A picture containing shirt

Description automatically generated

**Gurugram Police Summer Internship 2020**

**Tool Based Project**

**on**

|  |  |
| --- | --- |
| **Submitted By:** | |
| * **Aakriti Yadav** | * **Prabhdeep Singh** |
| * **Ankit Rathore** | * **Prachi Dixit** |
| * **Ankit Sharma** | * **Prateek Gupta** |
| * **Arpit Sharma** | * **Sana Chawla** |
| * **Jyoti Singh** | * **Shivani Rai** |

**Topic “Open Source Social Engineering Tool**”

**Supported by: Submitted To:**

***Society for Safe Gurgaon***

**SISL Infotech Pvt Ltd Mr. Rakshit Tandon**

**Team Members Overview**

* **Aakriti Yadav GPSI-Nte-002** 
  + BA LLB
  + USLLS, Guru Gobind Singh Indraprastha University, Delhi
* **Ankit Rathore GPSI-Tec-015**
  + Master’s in cyber law And Information Security
  + National law Institute University, Bhopal
* **Ankit Sharma GPSI-Tec-014**
  + BTech (CSE)
  + B K Birla Institute of Engineering and Technology, Pilani
* **Arpit Sharma GPSI-Tec-028**
  + Bachelor’s of Engineering (EXTC)
  + KJ Somaiya Institute of Engineering and Information Technology, Mumbai
* **Jyoti Singh GPSI-Tec-084**
  + M.Sc. Forensic Science
  + SHUATS, Allahabad (U.P)
* **Prabhdeep Singh GPSI-Tec-164**
  + BTech (CSE)
  + Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib
* **Prachi Dixit GPSI-Tec-165**
  + BTech in CS with specialisation in Cyber Security and Digital Forensics
  + Vellore Institute of Technology, Bhopal
* **Prateek Gupta GPSI-Tec-173**
  + BTech (Chemical Engineering)
  + Indian Institute of Technology, Roorkee
* **Sana Chawla GPSI-Sch-074**
  + Science
  + The Cathedral and John Connon School, Mumbai
* **Shivani Rai GPSI-Tec-211**
  + M.Sc. in Information and Cyber Security
  + Makhanlal Chaturvedi University, Bhopal

**Undertaking for Originality for the work**

I, **Prachi Dixit** give undertaking that the group project titled “**Open Source Social Engineering Tool**” submitted by me, towards the partial fulfilment of the requirements for the certificate of Gurugram Police Summer Internship 2020, is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. I understand that in the event of similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

Student Name:

Prachi Dixit

Date :

11th July 2020

**ACKNOWLEDGEMENT**

We express our sincere gratitude to Gurugram police for providing us an opportunity to be part of this Gurugram Police Summer Internship 2020 to complete this Tool Based Project on “**Open Source Social Engineering Tool**”

We sincerely thank

1. Sh. K.K. Rao (Commissioner of Police, Gurugram)
2. Ms Nikita Gehlawat (DCP)
3. Ms Pankhuri Kumar (ACP SO1)
4. Mr. Karan Goyal (ACP Cyber Crime)
5. Mr. Rakshit Tandon (Cyber Safety Advisor Cyber Crime, Director Council of Information Security, Consultant-Internet & Mobile Association of India)
6. Ms Akshita Jain (Systems Engineer)

for their guidance and encouragement in carrying out this project work.We also thank to all the experts invited to GPCSSI 2020.

Lastly, we thank almighty, our parents, our siblings and friends and whole Gurugram Police Team for their constant support without which this project would not be possible.

**TABLE OF CONTENTS**

1. [Abstract: 15](#_Toc43078359)
2. [Introduction: 16](#_Toc43078360)
3. [Main Contents: 20](#_Toc43078361)
4. [Need Of The Tool: 34](#_Toc43078362)
5. [Implementation: 35](#_Toc43078363)
6. [Conclusion: 36](#_Toc43078364)
7. [Reference: 37](#_Toc43078365)

**ABSTRACT**

During the past few years, Social Engineering has evolved rapidly and has become a prevalent technique in cybercrime and terrorism. Use of Open Source Intelligence, however, is a double-edged sword as the data may be harnessed not only by intelligence services to counter cyber-crime but also by the perpetrator of criminal activity who use it to undermine the victim. Since they have an enormous amount of information and billions of users available made them critical with respect to the user security and privacy. In order to practically evaluate the risks of information leakage trough the target organizations employees, when performing a penetration test, an ethical hacker must consider social engineering as a very important aspect of the performed test.

In this project we have worked on an integrated ‘Open Source Social Engineering Tool’, designed specifically to perform advanced information gathering processes to attain maximum open source information from a phone number, domain, Facebook, Twitter and Instagram usernames. The tool is based on the Python Programming Language and operates on various operating systems like Linux, Mac and Windows. Based on the above work, this tool attempts to address these conceptual deficiencies available in the cyberspace which can be a threat to the privacy of an individual.

**INTRODUCTION**

The 21st century has been marked by the increasing influence of the digital world on all aspects of modern society. The safety and security of that modern society is becoming a grave concern as the rapid development of technology and information has led to an increased availability of personal information online.

Law Enforcement and Security Agencies rely on information technology that facilitates the collection of open source information open source means any information which is available publicly in a legal manner. Information that are unclassified and include sources ranging from newspapers, governmental reports, public data, maps, academic sites to blogs, social networking sites, apps and web-based communities.

With the evolution of the internet, a vast array of information has become retrievable with the click of a mouse. In addition to this accumulation of valuable data, the internet also contains large quantities of personal information, often posted online by people themselves through social networking sites, blogs or apps. Individuals regularly share personal information online, which is stored as digital data in databases or in the cloud.

Examples: In 2014 according to (Hoobs et al) large scale of user generated information include the 500 million Tweets per day on Twitter and the 98 million daily blog posts on Tumblr , as well as millions of individual personal Facebook pages. With the evolution of the digital information, it has been essential that law enforcement agencies and security agencies now harvest relevant content through investigations and regulated surveillance, to prevent and detect cyber activities.



Figure 1.1: Social Sites

**OPEN SOURCE INTELLIGENCE TOOL (OSINT)**

The term “open source” refers specifically to information that is legally available for public consumption. If any specialist skills, tools, or techniques are required to access a piece of information, it can’t reasonably be considered open source. Some source of information may be openly accessible without any authentication. Many law enforcement and security agencies use these tools to access the depth of the information of a target which is very useful for connecting the link between the target and the crime. OSINT tools have been used for a long time by the government and corporate world. The OSINT tools are used for reconnaissance: the open source information by the phone numbers, email ID, domain name etc and it also play a critical role in social sites, dating apps and blog etc.

**GPCSSI- OSINT TOOL**

****

Figure 1.2 :GPCSSI-OSINT

GPCSSI Tool is based on python programming. From this tool we access the information without any authentication required, we have 0-5 number from which every different number, we will get different information of a target. The choice of number from 0-5 value gives information of a target as follows:

1. Phone Number: When we enter the 1 value then type the target phone number and the information we get like type, Country Code, Country, Location and Carrier.
2. Instagram: When we enter 2 value then type target username and then we get profile like Name of the target, URL, Followers, Following, Posts, Bio, External URL, Private (type of profile), Profile Verified or not, Profile pic URL, Business Account, Connected to FB (One of the Social Sites), Joined recently or not.
3. Twitter: When we enter the 3 value then type username then we get account information like Name, Location, Description, Created on (Date),URL associated, Profile image, Total Followers, Status Count and Total Following.
4. Facebook: When we enter the value 4 then type target username we get account information like Name, Work details, Education details, Current City and Home-Town.
5. Domain: When we enter the value 5 then type domain name and we get information about domain like Email ID, First and Last name of a target person, Position (Designation), Link with Twitter account details and interlink with domain information also give.
6. Exit: When we enter 0 we simply exit from the GPCSSI-OSINT terminal.

**MAIN CONTENT**

The GPCSSI-OSINT Tool has five different components, each component functions based on a Python Code. It is a Command Line Tool which can be executed on a Mac, Windows, Linux operating systems

**Common features:**

* OSINT foot printing using external APIs, Google Hacking, phone books & search engines: Open Source Intelligence (OSINT) takes three forms: Passive, Semi-passive, and Active. There are several ways to deal with information in an intelligence context. Especially when it comes to foot printing. GPCSSI-OSINT practices **Passive Information Gathering**(or Passive Reconnaissance), which means that it will neither store data nor use any third-party sources. But it gathers information from many sources and filter the results to find the relevant information.
* **Use of API Keys**

GPCSSI-OSINT works on an API framework. API is the acronym for Application Programming Interface, which is a software intermediary that allows two applications to talk to each other. Any data can be shared with an application program interface. APIs are implemented by function calls (def)[[1]](#footnote-1)  composed of predefined syntax. For example, on a real estate website, one API might be used to publish available real estate properties by geography, while a second API provides the visitor with current interest rates and third API brings in a mortgage calculator. Exposing data with an API can improve the customer experience because it provides greater functionality and scope of services within a single application or other digital properties.

An API key is a unique identifier used to connect to, or perform, an API call. The API key process is similar to user authentication for web applications and mobile devices. The API call starts with one API calling another and then passing the API key to gain access. The API key signifies that the connecting API has a "password" or key and a defined set of access rights. Project authorization rules are created and managed by the API owner or source. API keys may serve as an initial authentication or security step by passing a secure authentication token. For example, an application that sends medical forms to patients would need to connect its own API to that of an application that stores medical forms. The owner of the medical forms API assigns an API key, which allows the first application to access medical forms and nothing else.

After being authenticated by the source it generates a request a http response code. i.e. “*response.status\_code ==200:* OK” which signifies that the user has established a connection with the source/API provider. Similarly, other response code have their own meaning:

## 301: “Moved Permanently”

## 401 : “Unauthorized access”

## 403 : “Forbidden”

## 404 : “Not Found”

## 200 : “Found”

* Use Of If-Else Operator:

GPCSSI-OSINT uses the if-else operator for decision making. It is a type of combined statement. An **else** statement contains the block of code that executes if the conditional expression in the **if** statement resolves to 0 or a FALSE value. The *else* statement is an optional statement and there could be at most only one **else** statement following **if**.

In the code for the GPCSSI-OSINT tool if the request is successful and response obtained from the API Call is equal to 200 then the mentioned information such as number, line type etc gets printed on the screen. Otherwise else operator will get executed and an Error message denoting “Invalid (mobile number/profile)” will get printed.

1. **PHONE NUMBER:**

Phone number plays a vital role in Social Engineering and Open Source Intelligence investigations. Almost everyone carries a phone and the phone is now linked with an individual’s life and it has the most important information of a person. In OSINT investigations, it is important to find the most relevant information available on open source platforms about the subject’s phone number. One aspect of the GPCSSI-OSINT is an advanced tool to scan phone numbers using only free resources. The goal is to first gather standard information such as country, area, carrier and line type on any international phone number with a very good accuracy. The tool then searches for footprints on search engines to try to find the VoIP provider or identify the owner. The number itself shows some information like country, city/state and sometimes carrier; while the other information can be extracted by using the open platform available.

A screenshot of a cell phone

Description automatically generated

**Features:**

* Checks if the phone number exists.
  + i.e. If we put an incomplete phone number (+91 7333975559$$) then it will not be able to process the request.
* Gather standard information such as type, country code, country, location and carrier:  
  *Type*: It provides information about the carrier which is used to expend services.  
  *Country code:* It reveals the country to which the number belongs. The tool takes an input in international format  ie."+ 93 xxx-xxx-xxxx " in which +93 signifies (Afghanistan) international calling codes assigned by International Telecommunication Union (ITU). The first four digits initially indicated an operator's code, while the remaining six digits are unique to the subscriber. However, with mobile number portability in place, the first four digits no longer indicate a particular operator.
* *Location*: Based on number provided it fetches the city/state in which the number is registered
* *Carrier*: It provides information about the Internet Telephony Service Provider(ITSP) which the subject is using to make calls and other services. An Internet Telephony Service Provider (ITSP) offers digital telecommunications services based on Voice over Internet Protocol (VoIP) that are provisioned via the Internet[[2]](#footnote-2).
* Uses custom formatting for more effective OSINT reconnaissance.
* Run your own web instance as a service.
* Programmatic usage with Go modules.

**2. INSTAGRAM**

The Instagram OSINT Tool gets a range of information from an Instagram account that you normally wouldn't be able to get from just looking at their profile. This tool can be used by Infosec Researchers, Penetration Testers, Bug Hunters and Cyber Crime Investigators to find deep information about their target. It also aggregates all the raw data, visualizing it on a dashboard.

**Features:**

* Performs OSINT on a username and finds out information of that user profile.
* Correlates and collaborates the result, showing them in a consolidated manner and gives information related about:
* Username
* Profile Name
* URL
* Followers
* Following
* Number of Posts
* Bio
* Profile Picture URL
* Is it a Business Account?
* Is it connected to a FB account?
* External URL
* Joined Recently?
* Business Category Name
* Is it private?
* Is it Verified?
* Downloads Public Photos.

**Working:**

* GPCSSI-OSINT works on API framework. API is the acronym for Application Programming Interface, which is a software intermediary that allows two applications to talk to each other. Any data can be shared with an application program interface.
* This tool has a module involved to scrape social media account details from Instagram.

**Modules Used:**

* ***Bs4* :** Beautiful soup is a Python library for pulling data out of HTML and XML files.
* ***Time* :** This module provides various time-related functions. For related functionality, see also the [datetime](https://docs.python.org/3/library/datetime.html#module-datetime) and [calendar](https://docs.python.org/3/library/calendar.html#module-calendar) modules.
* ***Json* :** JSON can store Lists, bools, numbers, tuples and dictionaries. The JSON module is mainly used to convert the python dictionary above into a JSON string that can be written into a file.
* ***Random* :**This module implements pseudo-random number generators for various distributions
* ***String* :**The [string](https://docs.python.org/2/library/string.html#module-string) module contains a number of useful constants and classes, as well as some deprecated legacy functions that are also available as methods on strings.
* ***OS* :** provides functions for interacting with the operating system.
* ***Sys* :** It provides information about constants, functions and methods of the Python interpreter.
* ***requests*** **:** Make a request to a web page, and print the response text.
* ***Collections* :** They are used to improve the functionalities of the built-in collection containers.
* ***Urllib.request* :** This module defines functions and classes which help in opening URLs (mostly HTTP).
* Here we have defined a constructor function named (def: get\_profile) which uses required syntax for an application to be called.
* It is getting the request from “<https://www.instagram.com/>” with the help of a module request.
* The code is fetching Username, profile name, URL, Followers, Following, Number of post, Bio, Joined Recently, Profile Picture URL, Business Account, Connected to a FB account or not, External URL, Joined Recently, Business Category Name, private or not, Is Verified or not , Public Posts for public account.
* For private accounts, this code is fetching all details like a public account except downloading private posts but it can download profile pictures from private accounts.

A screenshot of a cell phone

Description automatically generated

1. **TWITTER**

There are currently 1.3 billion accounts on Twitter. Twitter is a hub for sharing and receiving information. The other feature of the GPCSSI-OSINT is one that scans and retrieves information from the Twitter account associated with a username that is entered.

How Twitter works:

The basic purpose of Twitter is that it allows you to post a short message which is then sent to all your contacts or ‘followers’ on Twitter. The most basic aspect of Twitter is a ‘Tweet’ which is the name given to the short message a user posts on his/her Twitter account. A broadcast on Twitter can either be restricted to only your followers or can be made visible to everyone on Twitter if the account is public. Tweets can include text and media files. Every Twitter handle includes some or all of the following information: full name of the user, a user created username, date of birth and of account creation, number of followers and following as well as other additional information like designation, links to websites etc.

**A screenshot of a cell phone

Description automatically generated**

**The Twitter API:**

Twitter provides programmatic access of its APIs to companies, developers as well as the twitter users to facilitate the sharing of information and allow them to integrate Twitter into their own software. Twitter provides access to the data that the users themselves have chosen to share with the public. The Twitter API consists of a wide range of endpoints ( an endpoint is the address that corresponds with the information the user wants from Twitter) which include: Accounts and Users, Tweets and replies, Direct Messages and Advertisements .

**Features:**

* The code makes use of the ‘Tweepy’ module. Tweepy is an Open Source python library which is used to access the Twitter API and includes classes and methods that represent the models of the Twitter API, allowing for easy implementation of these models in a Python code. The GPCSSI-OSINT uses the following attributes of the Tweepy module:
  + 1. The OAuthHandler attribute is used to set the credentials to be used for each API Call to authenticate the request.
    2. The API class provides access to the functionality of the Twitter API. The methods of the API Class include Methods for: users, followers, likes etc and these methods return details of the account associated with the username inputted.
    3. The Stream class is used to actively ‘listen’ for tweets of the specific user account in real time. The Stream class of Tweepy uses the Twitter API to get the tweets of the user which match specific criteria (for example, only the media files) and the StreamListener attribute is used to process the tweets received.

**Working:**

* The code creates a directory whose path is the *username* and creates three files in this directory:
  + 1. A username.txt: the tool writes the information gathered using the Tweepy attributes which includes the name of the user, the date of creation of the account, number of followers and following, location and the profile image and writes them onto a username.txt file in the created directory
    2. The user timeline method is used to read the tweets and retweets, which are then written to a username.csv file along with the number of characters, date on which they were tweeted, number of likes and their text content. DataFrame attribute of the pandas module is used to format the file and the tweets, data, number of characters are stored in NumPy arrays.
    3. The third file is the username.jpg file that contains any type of media file in the tweets of the user, for example: jpg, mp4 files. The code first checks if any media files are present and only then creates the .jpg file to store these media files.

A screen shot of a computer

Description automatically generated

A screenshot of a cell phone

Description automatically generated

1. **FACEBOOK:**

Facebook is one of the most well-known and popular social media sites, which makes it an attractive target for Social Engineering attacks. The GPCSSI-OSINT includes a tool which extracts the Facebook profile of an associated username and provides information which includes the Name of the user, the current place of employment/education, city and hometown.

**Features/Working:**

* It works using the same API Module that was described earlier
* In addition, the Facebook module of the tool makes use of the bs4/Beautiful Soup Python library to parse the HTML source code of the Facebook webpage of the particular user whose username was entered. It function is to access specific data such as user ID from the webpage, remove the HTML markup from the data and save the information so that it can be displayed in the output. It is essentially a tool for web scraping (the extraction and content and data from a website) which enables us to parse through documents that have been accessed from the web.
* A picture containing green, sitting

  Description automatically generatedThe code uses the attributes of Beautiful Soup to identify specific information such as: ID, education, company from the Facebook webpage of the particular user associated with the username and that information is then converted to text and printed in the output.

1. **DOMAIN:**

* GPCSSI-OSINT domain reconnaissance is an Open Source Intelligence (OSINT) Framework for scanning IP Addresses, Emails, Websites, Organizations and find out information from different sources. This tool can be used by Infosec researchers, Penetration Testers, Bug Hunters and Cyber Crime Investigators to find deep information about their target. It also aggregates all the raw data visualized on the dashboard.

**Features:**

* Performs OSINT on a Domain / Email / IP Address and finds out information from different sources
* Correlates and collaborates the results, showing them in a consolidated manner. This information includes the email id, first and last name of the person, position, LinkedIn identity and twitter handle associated with a domain name
* Uses specific script/launch automated (json) OSINT for consolidation of the data.
* Provides an easy command line console.

**Working:**

* GPCSSI-OSINT works on an API framework. API is the acronym for Application Programming Interface, which is a software intermediary that allows two applications to talk to each other. Any data can be shared with an application program interface.
* This tool has a module involved to scrape social media account details from Instagram, Facebook, LinkedIn and Twitter.
* Here we have defined a function named maildb, which uses required syntax for an application to be called.
* It accepts the input in the format of “@” or “.com” (which is controlled by the use of an if-else function). If the input is in any other format it will print an error message showing ‘Invalid Email Address’.
* It also works on the same API concept which was earlier defined
* This module also uses the json syntax to fetches information from various data breaches for example: email, phone no. etc. **JSON[[3]](#footnote-3)** stands for JavaScript Object Notation. It is a lightweight data-interchange format that is used to store and exchange data. It is a language-independent format and is very easy to understand since it is self-describing in nature. There is a built-in package in python that supports JSON data which is called json. The data in JSON is represented as quoted-strings consisting of key-value mapping enclosed between curly brackets {}.
* Domain module performs various scans for domain check for vulnerability and spider crawlers and fetches information about mainly email id, first and last name of the person, position, LinkedIn identity, Twitter handle and we can add many more attributes for related information which is linked to that email id or domain depending on the requirements.
* It also uses a metadata analyser to relate the different attributes.

A screenshot of a cell phone

Description automatically generated

**NEED OF THE TOOL**

In this Modern era where technology is expanding day by day there comes higher risk of threats also. This creates increase vulnerability in the area of cyber, information security etc. The people, organisations, businesses are exposed to such attacks without even realizing. Nowadays data protection is becoming a huge priority for any person or organisation , so such tools are necessary for the protection of their information security .

This tool is the compilation of the open source tools available online, So instead of getting to separate tools for information one can resort to this tool for easy and quick access. This will make the investigation process little convenient for the person. This tool will basically help in getting information about the person through their phone numbers, emails, Instagram accounts, twitter account, and Facebook account. This will help you to search what information is out there of the person. The tool has been designed with the motive to help the law enforcement agencies in their investigation procedure.

**IMPLEMENTATION**

The 5 individuals modules described above have been merged into a single file called finalmain.py using the if-else operator.

The GPCSSI-OSINT tool is made for the Linux, Windows and/or Mac environment.

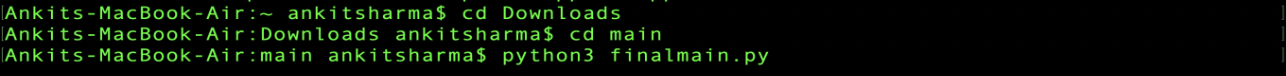
The tool has certain requirements which are listed below and can be installed using the command:

***pip3 install -r requirements.txt***

Requirements:

* requests
* bs4
* tweepy
* wget
* pandas
* numpy

The tool is operated from the Command Line by running the following code :



The source code of the tool can be found on the following GitHub link :

<https://github.com/prabhdeep1/GPCSSI-OSINT>

**CONCLUSION**

For the purpose of addressing conceptual deficiencies of social engineering in cybersecurity since data is the new fuel in the digital world, this tool attempts to provide a new aspect of using open source intelligence in a fruitful way within legal ambit to facilitate reconnaissance problem. This tool-based research facilitated us with the evolution of the cyber world and many relevant conceptual problems of social engineering in cybersecurity are investigated, analysed and discussed in a systematic way, which serve as materials for the defining work. The methodology of research is based on the analysis of merits and demerits of different kinds of categorization theories. The proposed report eliminates the conceptual inconsistencies, vagueness and confusion; covers the mainstream intension in conceptual evolution; clarifies the conceptual boundary; and avoids the overgeneralization, abuse and decomposition of content. Although providing a conceptual paradigm of social engineering in cybersecurity is difficult, this project makes its best efforts to achieve our hypothesis of collecting maximum information from open source intelligence available. In future work, we will systematically research the human vulnerabilities and their effect on the mechanisms used in social engineering.

**REFERENCES**

1. <https://medium.com/@SundownDEV/phone-number-scanning-osint-recon-tool-6ad8f0cac27b>
2. <https://github.com/sundowndev/PhoneInfoga>
3. <https://whatis.techtarget.com/definition/APIey#:~:text=%20The%20steps%20below%20represent%20general%20steps%20common,There%20are%20generally%20two%20types%20of...%20More%20>
4. <https://help.twitter.com/en/rules-and-policies/twitter-api>
5. <https://realpython.com/twitter-bot-python-tweepy/#review-of-tweepy-functionality>

1. <https://www.programiz.com/python-programming/function> [↑](#footnote-ref-1)
2. <https://en.wikipedia.org/wiki/Internet_telephony_service_provider> [↑](#footnote-ref-2)
3. [↑](#footnote-ref-3)