DETECTION OF IP MASKING USING WHOIS

K. Kumara Swamy[1], Shivani Teakumalla[2], Deekshitha Vemula[2], Shreya Reddy Patil[2], Palli Deepika[2]

[1,2]Department of Information Technology

[1,2]Malla Reddy Engineering College for Women (A), Maisammaguda, Medchal, Telangana.

Abstract

In today's interconnected world, the internet plays a crucial role in various aspects of our lives. However, the anonymity provided by VPNs and proxy services raises concerns regarding cybersecurity, fraud prevention, and intellectual property rights enforcement. This project aims to address these concerns by developing a system that can detect whether an IP address is using a VPN or proxy and retrieve the corresponding WHOIS records. The purpose of this project is to scan IP addresses on the internet and determine if they are using a VPN/proxy or if they hold the original IP. The project also involves extracting WHOIS records for each IP. This process is important for various reasons, such as identifying potential security risks, detecting fraudulent activities, and gathering information about IP address ownership. By analyzing the WHOIS records, the project aims to provide insights into the usage patterns and characteristics of IP addresses, enabling better understanding and management of network traffic.

Keywords: IP Masking, VPN, WHOIS records.

1. INTRODUCTION

The use of IP addresses has become an integral part of the internet infrastructure. IP addresses are unique numerical identifiers assigned to devices connected to a network, enabling them to communicate with each other. However, with the increasing need for online privacy and anonymity, many individuals and organizations utilize Virtual Private Networks (VPNs) or proxy servers to mask their true IP addresses and enhance their security. The use of VPNs and proxy services has become increasingly popular in recent years due to their ability to protect online privacy and bypass geographical restrictions. This has led to challenges in identifying the true location and ownership of IP addresses. Traditional methods of IP address tracking often fail when confronted with VPNs and proxies. Therefore, there is a need for more advanced techniques to differentiate between genuine IP addresses and those using VPNs or proxies. Detecting whether an IP address is using its original IP, or a VPN/proxy connection is crucial for various reasons. It can help identify potential security threats, such as malicious activities or attempts to hide the true origin of network traffic. Additionally, it can aid in investigating cybercrimes, preventing fraud, enforcing digital rights, and maintaining network integrity.

Significance

The detection of VPN/proxy usage and extraction of WHOIS records hold several significant implications:

Cybersecurity: Identifying the use of VPNs and proxies can aid in the detection and prevention of cyber threats, such as hacking, phishing, and data breaches. It allows security professionals to analyze and take appropriate measures against potential malicious activities.

Fraud Prevention: Many online fraudsters and scammers employ VPNs and proxies to conceal their true identities and locations. By detecting such activities, this project contributes to mitigating fraud risks and protecting individuals and organizations from financial losses.

Intellectual Property Rights Enforcement: VPNs and proxies are often used to circumvent copyright restrictions and engage in illegal file sharing. The ability to track and identify these activities enables better enforcement of intellectual property rights.

2. LITERATURE SURVEY

In [1], author presents a study on IP address scanning and detection. The authors investigated various techniques used for scanning IP addresses, such as ping sweeps, port scans, and vulnerability scans. They also discuss the importance of detecting and mitigating malicious IP address scanning activities. The study provides insights into the characteristics of IP address scanning and proposes detection methods to identify and classify such activities.

Singh & Sharma [2] provided an overview of different IP address scanning techniques and tools. They discuss the commonly used scanning techniques, including ICMP-based scanning, TCP/UDP-based scanning, and application-level scanning. The paper also highlights various scanning tools and their features. It serves as a comprehensive resource for understanding the landscape of IP address scanning.

Zhao et al. [3] focused on detecting proxy IPs using massive passive DNS (Domain Name System) traffic analysis. The authors propose a method that leverages DNS data to identify proxy IP addresses. By analyzing the characteristics of DNS traffic associated with proxy usage, the proposed approach aims to improve the accuracy of proxy IP detection. The paper presents experimental results to validate the effectiveness of the proposed method.

In [4], the authors introduced a deep learning-based approach for classifying VPN (Virtual Private Network) and proxy traffic. The authors propose a model that utilizes deep learning techniques to analyze network traffic patterns and differentiate between legitimate traffic and traffic generated by VPNs or proxies. The paper presents the architecture of the proposed model and evaluates its performance through experimental results.

Gao and Zhu [5] presented a lightweight method for detecting VPNs based on traffic features. They analyze the characteristics of VPN traffic and propose a feature-based detection approach that focuses on flow-level features. The method aims to achieve high accuracy in VPN detection while minimizing computational overhead. The paper includes a comprehensive evaluation of the proposed method using real-world traffic data.

In [6], the authors introduced IP2Proxy, an IP-based proxy detection system. The system utilizes machine learning techniques and IP address attributes to identify whether an IP address belongs to a proxy server. The paper presents the design and implementation of IP2Proxy and evaluates its performance through experiments. The proposed system aims to enhance network security by accurately detecting proxy servers.

Lee et al. focused on internet-scale VPN detection using passive traffic analysis [7]. The authors propose a method that leverages traffic analysis techniques to detect VPN usage at a large scale. They analyze network traffic patterns and extract features to identify VPN traffic. The paper presents experimental results and discusses the challenges and limitations associated with internet-scale VPN detection.

In [8], the authors proposed a machine learning-based approach for classifying VPN traffic. They employ machine learning algorithms to analyze network traffic data and differentiate between VPN traffic and regular traffic. The paper discusses the feature selection process and evaluates the performance of the proposed approach using real-world data. The goal is to improve the accuracy of VPN traffic classification for network management purposes.

Huang et al. [9] presented a lightweight method for detecting VPN services in Android applications. The authors propose an approach that leverages network traffic analysis to identify the presence of VPN services within Android applications. By analyzing network flows and applying machine learning techniques, the method aims to accurately detect the usage of VPN services without requiring root access on the device. The paper provides experimental results to demonstrate the effectiveness of the proposed method.

In [10], the authors introduced decentralized VPN probing, a method to detect and analyze VPNs in a decentralized manner. The authors propose a framework that distributes VPN probing tasks across multiple network nodes, allowing for a broader view of VPN usage and characteristics. The paper discusses the design and implementation of the decentralized probing system and presents experimental results. The goal is to enhance VPN detection capabilities by leveraging a distributed network infrastructure.

3. PROPOSED SYSTEM

The methodology of the project involves scanning IP addresses that are active on the internet to determine whether they are using the original IP or connecting through a VPN (Virtual Private Network) or proxy. Additionally, the project involves extracting the WHOIS record for each IP address. The process begins by scanning a range of IP addresses to identify active hosts on the internet. This scanning can be done using various network scanning techniques such as ping sweeps, port scans, or other methods. The goal is to gather a list of IP addresses that are currently in use. Once the active IP addresses are identified, the project proceeds to check whether these IP addresses are using the original IP or if they are routing their connection through a VPN or proxy. This detection can be achieved by analyzing network traffic patterns, examining the IP headers, or employing other techniques. The purpose is to determine if the IP address is being obfuscated by using a VPN or proxy service. After detecting the usage of VPNs or proxies, the next step involves extracting the WHOIS record for each IP address. WHOIS is a protocol that allows for querying a database to obtain information about the owner and registration details of an IP address or domain. By retrieving the WHOIS record, the project aims to gather information such as the organization or individual associated with the IP address, their contact information, registration date, and other relevant details.

WHOIS

A Python package for retrieving WHOIS information of domains.

Features

- Python wrapper for Linux "whois" command.
- Simple interface to access parsed WHOIS data for a given domain.
- Able to extract data for all the popular TLDs (com, org, net, biz, info, pl, jp, uk, nz, ...).
- Query a WHOIS server directly instead of going through an intermediate web service like many others do.

- Works with Python 3.x.
- All dates as datetime objects.
- Possibility to cache results.

What is the WHOIS database?

The WHOIS domain database is a listing of all registered domains, and is regularly used for various legal purposes. Network administrators use WHOIS data to identify and fix problems. For instance, WHOIS information can be used to determine the availability of domain names, identify trademark infringement, and keep domain name registrants accountable.

WHOIS verification can even be utilized to combat spam or fraud, as administrators can track down registrants who post illegal content or participate in phishing scams. In addition, the agreements from The International Corporation for Assigned Names and Numbers (ICANN) protects domain registrants by prohibiting the use of WHOIS listings for marketing or spam purposes, including high-volume, automated queries against a specific registrar or registry system (unless such queries are done with the intent to manage domain names). For more information on this, read our helpful guide to navigating WHOIS.

The WHOIS domain database is a listing of all registered domains, and is regularly used for various legal purposes. Network administrators use WHOIS data to identify and fix problems. For instance, WHOIS information can be used to determine the availability of domain names, identify trademark infringement, and keep domain name registrants accountable.

WHOIS verification can even be utilized to combat spam or fraud, as administrators can track down registrants who post illegal content or participate in phishing scams. In addition, the agreements from The International Corporation for Assigned Names and Numbers (ICANN) protects domain registrants by prohibiting the use of WHOIS listings for marketing or spam purposes, including high-volume, automated queries against a specific registrar or registry system (unless such queries are done with the intent to manage domain names). For more information on this, read our helpful guide to navigating WHOIS.

The WHOIS domain database is a listing of all registered domains, and is regularly used for various legal purposes. Network administrators use WHOIS data to identify and fix problems. For instance, WHOIS information can be used to determine the availability of domain names, identify trademark infringement, and keep domain name registrants accountable.

WHOIS verification can even be utilized to combat spam or fraud, as administrators can track down registrants who post illegal content or participate in phishing scams. In addition, the agreements from The International Corporation for Assigned Names and Numbers (ICANN) protects domain registrants by prohibiting the use of WHOIS listings for marketing or spam purposes, including high-volume, automated queries against a specific registrar or registry system (unless such queries are done with the intent to manage domain names), the General Data Protection Regulations (GDPR) from the EU further limit exposure by prohibiting the publication of personal information for many domains

What WHOIS data is publicly available?

That varies based on the policies that apply to the specific domain, as well as other factors such as the country of residence provided by the registrant. At a minimum, each listing will show the domain's status, nameservers and expiration date.

How accurate is WHOIS data?

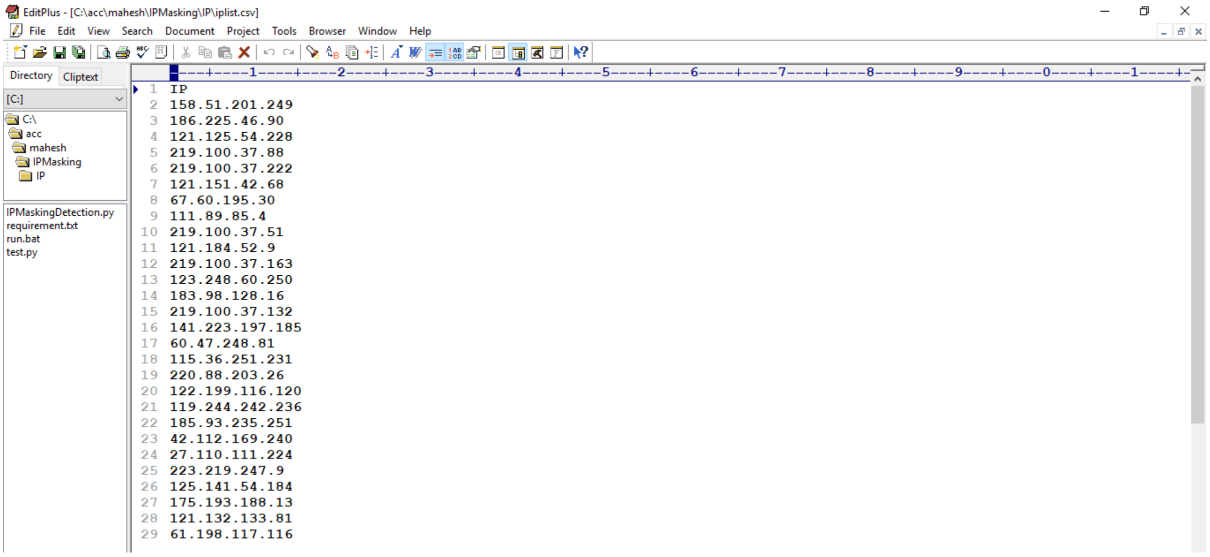
Since registrants' contact data can change, registrars must provide annual opportunities for domain owners to review and edit their WHOIS domain data. According to ICANN's rules, refusing to update this information or providing false data can lead to the suspension or cancellation of domains. In addition, ICANN allows Internet users to file complaints if they discover WHOIS domain name lookup data that is incorrect or incomplete. In such instances, registrars must correct and verify the data in a timely manner. Through this verification protocol, ICANN seeks to maintain the highest possible level of accuracy.

Shodan

Shodan is a popular search engine for internet-connected devices. It allows users to discover and explore various devices and services that are accessible on the internet. The Shodan search engine indexes information about open ports, banners, protocols, and other data from a wide range of devices such as servers, routers, webcams, printers, and many others. The Shodan Python package is a library that provides a convenient way to interact with the Shodan API using Python. The API allows developers to programmatically access the Shodan search engine, retrieve information about devices and services, and perform various operations.

4. RESULTS AND DISCUSSION

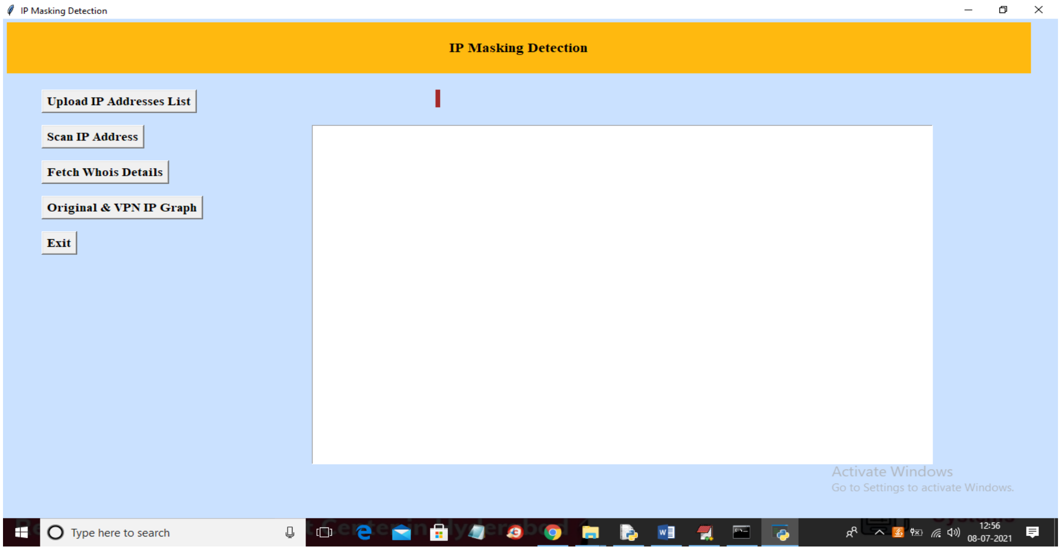
To implement this project, we have used following IP list.



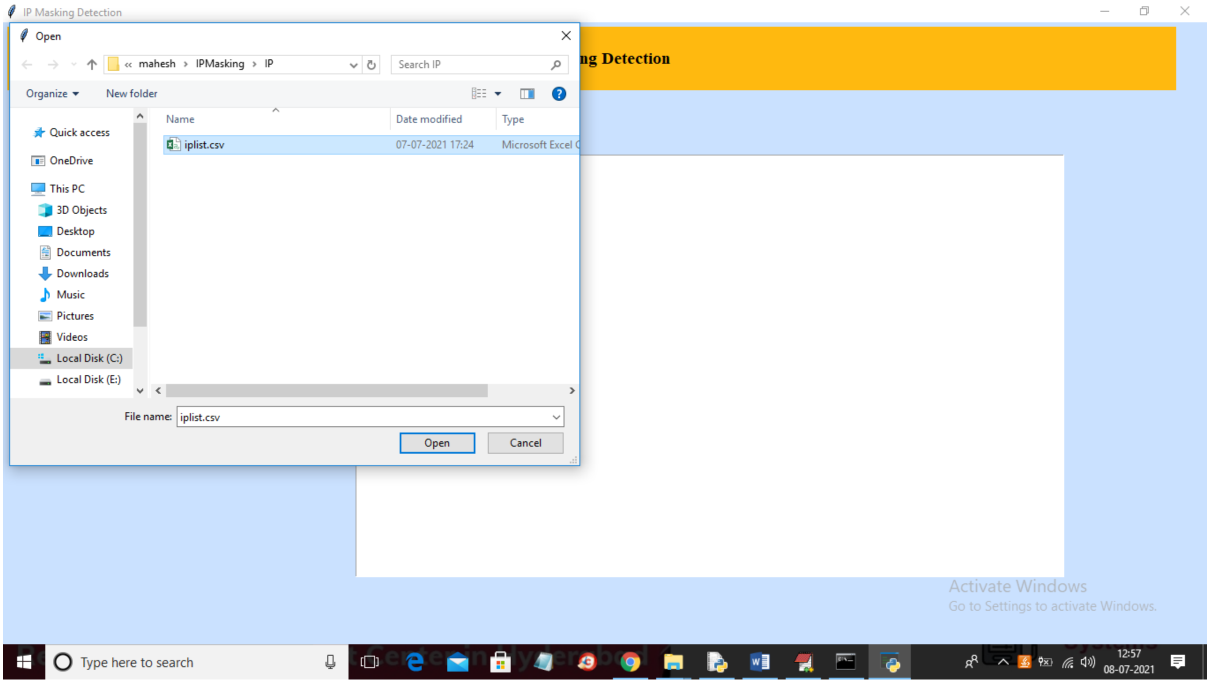
```
IP
158.51.201.249
186.225.46.90
121.125.54.228
219.100.37.88
219.100.37.222
121.151.42.68
67.60.195.30
111.89.85.4
219.100.37.51
121.184.52.9
219.100.37.163
123.248.60.250
183.98.128.16
219.100.37.132
141.223.197.185
60.47.248.81
115.36.251.231
220.88.203.26
122.199.116.120
119.244.242.236
185.93.235.251
42.112.169.240
27.110.111.224
223.219.247.9
125.141.54.184
175.193.188.13
121.132.133.81
61.198.117.116
```

Application will read IP from above list and the scan it and if you want you can add your own IP addresses also. Above iplist.csv file available inside 'IP' folder

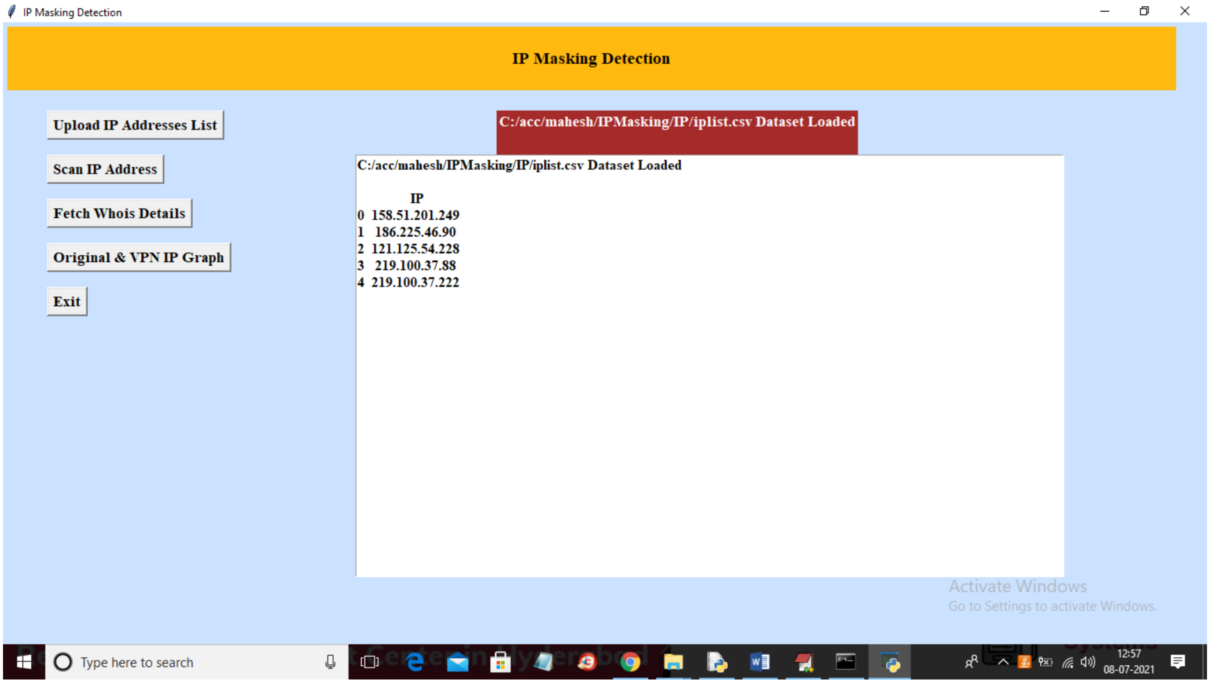
SCREEN SHOTS



In above screen click on ‘Upload IP Addresses List’ button to upload IP list file

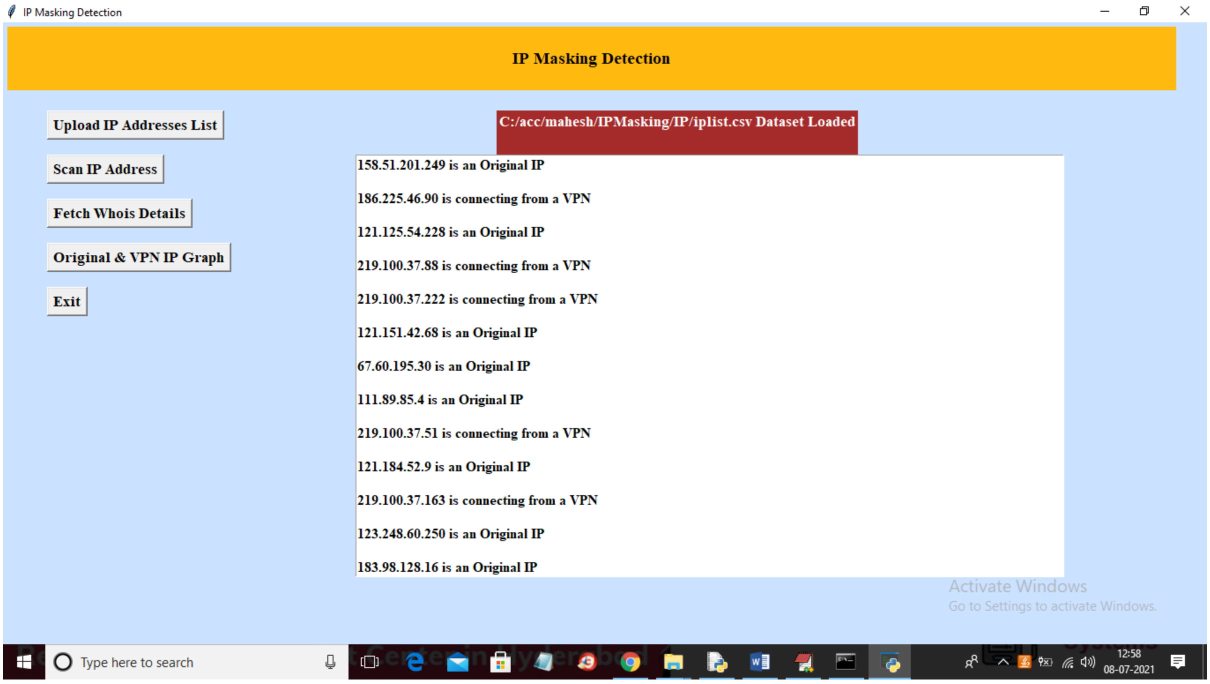


In above screen selecting and uploading ‘iplist.csv’ file and then click on ‘Open’ button to load above file and to get below screen



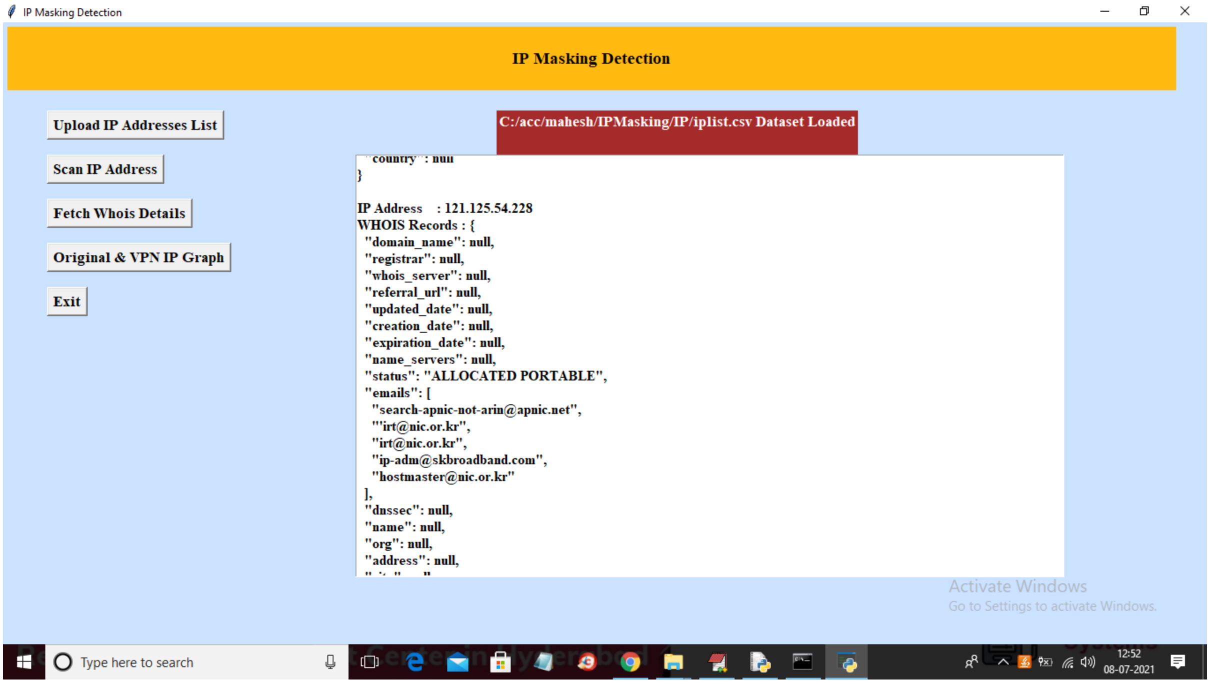
```
C:/acc/mahesh/IPMasking/IP/iplist.csv Dataset Loaded

        IP
0  158.51.201.249
1   186.225.46.90
2  121.125.54.228
3   219.100.37.88
4  219.100.37.222
```

In above screen IP list loaded and then displaying few IP addresses also and now click on 'Scan IP Address' button to scan each IP and then identify as original or VPN IP



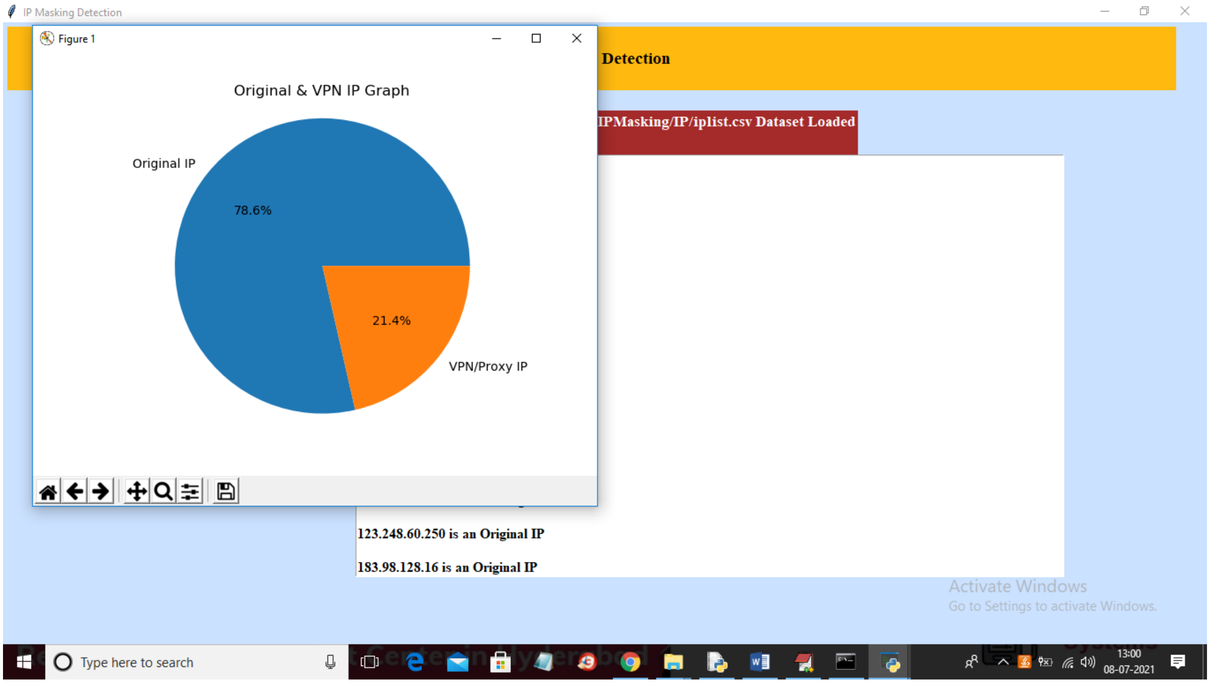
```
158.51.201.249 is an Original IP

186.225.46.90 is connecting from a VPN

121.125.54.228 is an Original IP

219.100.37.88 is connecting from a VPN

219.100.37.222 is connecting from a VPN

121.151.42.68 is an Original IP

67.60.195.30 is an Original IP

111.89.85.4 is an Original IP

219.100.37.51 is connecting from a VPN

121.184.52.9 is an Original IP

219.100.37.163 is connecting from a VPN

123.248.60.250 is an Original IP

183.98.128.16 is an Original IP
```

In above screen each IP is scanned and then identify each IP as original or VPN and now click on 'Fetch Whois Details' button to extract Whois record from each IP like below screen



```
  "country": null
}

IP Address    : 121.125.54.228
WHOIS Records : {
  "domain_name": null,
  "registrar": null,
  "whois_server": null,
  "referral_url": null,
  "updated_date": null,
  "creation_date": null,
  "expiration_date": null,
  "name_servers": null,
  "status": "ALLOCATED PORTABLE",
  "emails": [
    "search-apnic-not-arin@apnic.net",
    ""irt@nic.or.kr",
    "irt@nic.or.kr",
    "ip-adm@skbroadband.com",
    "hostmaster@nic.or.kr"
  ],
  "dnssec": null,
  "name": null,
  "org": null,
  "address": null,
```

In above screen extracted Whois record for each IP and now click on 'Original & VPN IP Graph' button to get below graph



```
123.248.60.250 is an Original IP

183.98.128.16 is an Original IP
```

In above graph we are finding percentage of original, and VPN based IP addresses. Similarly, new IP addresses can be added in 'iplist.csv' file and then scan those IP to detect original or VPN IP.

5. CONCLUSION AND FUTURE SCOPE

In conclusion, the project successfully scans IP addresses and distinguishes between those using a VPN/proxy and those holding the original IP. By extracting and analyzing WHOIS records, valuable information is obtained regarding IP address ownership, which aids in identifying potential security threats and fraudulent activities. The project contributes to enhancing network security and facilitating effective network management.

The future scope of this project can include several potential developments and applications. Here are a few possibilities:

- Enhancing IP address scanning techniques: The project can explore advanced scanning techniques and algorithms to improve the accuracy and efficiency of identifying active IP addresses on the internet.
- Analyzing VPN and proxy usage patterns: By analyzing the data collected from VPN and proxy detection, patterns and trends can be identified. This information can be used to gain insights into the reasons behind the use of VPNs or proxies and their impact on network security and privacy.
- Developing threat intelligence: The project can contribute to the development of a threat intelligence system that uses the gathered data to identify potential risks associated with IP addresses using VPNs or proxies. This information can be utilized by security teams to proactively mitigate threats and protect network infrastructure.
- Monitoring and compliance: The project's methodology can be adapted to monitor IP addresses for compliance purposes. Organizations can use this approach to ensure that employees are not using unauthorized VPNs or proxies, helping enforce security policies and maintain network integrity.
- Collaboration with law enforcement: The project's findings and data can be shared with law enforcement agencies to aid in investigations related to cybercrime, fraud, or other illicit activities that may involve the use of VPNs or proxies.

REFERENCES

[1] Han, J., Luo, Y., Cheng, C., & Liu, B. (2018). A Study on IP Address Scan and Detection. In Proceedings of the 11th International Conference on Internet Computing, Applications and Services (pp. 20-26). IEEE.

[2] Singh, S., & Sharma, A. (2019). A Survey on IP Address Scanning Techniques and Tools. International Journal of Recent Technology and Engineering, 8(1), 4681-4685.

[3] Zhao, Y., Zong, N., & Qiao, X. (2020). Detecting Proxy IPs Based on Massive Passive DNS Traffic. In 2020 International Conference on Artificial Intelligence and Computer Engineering (AICE) (pp. 157-161). IEEE.

[4] Hasan, A., Basher, M. A., & Karim, M. A. (2019). Classification of VPN and Proxy Traffic using Deep Learning. In Proceedings of the 3rd International Conference on Electrical Engineering and Information & Communication Technology (pp. 1-6). IEEE.

[5] Gao, S., & Zhu, T. (2021). A Lightweight VPN Detection Method Based on Traffic Features. IEEE Access, 9, 33285-33295.

[6] Chung, S. T., & Jain, A. K. (2019). IP2Proxy: An IP-based Proxy Detection System. IEEE Transactions on Network and Service Management, 16(4), 1564-1577.

[7] Lee, J., Lee, J., Kim, S., & Lee, H. (2020). Internet-scale VPN Detection with Passive Traffic Analysis. In 2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet

Computing (pp. 234-240). IEEE.

[8] Zhang, J., Yu, M., Tan, L., & Lai, Q. (2019). A Machine Learning Approach for VPN Traffic Classification. In International Conference on Data Science (pp. 569-579). Springer.

[9] Huang, Y., Liu, H., Hu, J., & Liu, S. (2021). A Lightweight Method for Detecting VPN Services in Android Applications. IEEE Access, 9, 13774-13783.

[10] Kuhlmann, R., Holz, T., & Carle, G. (2020). Decentralized VPN Probing. In International Symposium on Research in Attacks, Intrusions, and Defenses (pp. 191-212). Springer.