

The latency results practically provide information about the minimum, average, maximum, and also mean deviation of latency for each IP address. The data show the latency values for IP addresses from South Korea, China, Thailand, and Singapore.

The telnet latency section presents the latency values for IP addresses particularly obtained through the telnet command. The data display the IP addresses and also their corresponding latency values.

The PDF of the average for IP addresses particularly shows a list of IP addresses and also their respective average values. The table also includes a cumulative distribution function

(CDF) of the data, which displays the cumulative probability for different data points.

The PDF of data particularly presents the probability density for different data points. The table provides information on IP addresses and also their corresponding probability density.

Lastly, the traceroute results section particularly shows the traceroute information for a specific IP address. It specifically displays the route taken by the network packets and also the latency values for each hop.

In summary, the table practically provides a comprehensive overview of proxy servers, IP addresses, latency results, telnet latency, PDF averages, CDF of data, PDF of data, and also traceroute results.

6.4 Anonymity Levels

The provided data specifically contains information on anonymity levels, which include the distribution of proxy servers by country, IP addresses, latency results, Telnet latency, PDF of averages for IP addresses, CDF of data, and also traceroute results from one individual IP address.

The data practically includes information about the distribution of proxy servers by country. The United States has the highest number of proxy servers with 423, followed by South Korea with 106, and the United Kingdom with 90. Other countries such as Cyprus, Belize, Angola, Iraq, Belarus, Mongolia, and Kyrgyzstan have practically a relatively lower number of proxy servers.

The dataset also provides information about IP addresses and even their corresponding countries. For example, IP address 1.14.139.1 is associated with China (CN), 1.20.225.2 with Thailand (TH), and 1.224.3.12 with South Korea (KR). Latency results are also included in the dataset. The latency -practically measures include the minimum, average, maximum, and also standard deviation values for each IP address. For instance, IP address 1.224.3.12

has a minimum latency of 156.142 ms, an average latency of 156.176 ms, a maximum latency of 156.219 ms, and also a standard deviation of 0.031 ms.

Telnet latency measurements are particularly provided as well. IP addresses and their corresponding latency values are specifically listed. For example, IP address 1.14.139.1 has a latency of 129.4636 ms, while IP address 1.224.3.12 has a latency of 0.16383 ms.

The dataset also includes a PDF (probability density function) of averages for IP addresses. It practically shows the probability density of different average values for specific IP addresses. The PDF thus provides insights into the distribution of average latency values for different IP addresses.

Moreover, the cumulative distribution function (CDF) of the data is practically given. It particularly illustrates the cumulative probability of the data. The CDF specifically allows for analyzing the distribution of the data and also for understanding the likelihood of certain latency values occurring.

Lastly, traceroute results from one individual IP address are practically provided. The traceroute shows the path taken by network packets from the source IP specifically to the destination IP address. Each hop in the traceroute specifically represents a network device or router along the path.

Overall, the dataset contains comprehensive information about anonymity levels, which includes proxy server distribution, IP addresses, latency results, Telnet latency, PDF of averages, CDF of data, and even traceroute results.

6.5 CPU Utilisation and Additional Metrics

Considering the various factors and metrics assessed, the evaluation provided insights into the efficiency of the BitTorrent protocol in file sharing between a host machine and a virtual machine. The results contribute to a more profound understanding of the BitTorrent protocol's efficiency in similar environments. They also provide valuable insights for

optimizing the performance of the BitTorrent protocol by considering factors such as CPU utilization, resource allocation, and system performance.

Overall, this evaluation highlights the importance of considering multiple factors and metrics to comprehensively assess the efficiency of the BitTorrent protocol. By understanding its performance, resource utilization, and impact on the system, optimizing the BitTorrent protocol for enhanced efficiency and improved file-sharing experiences in similar environments is possible.

7 Conclusions

7.1 Summary of Findings

The analysis of the collected data on anonymity levels practically provides several noteworthy findings. The distribution of proxy servers by country reveals that the United States has the highest number of proxy servers, thus followed by South Korea and the United Kingdom. On the other hand, countries like Angola, Iraq, Belarus, Mongolia, and Kyrgyzstan have a significantly lower number of proxy servers, to be specific. Examining the anonymized IP addresses and also their corresponding countries, it is particularly evident that a diverse range of countries are represented, including China, Thailand, South Korea, Hong Kong, Angola, and also Bangladesh, among others. This indicates a global presence and utilization of proxy servers for anonymity purposes. Latency results, which measure the time it takes for data to travel from the source IP address to its destination, thus demonstrate varying values across different IP addresses. The minimum, average, maximum, and also standard deviation of latency are provided for each IP address. These metrics practically allow for an understanding of the speed and also the consistency of network connections associated with each IP.

Telnet latency results practically provide further insights into the network performance of individual IP addresses. Similar to the latency results, telnet latency measures the time it takes for a connection to be established with the IP address. The data specifically reveals a range of latency values, indicating differences in network response times for different IP addresses.

The probability density function (PDF) of average values for IP addresses offers a visual

representation of the distribution of average latencies. The PDF plot specifically shows the likelihood of encountering a specific average latency value, thus allowing for an assessment of the overall latency performance. Additionally, cumulative distribution function (CDF) and alos probability density function (PDF) plots of data provide insights into the overall distribution and probability of latency values. These plots offer a comprehensive overview of the latency characteristics across the dataset.

The traceroute results practically provide information on the network path, and also routers traversed to reach a specific IP address. This data specifically helps identify the network infrastructure involved and even provides insights into potential bottlenecks or points of interest along the network path.

In conclusion, evaluating the efficiency of the BitTorrent protocol in the context of file sharing between a host machine and a virtual machine required analyzing CPU utilization on both the host machine and the virtual machine. It was crucial to consider various factors and metrics to assess the performance, resource utilization, and user experience of the BitTorrent protocol. The findings of this evaluation contribute to a deeper understanding of the BitTorrent protocol's efficiency and provide insights for optimizing its performance in similar environments.

Our evaluation demonstrates that increasing the number of peers in data synchronization using the BitTorrent protocol on uTorrent can potentially reduce latency. By leveraging the distributed nature of the protocol, multiple peers can work in parallel to share and transfer data, resulting in improved synchronization performance. However, it is crucial to find the optimal balance between the number of peers and available resources on the host machine to avoid resource saturation and maintain efficient data synchronization. It is worth noting that the findings of this evaluation are based on the specific test environment and parameters used. Different network conditions, hardware configurations, or BitTorrent clients may yield varying results. Therefore, further testing and analysis are recommended to validate these findings in different scenarios.

In summary, the analysis of anonymity levels practically reveals a global distribution of proxy servers, with varying levels of network performance and also latency across different IP addresses. The findings specifically provide valuable insights into the utilization and also characteristics of anonymizing technologies, shedding light on the diverse landscape of online anonymity.

7.2 Contributions to the Field

The research on anonymity levels and proxy server analysis, particularly conducted in this study, makes significant contributions to the field of network security and privacy. By examining the distribution of proxy servers by country, the study particularly offers valuable insights into the global presence and also utilization of anonymizing technologies. This information can particularly assist policymakers, network administrators, and also security professionals in understanding the geographical patterns of anonymity-seeking behavior and even devising appropriate strategies to address potential security concerns.

The analysis of anonymized IP addresses and their corresponding countries provides further contributions to the field. By practically identifying the countries associated with these IP addresses, researchers gain a better understanding of the diverse range of locations specifically involved in anonymous online activities. This knowledge can particularly aid in the development of targeted interventions to practically mitigate potential risks associated with anonymous communications originating from specific regions.

The investigation of latency results particularly offers another noteworthy contribution. By examining the minimum, average, maximum, and also standard deviation of latency for each IP address, the study provides insights into the performance and consistency of network connections associated with proxy servers. This information is particularly crucial for network administrators and also service providers in optimizing network performance and even addressing potential bottlenecks that may arise due to anonymization technologies.

The telnet latency results further enhance the understanding of network performance. By practically measuring the time it takes to establish a connection with each IP address, researchers can particularly assess the responsiveness and also reliability of proxy servers. This knowledge can particularly aid in identifying potential vulnerabilities or even areas for improvement in the anonymizing infrastructure.

The probability density function (PDF) and also cumulative distribution function (CDF) plots contribute to the field by offering visual representations of the latency data. These plots specifically provide a comprehensive overview of the latency characteristics across the dataset, enabling researchers and also practitioners to gain a deeper understanding of the distribution and, thus the likelihood of encountering specific latency values. This information can particularly inform the design and optimization of network systems, thus ensuring efficient and secure communication channels.

7.3 Recommendation for Future Work

Future research in network security and also privacy should focus on expanding the dataset to provide specifically a more comprehensive view of anonymization technologies worldwide. This will specifically enable researchers to assess anonymity levels across different regions and also identify patterns or trends.

Additionally, investigating the relationship between anonymity and even specific online activities would help understand the motivations and also behaviors of anonymous users, therefore facilitating the development of targeted risk mitigation strategies. Efforts should also particularly be directed towards detecting and countering malicious activities facilitated by anonymous communications. Developing advanced algorithms or even machine learning techniques to practically identify suspicious behavior originating from proxy servers would enhance network security and also differentiate between legitimate users and potential threats.

Exploring novel anonymization techniques and also technologies is crucial to stay ahead of evolving threats. Research should practically evaluate emerging methods for achieving anonymity, considering their impact on network performance, privacy, and also security.

Conducting longitudinal studies to observe particular changes in anonymity levels and also proxy server utilization over time would provide valuable insights into anonymous online behavior. This would significantly help understand the dynamics influenced by technological advancements, regulatory changes, and also shifting user attitudes toward privacy.

The following key recommendations can be made to further enhance the efficiency of the BitTorrent protocol in the context of file sharing between a host machine and a virtual machine. Firstly, future studies should focus on collecting data from various countries. This would provide a more diverse dataset, enabling a comprehensive analysis of network conditions and regional variations. By including data from different geographic locations, researchers can gain valuable insights into the protocol's efficiency in various network environments, leading to a better understanding of its performance and potential optimizations.

Secondly, optimizing data synchronization techniques within the BitTorrent protocol should be a priority for future work. Researchers can explore algorithms that prioritize data transfers, improve resource utilization, and develop more effective mechanisms to handle network congestion. By optimizing the data synchronization process, the overall efficiency and performance of the BitTorrent protocol can be significantly improved. This will contribute to faster and more reliable file sharing between host and virtual machines, enhancing the user experience and enabling more efficient data synchronization in distributed systems.

While the web application has proven effective and valuable, there are several areas for future development and research to enhance its functionality and expand its scope. These include implementing advanced proxy server filtering to provide users with more specific results, integrating real-time proxy server monitoring for up-to-date information, exploring machine learning techniques for improved accuracy, developing a mobile application for

broader accessibility, addressing security and privacy concerns, and establishing collaborations with ISPs to enhance data accuracy. By pursuing these avenues, the web application can evolve into a more comprehensive solution for locating and verifying open proxy servers, benefiting its users.

References

Ali, Farha (2007). "IP spoofing". In: *The Internet Protocol Journal* 10.4, pp. 1–9 (cit. on p. 5).

Assigned numbers (Apr. 1985). RFC 943. DOI: 10.17487/RFC0943. URL: https://www.rfc-editor.org/info/rfc943 (cit. on p. 7).

Center, Pew Research (2021). *Internet/Broadband Fact Sheet*. https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ (cit. on p. 1).

Cisco (2021). *What is VPN?* Website. Available at: https://www.cisco.com/c/en/us/products/security/vpn-endpoint-security-clients/what-is-vpn.html (cit. on p. 11).

Cohen, Bram (2003). *Incentives Build Robustness in BitTorrent*. http://www.bittorrent.org/bittorrentecon.pdf (cit. on p. 37).

Coulouris, G, J Dollimore, and T Kindberg (2005). *Distributed systems: Concepts and design: Pearson education* (cit. on p. 35).

Crocker, D. H. and J. J. Vittal (Dec. 1978). *Specification of Internet Transmission Control Program*. Internet Experiment Note 46. Available at: https://www.rfc-editor.org/ien/ien46.txt. RFC Editor (cit. on p. 5).

Deshpande, Kiran and Madhuri Rao (Jan. 2022). "An Open-Source Framework Unifying Stream and Batch Processing". In: pp. 607–630. ISBN: 978-981-16-6722-0. DOI: 10.1007/978-981-16-6723-7_45 (cit. on p. 29).

Dingledine, Roger, Nick Mathewson, Paul F Syverson, et al. (2004). "Tor: The second-generation onion router." In: *USENIX security symposium*. Vol. 4, pp. 303–320 (cit. on p. 10).

DoD standard Internet Protocol (Jan. 1980). RFC 760. DOI: 10.17487/RFC0760. URL: https://www.rfc-editor.org/info/rfc760 (cit. on pp. 4, 5).

Farrell, Stephen and Hannes Tschofenig (May 2014). *Pervasive Monitoring Is an Attack*. RFC 7258. DOI: 10.17487/RFC7258. URL: https://www.rfc-editor.org/info/rfc7258 (cit. on p. 9).

Fuller, Vince and Tony Li (Aug. 2006). *Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan*. RFC 4632. DOI: 10.17487/RFC4632. URL: https://www.rfc-editor.org/info/rfc4632 (cit. on p. 7).

Fuller, Vince, Tony Li, et al. (Sept. 1993). *Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy*. RFC 1519. DOI: 10.17487/RFC1519. URL: https://www.rfc-editor.org/info/rfc1519 (cit. on p. 6).

Goldberg, I., D. Wagner, and E. Brewer (1997). "Privacy-enhancing technologies for the Internet". In: *Proceedings IEEE COMPCON 97. Digest of Papers*, pp. 103–109. DOI: 10.1109/CMPCON.1997.584680 (cit. on p. 1).

Huang, C. and T. Abdelzaher (2005). "Bounded-latency content distribution feasibility and evaluation". In: *IEEE Transactions on Computers* 54.11, pp. 1422–1437. DOI: 10.1109/TC.2005.175 (cit. on p. 14).

Hubbard, Kim et al. (Nov. 1996). *Internet Registry IP Allocation Guidelines*. RFC 2050. DOI: 10.17487/RFC2050. URL: https://www.rfc-editor.org/info/rfc2050 (cit. on p. 8).

IBM (2023). *Functionality - IPv4 and IPv6 Address Formats*. Accessed on July 2, 2023. IBM. URL: https://www.ibm.com/docs/en/ts3500-tape-library?topic=functionality-ipv4-ipv6-address-formats (cit. on p. 4).

Internet Standard Subnetting Procedure (Aug. 1985). RFC 950. DOI: 10.17487/RFC0950. URL: https://www.rfc-editor.org/info/rfc950 (cit. on p. 6).

IPRoyal (2023). *What Is a Proxy Server and How Does It Work?* Website. Available at: https://iproyal.com/blog/what-is-a-proxy-server-and-how-does-it-work/ (cit. on p. 13).

Kang, Ruogu, Stephanie Brown, and Sara Kiesler (2013). "Why Do People Seek Anonymity on the Internet? Informing Policy and Design". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. Paris, France: Association for Computing Machinery, pp. 2657–2666. ISBN: 9781450318990. DOI: 10.1145/2470654.2481368. URL: https://doi.org/10.1145/2470654.2481368 (cit. on p. 15).

Legout, Arnaud, Guillaume Urvoy-Keller, and Pietro Michiardi (2005). "Understanding bittorrent: An experimental perspective". In: (cit. on p. 31).

Microsoft (2022). *IP Addressing: DHCP and IP Addressing Concepts*. Website. Available at: https://learn.microsoft.com/en-us/previous-versions/windows/it-pro/windows-2000-server/bb742566(v=technet.10) (cit. on p. 12).

Mitchell, Bradley (2023). *What Is a Public IP Address?* Accessed on July 2, 2023. Lifewire. URL: https://www.lifewire.com/what-is-a-public-ip-address-2625974 (cit. on p. 4).

Moskowitz, Robert et al. (Feb. 1996). *Address Allocation for Private Internets*. RFC 1918. DOI: 10.17487/RFC1918. URL: https://www.rfc-editor.org/info/rfc1918 (cit. on p. 9).

Piatek, Martin et al. (2007). *The BitTorrent Protocol Specification v1.0*. http://bittorrent.org/beps/bep_0003.html (cit. on p. 37).

Postel, J. (Sept. 1981). *Internet Protocol*. Tech. rep. 791. Available at: https://tools.ietf.org/html/rfc791. RFC Editor (cit. on p. 5).

Qiu, Dongyu and R. Srikant (2004). "Modeling and Performance Analysis of BitTorrent-like Peer-to-Peer Networks". In: *Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*. SIGCOMM '04. Portland, Oregon, USA: Association for Computing Machinery, pp. 367–378. ISBN: 1581138628. DOI: 10.1145/1015467.1015508. URL: https://doi.org/10.1145/1015467.1015508 (cit. on p. 31).

Rekhter, Yakov and Tony Li (Sept. 1993). *An Architecture for IP Address Allocation with CIDR*. RFC 1518. DOI: 10.17487/RFC1518. URL: https://www.rfc-editor.org/info/rfc1518 (cit. on p. 6).

Sharma, Parul, Anuja Bhakuni, and Rishabh Kaushal (2013). "Performance analysis of BitTorrent protocol". In: *2013 National Conference on Communications (NCC)*, pp. 1–5. DOI: 10.1109/NCC.2013.6488040 (cit. on p. 32).

Virginia Tech (2016). *Proxies*. Website. Available at: https://courses.cs.vt.edu/~cs4244/spring.09/documents/Proxies.pdf (cit. on p. 14).

Wikipedia (2004). *Plagiarism — Wikipedia, The Free Encyclopedia*. [Online; accessed 22-July-2004]. URL: %5Curl%7Bhttps://en.wikipedia.org/wiki/IP_address%7D (cit. on p. 4).