CP-I Project Report on

INFOGREPPER

at

U. V. Patel College of Engineering

Internal Guide:
Dr. Paresh M. Solanki

Prepared By:
Mr. Parth Sharma (20012011159)
Mr. Krish Patel (20012011103)
Mr. Faizal Kadiwal (20012011049)

B. Tech Semester-V
(Computer Engineering)
Nov-Dec, 2022

Submitted to,
Department of Computer Engineering
U.V. Patel College of Engineering
Ganpat University, Kherva - 384 012

U.V. PATEL COLLEGE OF ENGINEERING

Ganpat University | U.V. Patel College of Engineering

॥ विद्यया समाजोत्कर्षः ॥

25 Years excellence in innovative technical education in shaping engineers

CERTIFICATE

TO WHOM SO EVER IT MAY CONCERN

This is to certify that Mr. PARTH SHARMA student of **B.Tech. Semester-V (Computer Engineering)** has completed his/her full semester on site project work titled "**InfoGrepper**" satisfactorily in partial fulfillment of the requirement of Bachelor of Technology degree of Computer Engineering of Ganpat University, Kherva, Mehsana in the year 2022-2023.

Dr. Paresh M. Solanki

College Project Guide and Head, Computer Engineering

U.V. PATEL COLLEGE OF ENGINEERING

Ganpat University | U.V. Patel College of Engineering

॥ विद्यया समाजोत्कर्षः ॥

25 Years excellence in innovative technical education in shaping engineers

CERTIFICATE

TO WHOM SO EVER IT MAY CONCERN

This is to certify that Mr. KRISH PATEL student of **B.Tech. Semester-V (Computer Engineering)** has completed his/her full semester on site project work titled "**InfoGrepper**" satisfactorily in partial fulfillment of the requirement of Bachelor of Technology degree of Computer Engineering of Ganpat University, Kherva, Mehsana in the year 2022-2023.

Dr. Paresh M. Solanki

College Project Guide and Head, Computer Engineering

U.V. PATEL COLLEGE OF ENGINEERING

Ganpat University
॥ विद्यया समाजोत्कर्षः ॥
U.V. Patel College of Engineering
CELEBRATING 25 YEARS OF EXCELLENCE 1997 - 2022
25 Years excellence in innovative technical education in shaping engineers

CERTIFICATE

TO WHOM SO EVER IT MAY CONCERN

This is to certify that Mr. FAIZAL KADIWAL student of **B.Tech. Semester-V (Computer Engineering)** has completed his/her full semester on site project work titled "**InfoGrepper**" satisfactorily in partial fulfillment of the requirement of Bachelor of Technology degree of Computer Engineering of Ganpat University, Kherva, Mehsana in the year 2022-2023.

Dr. Paresh M. Solanki

College Project Guide and Head, Computer Engineering

ACKNOWLEDGEMENT

We express our sincere gratitude towards internal guide **Dr. Paresh Solanki** for his constant help, encouragement, suggestions, and inspiration throughout the project work. Without his invaluable advice, suggestions, and assistance it would not have been possible for me to complete this project work.

Thank you, Sir

Abstract

We believe that alongside Data Science, Hacking is also the next big thing of the 21st century. Reconnaissance is the first stage of ethical hacking, where you collect data about the target system. The goal of reconnaissance is to identify as many potential attack vectors as possible.

Our noble purpose is to give users a toolkit that will act as a gateway for them to enter this hacking world hassle-free irrespective of any fear regarding their internet freedom, tools that are completely ethical to use and to acknowledge users that hacking, when used appropriately, can make their digital lives better and even add some positive value in their lives.

The sole purpose of this project is to provide the user with a toolkit containing 8 carefully curated tools to be used by users to get information in a completely safe way. The user can register and log in using his credentials. They can use any of the tools to get the desired output also they can view the outputs of tools that they have used previously.

Table of Contents

Chapter 1 – INTRODUCTION

1.1 Project Overview

We as a team believe that alongside Data Science and Machine Learning, Hacking is also the next big thing of the 21st century. We pretty much had a good idea about how an individual shows curiosity when hearing the term hacking or anything revolving around hacking, take it as phishing, scamming, reverse engineering or any fancy word which may sound fancy but deep down takes a lot to achieve it. Our noble purpose is to give users a toolkit which will act like a gateway for them to enter this hacking world hassle free irrespective of any fear regarding their internet freedom, tools which are completely ethical to use and to acknowledge users that hacking when used appropriately can make their digital lives better and even add some positive value in their lives. I personally had been into the digital world of Hacking for around 3 years now, so I pretty much know the basics and how any end user having little or no knowledge regarding it should be introduced to it in an interactive and simple way so that user may not get a sense of fear afterwards whenever they read about hacking portrayed in a negative light. The sole purpose of this application is to give users a toolkit to interact in their dayto-day life with complete security and ethics followed.

1.2 Problem Statement

There are numerous tools regarding information gathering in cyberspace like Nmap, Masscan, Wireshark, Sherlock but all this tools are for different purpose and usage like one is for scanning open ports and services while other one is use in performing network pentesting but as a user in cybersecurity field one has to go through all this tools to get good amount of information regarding it's target. So inshort there's no collective tool for information gathering in the domain.

1.3 Project

A Django Application comprises of 8 different set of information gathering tools which fulfills the need of three different types of users – Penetration Testers, SOC Analyst, OSINT Researchers. Our toolkit fulfills the user requirements by collecting information like website and server crawling, performing recon operations to fetch domain names associated with that particular IP Address and to performing searches on multiple torrent sites like torlock.com and thepiratebay10.org for supplied keywords.

1.4 Project Scope

Project objectives : Create a working model of Django project with 8 tools which are listed below in literature survey.

Resources :

- Core team (three people), 15 hours of work a week for 24 weeks
- Project Guide (one person), 5 hours of work a week for 24 weeks **Deliverables :**
- Complete creation and working of all tools in late April 2023
- Entire project deployed by June 2023

1.5 Literature Survey

We had titled our project/product as INFOGREPPER which literally translates to Information Grepper, with the last word paying a homage to my favorite terminal command utility grep. Our toolkit has 8

different tools to cater all kinds of cybersecurity audiences. We will now try to elaborate each tool one by one for better understanding.

I. ROBOTS TESTER

A text file having a default name robots.txt is created to tell search engine web crawlers to list down which URL websites can easily access your site. This is used mainly to avoid request overloading your site. With this Python script, users can easily filter out all URLs present in robots.txt file, and can know whether they can access them or not.

II. MASSCAN TO NMAP

A tool consisting of three process cycle -

- Run MASSCAN to find the TCP & UDP ports that were open on a target.
- Copy and paste those open ports, and ask NMAP to run those ports.
- Find the service versions of those ports and run the enumeration scripts.

An efficient Python script was designed to perform all these processes into one shot with just giving a single IP Address as a parameter. This script lets users print out the NMAP results and also saves a copy in your local directory as well.

III. SHODAN RECON

This tool takes a set of IP addresses and performs recon operation to fetch domain names associated with that particular IP Address. It is designed specifically when there is no Reverse DNS record set. This tool will become handy whenever the IP address will not directly resolve to the web app and may require a domain name.

IV. DORKSCAN

Dork is a keyword which can enable users to gain access to information that corporations did not intend to make publicly available like user records and confidential data that is hosted on their internal server. Dorkscan will take a single dork or a list of dorks as its arguments, after that the script will find the results we want according to our applied dorks and save them to a text file for further processing. We had designed a Selenium powered Python script to automate searching the web for vulnerable applications.

V. PYCAT

An efficient NETCAT replacement tool purely written in Python that automatically scans for hosts that are up on the local network. Simply run the script without any arguments to automatically start scanning the network hosts.

VI. TORRENT METASEARCHER

This script searches on multiple Torrent sites like torlock.com and thepiratebay10.org for supplied keywords and returns the results grouped by categories in JSON Format file.

VII. IP GEO

A Python Script for fast, accurate and reliable website details. This script is designed in such a way that the geolocation of a server can be searched on the basis of the provided Site/Domain Name of any website or server which is hosted on the web.

VIII. WEB URL SCRAPER

As the name implies this script accesses the website and returns all the links present in that website. The main focus of this script is to retrieve all the webpage links from a given URL of any website.

1.6 Overview

On the 1st day here in college, we all are asked to tell us about our hobbies. The moment I uttered the word Hacking, I had seen a spark of excitement to learn about it in some of my colleagues. Actually thanks to the Internet, the way hacking is portrayed in movies in general is itself cool and worth eyecatching for any individual to get his/her hands on it. Hacking according to me is to exploit a system for personal advantages.
Now the trick part comes, the advantage may vary from hacker to hacker, the downside resembles that it is always not beneficial to the targeted person. Not many individuals who wish to learn about hacking proceeds, because of having a constant fear of getting hacked while doing so. Hacking, if done ethically, is the strongest digital weapon available on the internet. The recent example of Uber showed us how hacking can be that derogatory if misused. Several OS have several toolkits for users to exploit for their advantages, some OS itself are the toolkits designed specifically for hacking. We here as a team and proud flagbearer of performing hacking with complete ethics present you a toolkit containing 8 carefully curated tools to be used by users to get information in a complete safe way.

1.7 Tools and Technology

1. DJANGO FRAMEWORK

A high level python web framework that will enable us to create rapid development of secure, maintainable and scalable progressive web applications.

2. PYTHON LANGUAGE

An advantageous programming language for cybersecurity due to its great and reliable performance in many cybersecurity functions, including malware analysis, scanning, and reconnaissance.

It is also very user-friendly in nature and has elegant simplicity, making it the perfect language choice for our web application.

3. BASH LANGUAGE

The default shell language in most server computing environments in linux in addition to having abilities for system administration and automation of tasks, including security.

It also holds great advantage over integration with python making a perfect choice for our software.

Chapter 2. FEASIBILITY STUDY

2.1 Study of the Current System

SIGIT – Simple Information Gathering Toolkit

SIGIT or Simple Information Gathering Toolkit is an automated tool used for Information Gathering and OSINT analysis. This tool collects various types of information like Username Reconnaissance, IP Location finder, Reverse IP, Mail Finder through Name, and many more. All this information can be used to understand the target domain or individual more clearly and prepare the methodology according to that information. SIGIT tool is available on the GitHub platform, it's free and open-source to use.

Installation of SIGIT Tool on Kali Linux OS

Step 1: Use the following command to install the tool in your Kali Linux operating system. git

clone https://github.com/termuxhackers-id/SIGIT.git

Step 2: Now use the following command to move into the directory of the tool. You have to move in the directory in order to run the tool.

cd SIGIT

Step 3: Change the permissions of sigit.py by using the following command. sudo

chmod 777 sigit.py

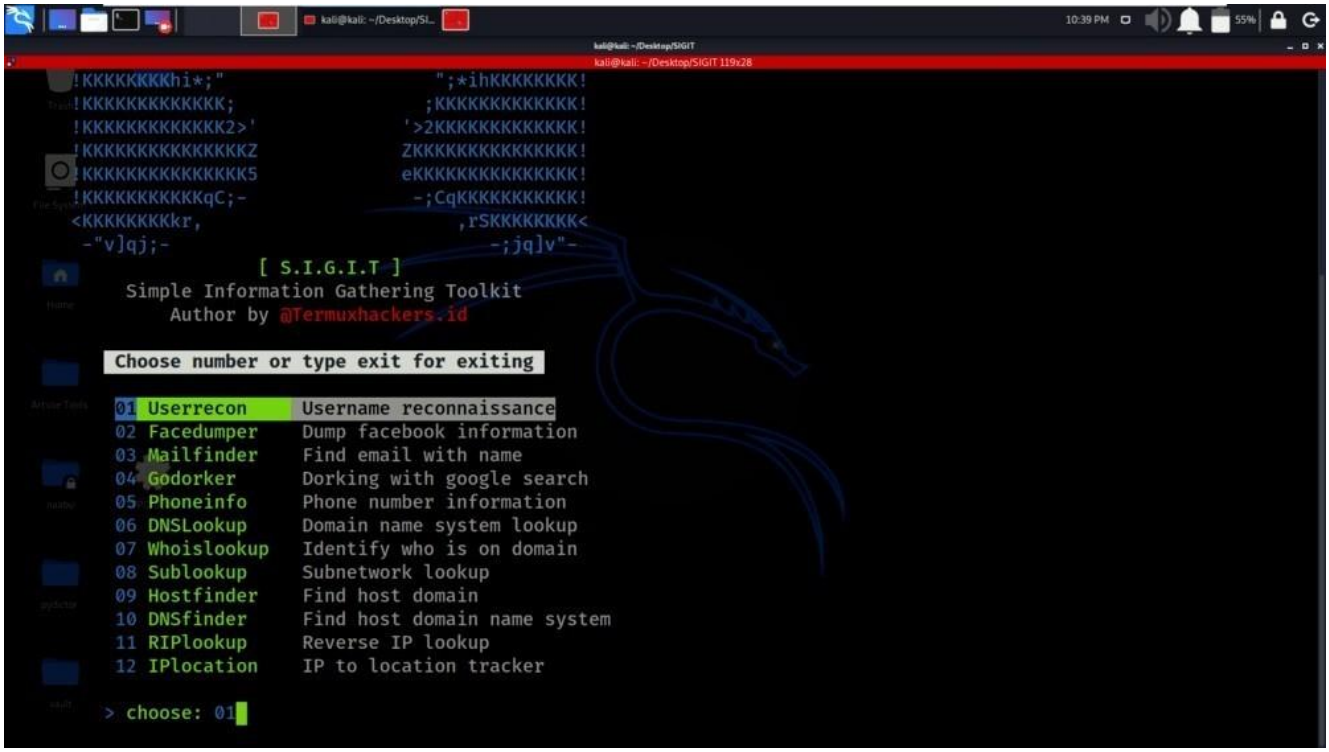
Step 4: Now use the following command to run the tool. python3

sigit.py

Working with SIGIT Tool on Kali Linux OS

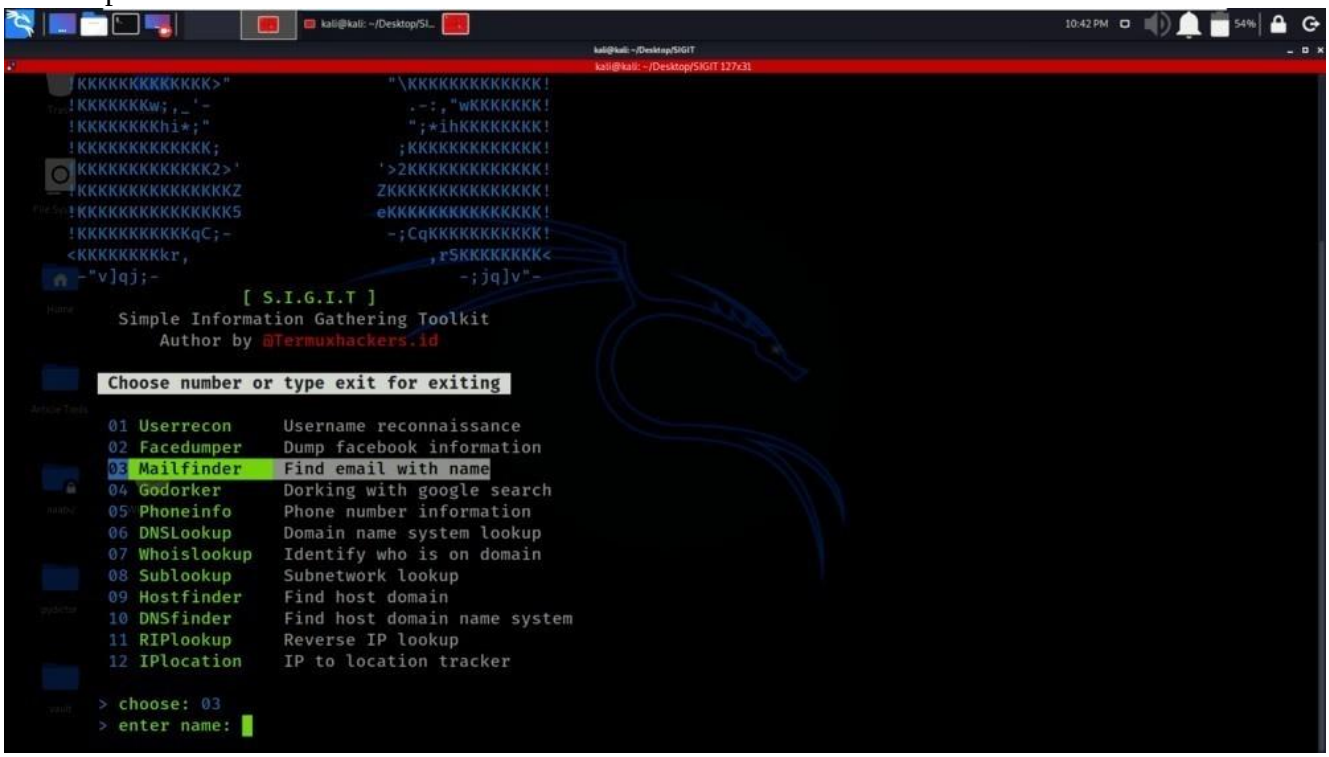
Example 1: Username reconnaissance

Select option 01



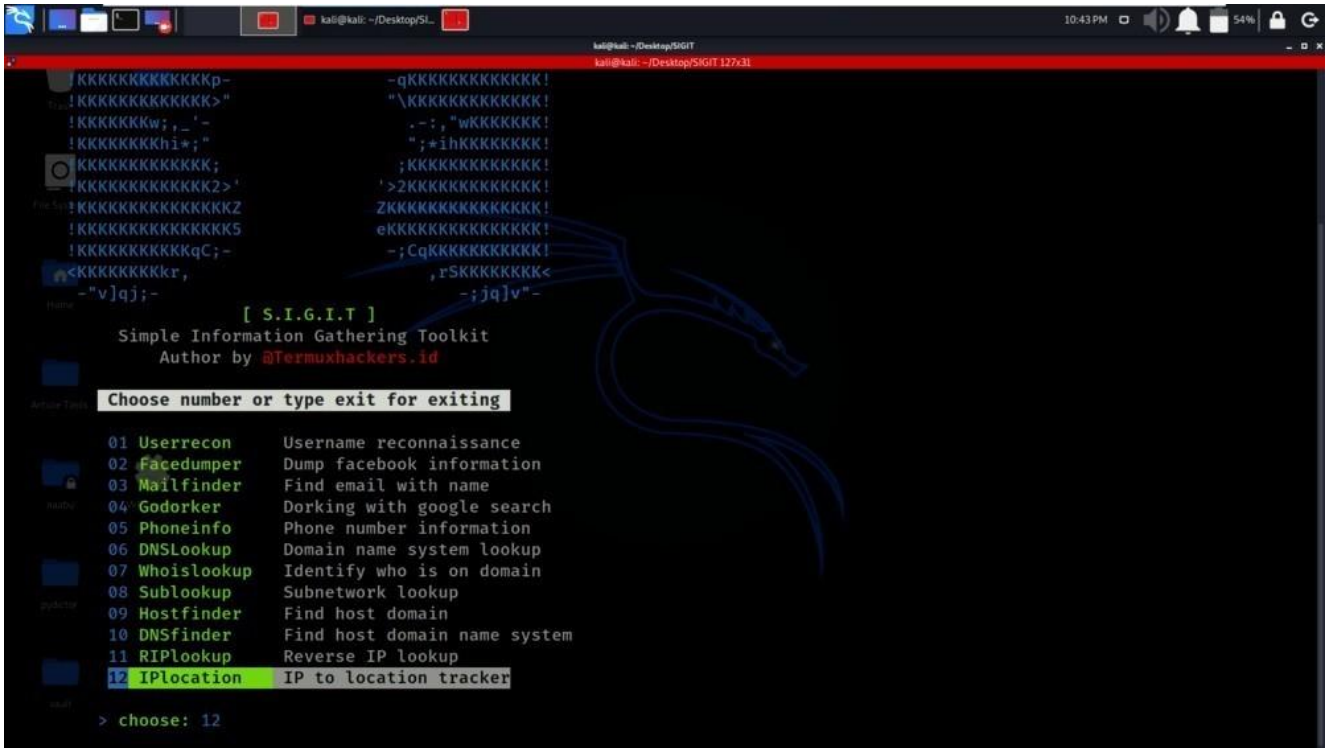
Example 2: Find the email with the name

Select option 03

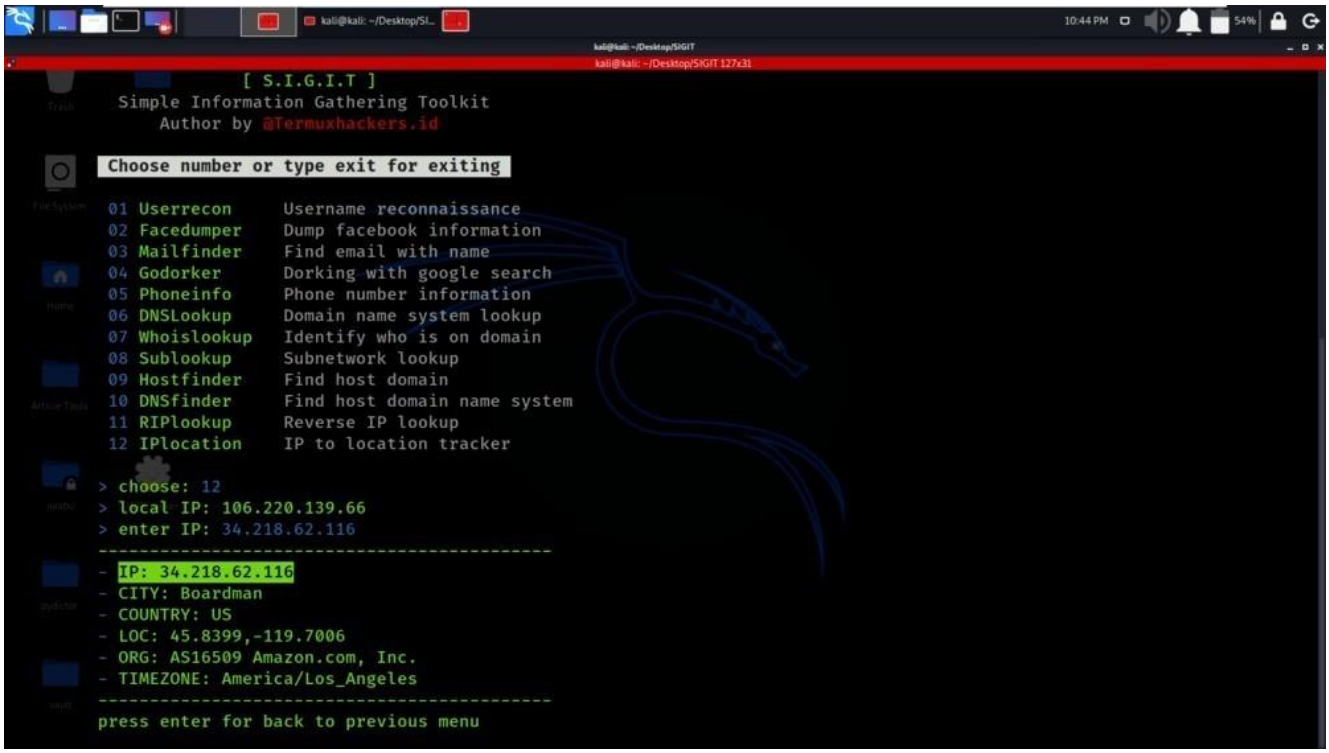


Example 3: IP to location tracker

Select option 12



We have got the location of IP Address.



Features :

- userrecon - username reconnaissance
- facedumper - dump facebook information
- mailfinder - find email with specific name
- godorker - dorking with google search

2.2 Problem and Weakening Of Current System

The toolkit is very good in terms of reliability but there are still some bugs present which are needed to be fixed and in-addition there are not more number of tools with vast area of scope. Also toolkit comes in form of a package which is not available to get used by Windows or MAC users resulting in less reach of the toolkit.

2.3 Requirement of New System

A New System should be made which can be used by user of any platform regardless of his/her operating system. And as our tool is going to be a django application any user can access our application worldwide by just typing the url resulting in global reach of our software.

2.4 Technical Feasibility

Infogrepper is a complete web based application. The main tools and technologies that are associated with infogrepper are :

i. Django Framework
ii. Python Language
iii. Bash Language
iv. SQL
v. VSCode
vi. Diagram Drawing Tools -
    a. NCLASS
    b. Asana
    c. Visio
    d. draw.io

Each of the technologies are freely available and the technical skills required are manageable. Time limitations of the product development and the ease of implementing using these technologies are synchronized.

Initially the project will be hosted locally but for the later implementations it will be hosted in a paid web hosting space with a sufficient bandwidth. Bandwidth required in this application is very low, since it doesn't incorporate any multimedia aspect.

From these it's clear that our project is technically feasible.

2.5 Resource and Time Feasibility

Resources that are required for the infogrepper project includes :

I. Programming Device (Laptop and/or PC)
ii. Hosting Space (Freely Available) iii.
Programming Tools (Freely Available) iv.
Programming Individuals

So it's clear that project infogrepper has the required resource feasibility.

2.6 Social/Legal Feasibility

Infogrepper uses freely available development tools and provide the system as an open source system. Only the maintenance cost will be charged from the potential users. Tools and Technologies used in the whole project are open-sourced or are free to use without any payment subscription.

2.7 Financial Feasibility

Being a web application Infogrepper will have an associated hosting cost. Since the system doesn't consider of any multimedia data transfer, bandwidth required for the operation is very low.
The system will follow the freeware system standards. No cost will be charged from the potential customers/users. Bug fixes and maintaining tasks will have an associated cost. At the initial stage the potential market space will be individual penetration testers but at the larger scale the project can be used in any large organizations.
From these it's clear that the project infogrepper is financially feasible.

Chapter 3. SYSTEM REQUIREMENTS STUDY

3.1 Functional Requirement

- Browser – Brave, DuckDuckGo, Chrome, Microsoft Edge or any other browser.
- Minimum CPU or processor speed - Intel or AMD processor with 64-bit support; Recommended - 2 GHz or faster processor.
- Minimum free storage space – 512MB of free disk space.
- Wireless connectivity - Internet connection required for software activation
- Operating system – Linux, Windows and Mac

Each tool is bound with certain open source python libraries to get work done hassle free. I will mention the libraries corresponding to tools for better clarification.

I. For ROBOTS TESTER we will require these 2 libraries -

- requests for manipulating requests
- urllib3 for manipulating urls

II. For MASSCANTONMAP we will use these 2 libraries -

- argparse for passing the arguments like ports and ip
- subprocess for multithreading for masscan and nmap

III. For SHODAN RECON we will consider these 2 libraries -
- shodan module for shodan related queries
- ip tools for converting ip requests to dns requests

IV. For DORKSCAN we will put in use of these 2 libraries -

- GeckoDriver for proxifying WebDriver and Firefox
- Selenium for automated testing

V. For PYCAT we will need these 3 libraries -

- socket for socket programming
- threading for multithreading
- subprocess for handling processes parallely.

VI. For TORRENT METASEARCHER we want these 2 libraries

- BeautifulSoap for web scraping purposes
- aiohttp for asynchronous network operations

VII. For IP GEO we will necessitate the following library -

- requests for manipulating IP Requests

VIII. For WEBURL_SCRAPPER we will take these 3 libraries -

- urllib for url functions
- BeautifulSoup for web scraping purposes
- SSL to avoid certificate verifications

3.2 Non-Functional Requirements

1. Speed – With usage of robust Django framework our application ensures speed to users.

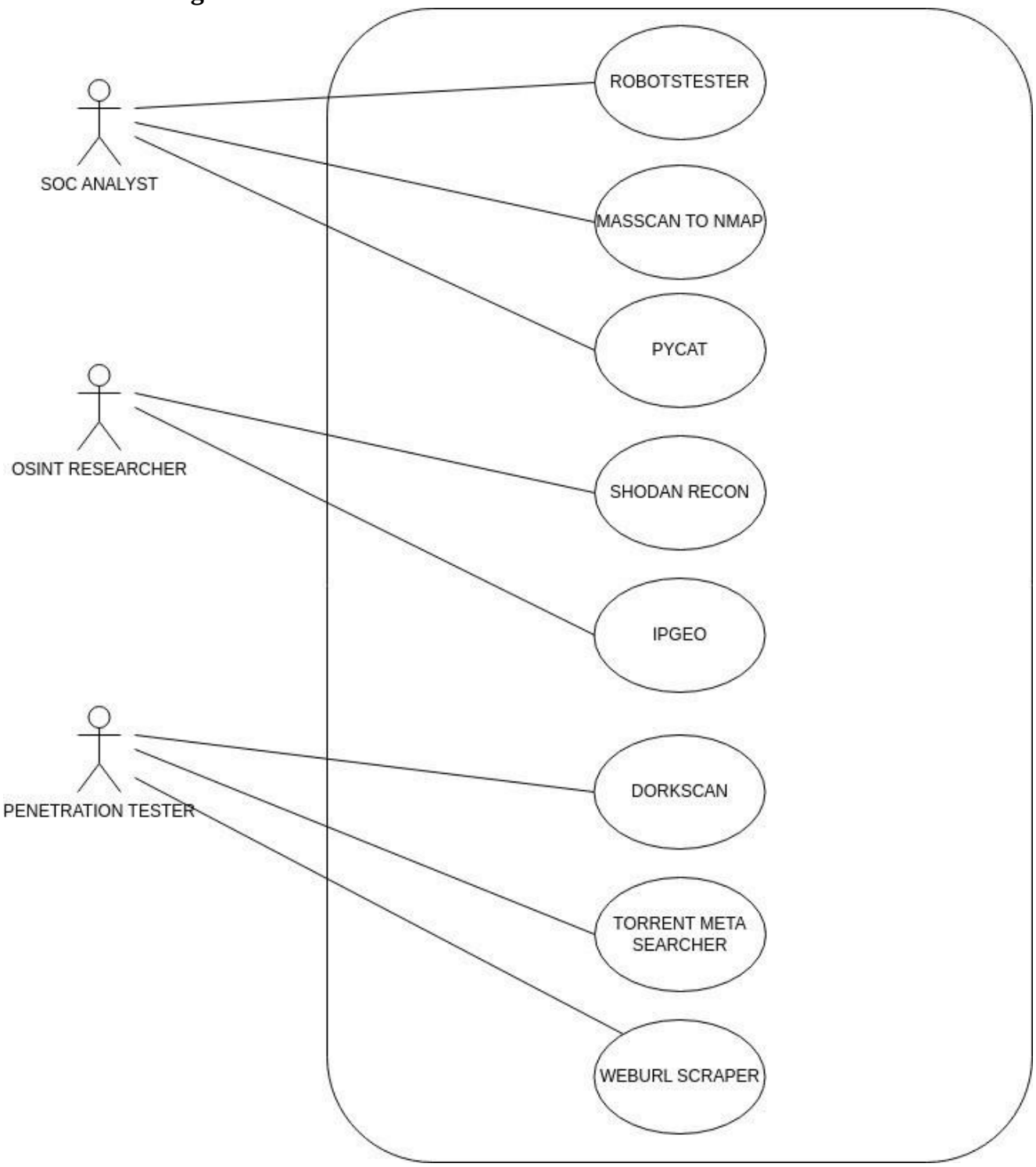
2. Security – Login credentials are used to access our application, hence giving it a secure gateway.

3. Portability – Our application is accessed by everyone with application URL, thus making it portable

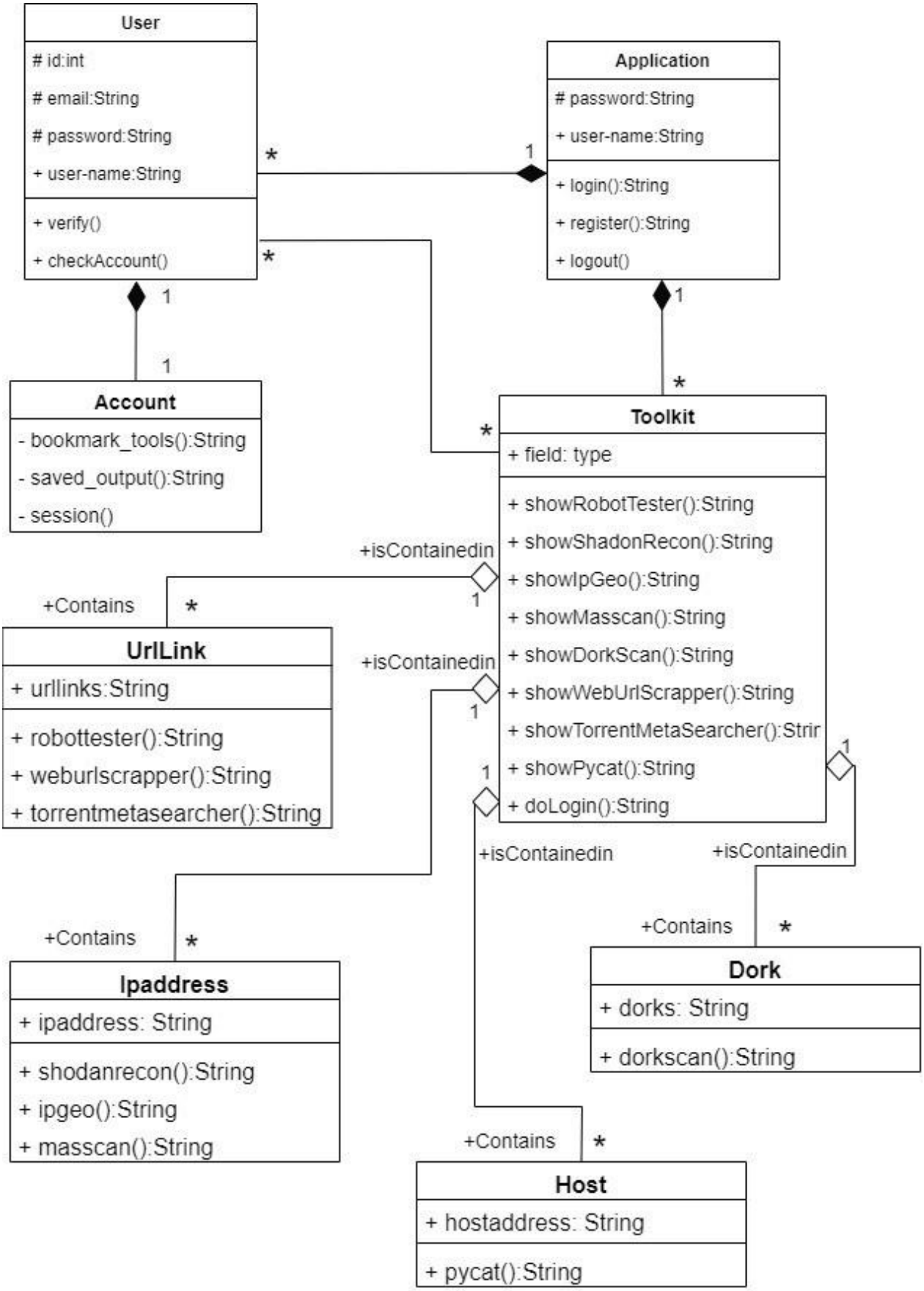
4. Capacity – Very low capacity for any device to access as it is web application

Chapter 4. SYSTEM DESIGN

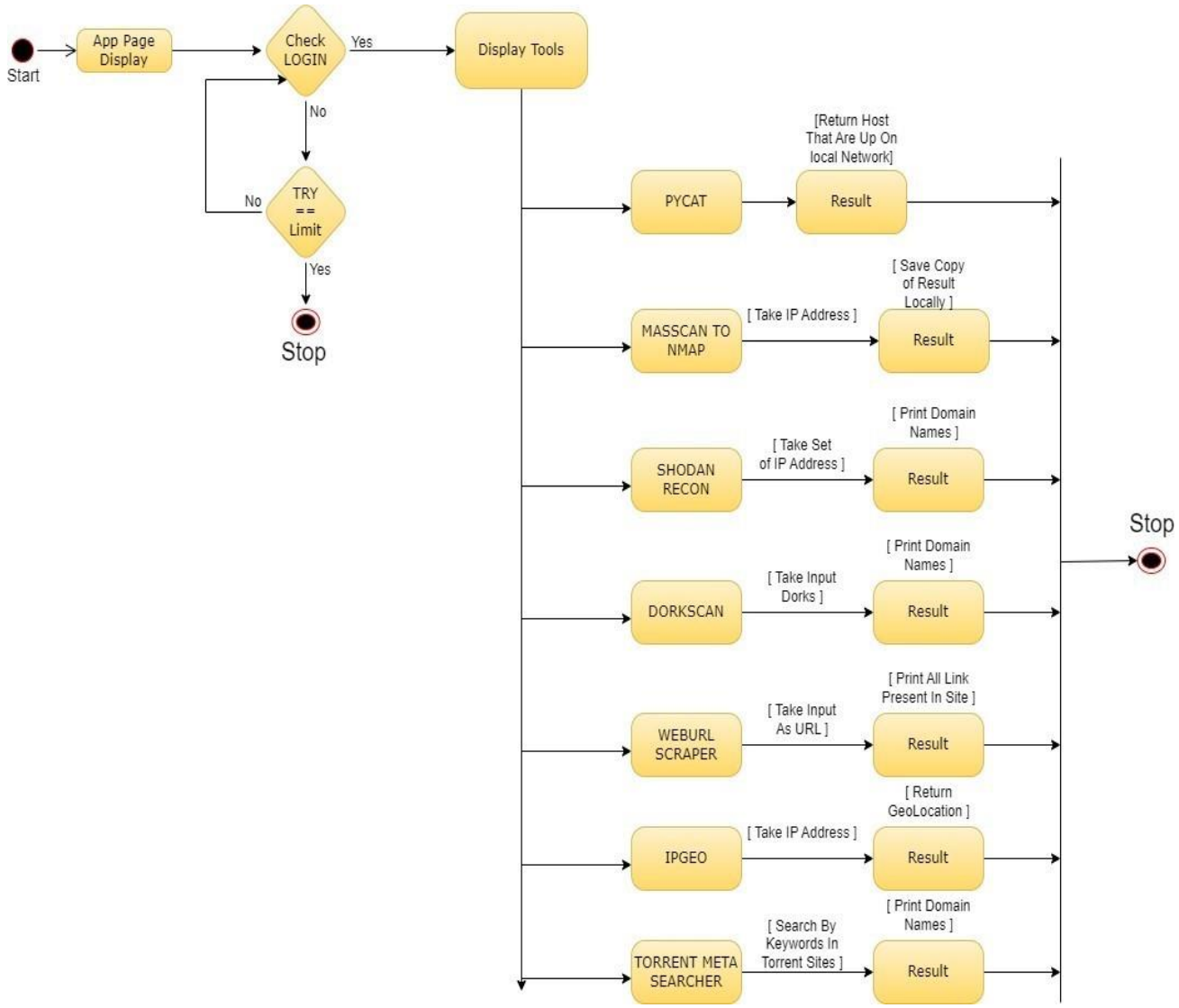
4.1 Use Case diagram



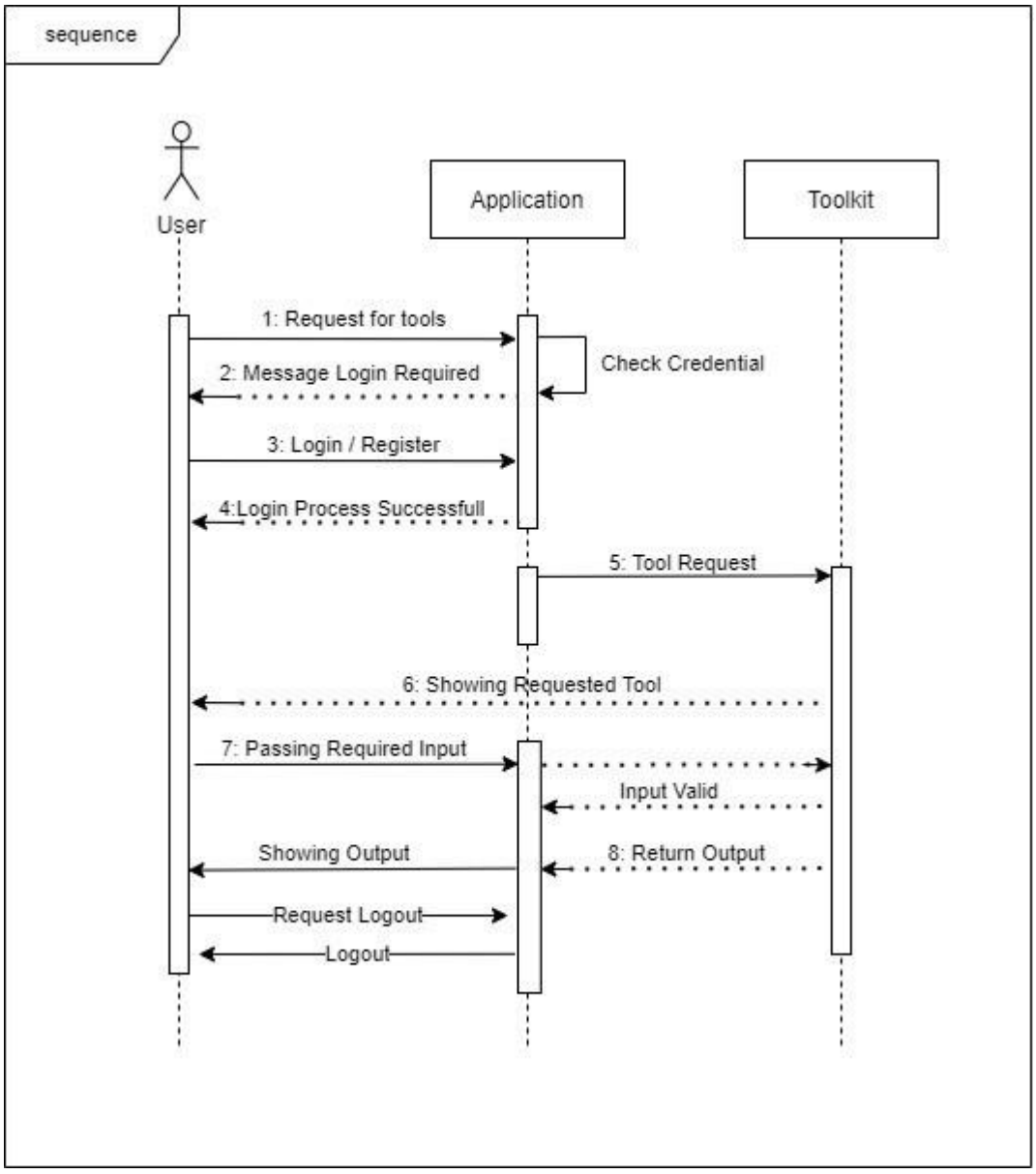
4.2 Class Diagram



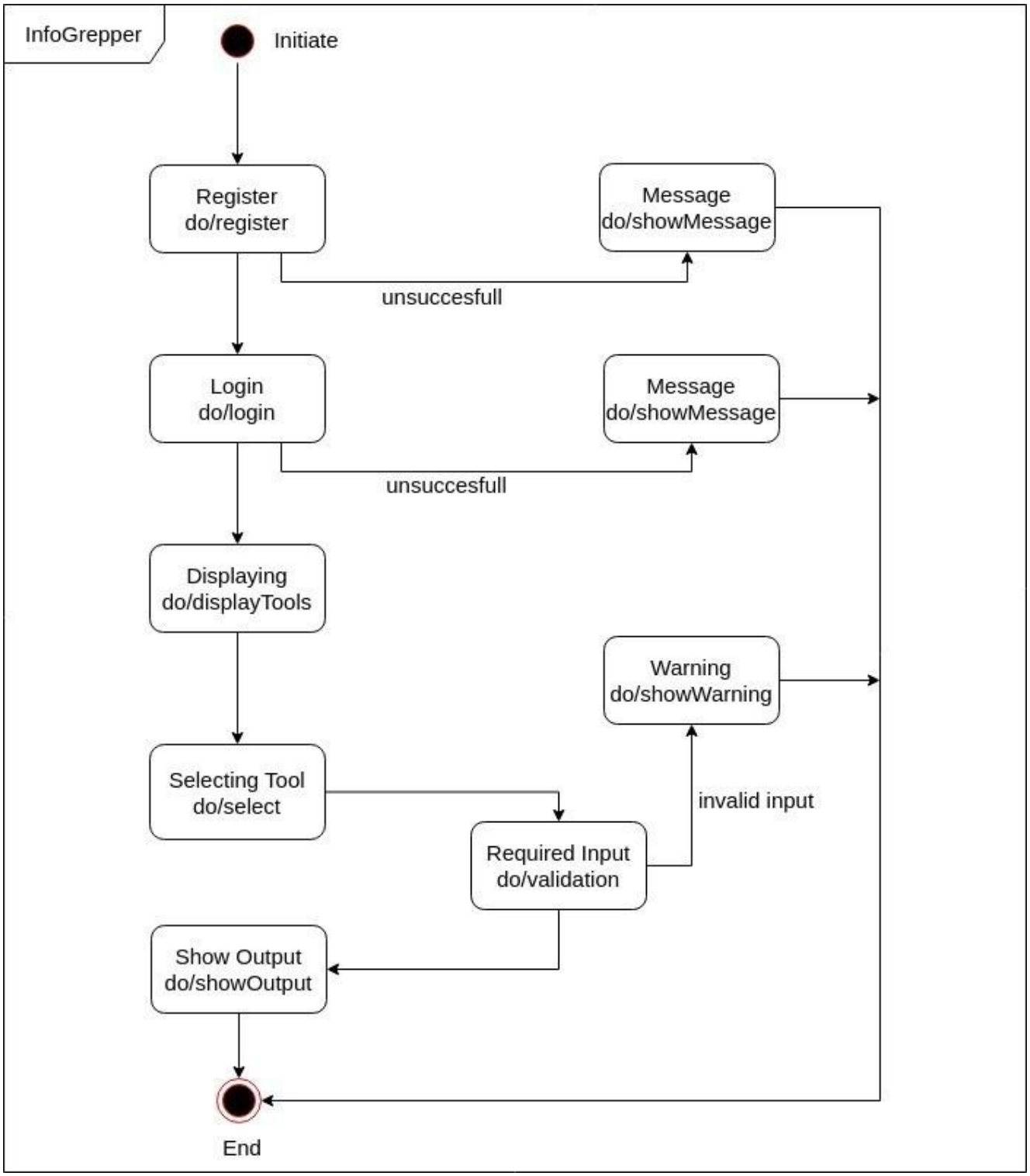
4.3 Activity Diagram



4.4 Sequence Diagram



4.5 State Diagram



5. Conclusion

With the continued evolution in technology, Hacking is becoming more and more essential now-a-days. Hacking helps keeping businesses and government organizations out of trouble caused by hackers who are trying to steal important data. By increasing digital network security, one can easily avoid security breaches by doing real-world testing.

It gives a great benefit if precautionary methods are taken in advance by all the firms. By working on safety, one can easily ensure that the clients and customers have all of their faith in one's organization. Hackers are clear and know all the potential entry points to enter the system. It is essential to repair those entry points to avoid a crisis.

It can assist owners in identifying problems inside the organization's firewall or system safety. It also enables companies to assess security from a hacker's perspective, rectifying any vulnerabilities while they threaten the achievement.

Regardless of the scandal involving the notion, information gathering assists firms and govern ments in protecting sensitive data from unfriendly hands. Their multiple benefits and relevance suggest that, as digitizing increases, stronger security techniques are essential to improve cyber crime.

Enhancement of technologies is also increasing security threats which have opened so many ways for a hacker to intrude whenever they want to. The ways of stealing data have also increased with time, and now hackers have devised creative ways to intrude and steal confidential and important data. All the firms who have taken preventive measures can save their image from getting spoiled when important data is leaked. These measures will help the firms to maintain their trust in the eyes of their clients.

6. REFERENCES

1.SIGIT : Simple Information Gathering Toolkit
https:// www.geeksforgeeks.org/sigit-simple-information-gathering-toolkit/

2. YAWAST : Open source web application information gathering toolkit
https://latesthackingnews.com/2019/02/27/yawast-open-source-web-application-information-gathering-toolkit/