OSINT FRAMEWORK FOR PENETRATION TESTING WITH AUTOMATION TECHNIQUES

ARUNMOZHI B, Assistant Professor
Department of Computer Science and Engineering.

KRISHNA PRAKASH S, Student of Computer Science and Engineering,
St. Joseph College of Engineering, Sriperumbudur, Chennai.

VINOTH KUMAR V, Student of Computer Science and Engineering,
St. Joseph College of Engineering, Sriperumbudur, Chennai.

Abstract:

Opensource intelligence OSINT is intelligence derived from publicly available information. The shifts in analysts and users will be investigated. We'll talk through data analysis, automated data collection, APIs and tools, algorithms like supervised and unsupervised, geolocational approaches, de-anonymization.

All of these factors, including ethics and context, interact inside OSINT. The need to ensure and encourage constructive usage of open intelligence is even stronger now that the playing field is levelling. One of the most difficult aspects of automating opensource intelligence from sources like social media is resolving named entities in order to identify danger actors and anticipate their future behavior. For example, my name is "Vinoth Kumar," and I am a member of the "human" category, which is a subset of "animal".

In this chapter, we look at two simple instances that make named entity resolution difficult: resolving lexical items inside a semantic context and how mistake propagation affects the accuracy of subsequent lexical analysis. As cybercrime becomes a more pervasive and changing danger, focus must shift to long-term solutions. Blocking these attacks does not provide long-term answers; it merely allows hackers to develop their attacks over time, which is very straightforward in today's climate.

As previously stated, OSINT refers to all information that is freely available to the public, including both online and offline resources.

Keywords: OSINT-Opensource Intelligence Tool, Python, Cyber Security.

Introduction:

The search, collection, analysis and application of information from open sources, as well as the methodologies and technologies used, is referred to as Opensource Intelligence OSINT. OSINT arose from the requirement for the military to gather useful and publicly available data. It is possible to find specific information that has some expertise provides advantage through the usage of an OSINT.A number of researches have been conducted proposing and developing new applications for OSINT in many fields.

A detailed literature review that explored the use of OSINT across time could not find your AI application. As a result, the goal of this project is to create a comprehensive literature evaluation on OSINT in order to examine the use of OSINT with AI. This effort was inspired by a need to address a research gap therefore it consolidated OSINT papers into publishing bases.

In terms of contribution, this paper includes a 9-step systematic literature assessment as well as aggregated data information to enable future OSINT investigations. There were 244 papers found in this study, which covered the period from January 1990 to October 2019. The systematic literature review has nine steps: keyword definition, query string definition, and publication base definition.

All information that may be found freely mostly via the internet without violating any copyright or privacy regulations is referred to as opensource intelligence OSINT. OSINT can be defined in this way to include a wide variety of sources. For example, publicly available information on social networking websites, posts on discussion forums and group chats, unprotected website directories, and any information that may be obtained by searching online. Keep in mind that most OSINT materials aren't accessible through traditional search engines like Google or Yahoo!, as many resources are buried deep into the deep and darknet, accounting for more than 96 percent of all web material.

The internet is the primary source of OSINT resources; in fact, many scholars distinguish between online and offline OSINT resources by using the phrase "Cyber OSINT" to refer solely to internet resources.

Blogs, social media websites, digital files photo, video, sound and their metadata, website technical foot printing, webcams, deep web like government records, weather records, vital records, criminal records, tax and property records, darknet resources.

Literature Survey:

A vast number of scholars have published papers in the Opensource Intelligence sector during the previous two decades. Opensource intelligence has grown in tandem with technological advancements. Many actors, including business corporations, antisocial elements, government agencies, law enforcement agencies, and others, have been enticed to integrate opensource intelligence for their benefit as a result of the massive data that has accumulated in the public domain as a result of the evolution of social media platforms. People may now easily locate and upload any sort of content because to Internet Accessibility (Edwards et al., 2017).

Fleisher (2008) presented a conceptual paper on how the growing popularity of open-source data and information affects competitive and marketing intelligence. This is a descriptive conceptual article that builds arguments from a review of three uncategorized collections of material in competitive and marketing intelligence, processing of intelligence, and analysis of market.

This article discusses the challenges they have in utilizing this data, as well as the successful techniques used by certain companies in incorporating and integrating open sources into competitive and marketing intelligence research processes. It's clear that the study was performed from the perspective of a marketing analyst and the use of intelligence gained through OSINT for the purpose of increasing marketing efforts, rather than from the perspective of the person who collects the data.

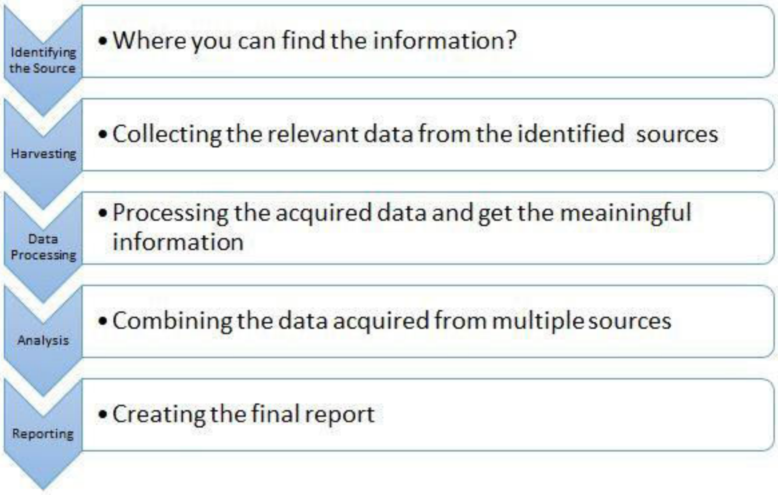
Lee & Shon (2016) proposed a new framework for cyber security threat inspection of critical infrastructure based on an Opensource intelligence. This framework included four steps: developing an opensource intelligence plan, preparing opensource intelligence, gathering information from opensource platforms, and producing security intelligence. Hayes & Cappa (2018) have demonstrated that OSINT may be used to do risk assessments for the company in order to prevent potential cyber-attacks on its critical infrastructure, which was part of the US electrical grid. A vulnerability assessment and various such open-source intelligence analysis procedures were carried out in order to profile the company's network, applications, devices, and critical IT resources.

Similar method for exploring website vulnerabilities was proposed by Wiradarma & Sasmita (2019). During the information gathering phases of penetration testing, OSINT tools like Maltego and others are used to get data on the victim a from open source. A system improvement recommendation was created by combining information from OSINT, penetration testing, and the ISO 31000 risk assessment standard.

Herrera-Cubides et al., (2020) conducted a study with an aim to investigate the evolution of production of research and study material in OSINT platform. This analysis looks at two of the material sources of OSINT such as research knowledge distribution databases and repositories pertaining to educational resources. This study provides academics with a roadmap to the current level of OSINT research and teaching, as well as a valuable metadata description in order to make resources more accessible and reusable in the educational ecosystem.

OSINT Process:

First, we need to have a good understanding of the process of sourcing and using opensource intelligence, as well as your organization’s security policies and procedures. An OSINT analyst usually follows the OSINT process as shown below.


Identifying the Source
• Where you can find the information?
Harvesting
• Collecting the relevant data from the identified sources
Data Processing
• Processing the acquired data and get the meainingful information
Analysis
• Combining the data acquired from multiple sources
Reporting
• Creating the final report

Tools used for OSINT:

This creates a large amount of data or information in a variety of formats like as audio, video, photos, and text that is free and open to the public unless prohibited by an organization or government. OSINT sources may be classified into six basic types of data flow: Print newspapers, magazines, and television from around the world, as well as between nations. Internet, online publications, blogs, citizen media such as mobile phone videos and user-generated material, YouTube, and other social media websites such as Facebook, Twitter, Instagram, and others. Due of its timeliness and accessibility, this source also outperforms a range of other sources.

1.Maltego:

Maltego offers solutions for open-source intelligence and visual link analysis.

2.Google Dork:

With a simple search query, Google Dorking, sometimes known as "Google hacking" might return information that is difficult to discover. The advanced search engine operator uses this search phrase to locate information that isn't easily available on a website. Google, often known as dorks, is an advanced search term operator that use Advanced Search Operators to locate information not easily available on websites. It's a slang phrase for Google hacking, and it's an extension of Google's extended operator.

3.Nmap:

Nmap is one of the most popular and commonly used network discovery and security auditing tools among security experts.

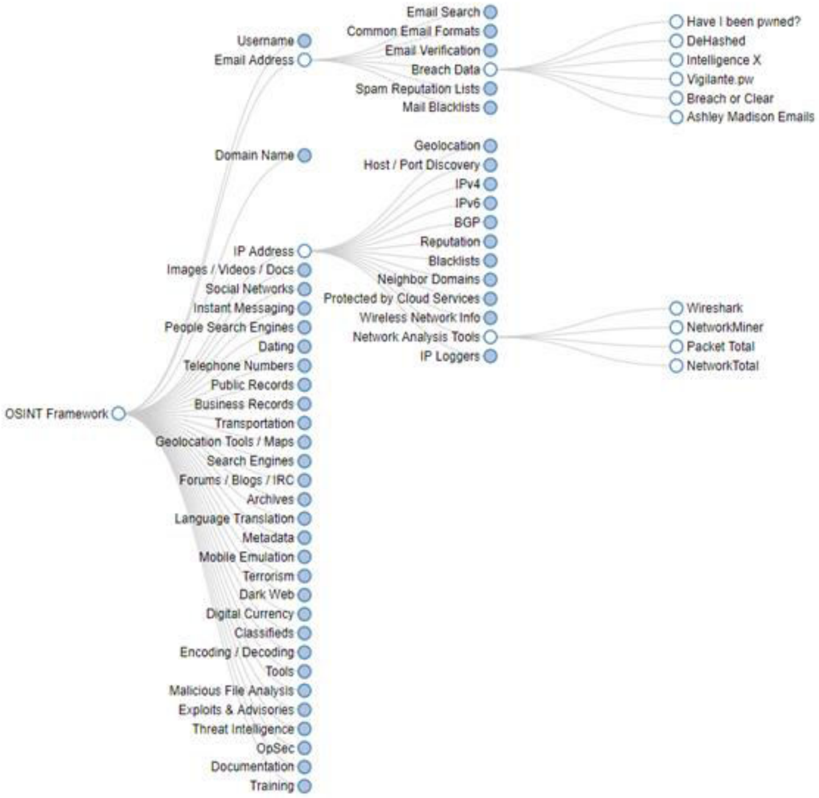
4.TheHarvester:

The Harvester is a simple-to-use, yet powerful and effective tool for usage in the early phases of a penetration test or red team engagement. Use it to acquire open-source information OSINT to assist identify a company's online external danger landscape. Using numerous public data sources, the programme collects emails, names, subdomains, IP addresses, and URLs.

OSINT Architecture:

In the below image, several categories are shown in the form of a tree, including email address, username, domain name, IP address, social networks, and so on. A sub-tree of relevant resources emerges when you click on any of the subjects.

Users can get email addresses, IP addresses, and phone data all in one location, which is why the OSINT architecture is so important for information discovery and cybersecurity.



Implementation:

- OSINT is a broad category of intelligence gathering and analysis. It does not have its own agency, rather units are dispersed among the Departments of Defence and State. When collecting information from the Internet, most OSINT collectors must exercise prudence. This might be as simple as utilising a VPN to hide their identify and collect data more quietly.

- This is where the importance of assessing sources in the whole OSINT collecting and analysis process becomes apparent. To determine a real process or reveal a fraudulent process that would affect future prediction, an OSINT analyst requires intelligence evaluation. Finally, the analysts must find a way to put the assessed intelligence to work in a final classified, unclassified, or proprietary intelligence product.

– Identification of source — Where you can find the information?

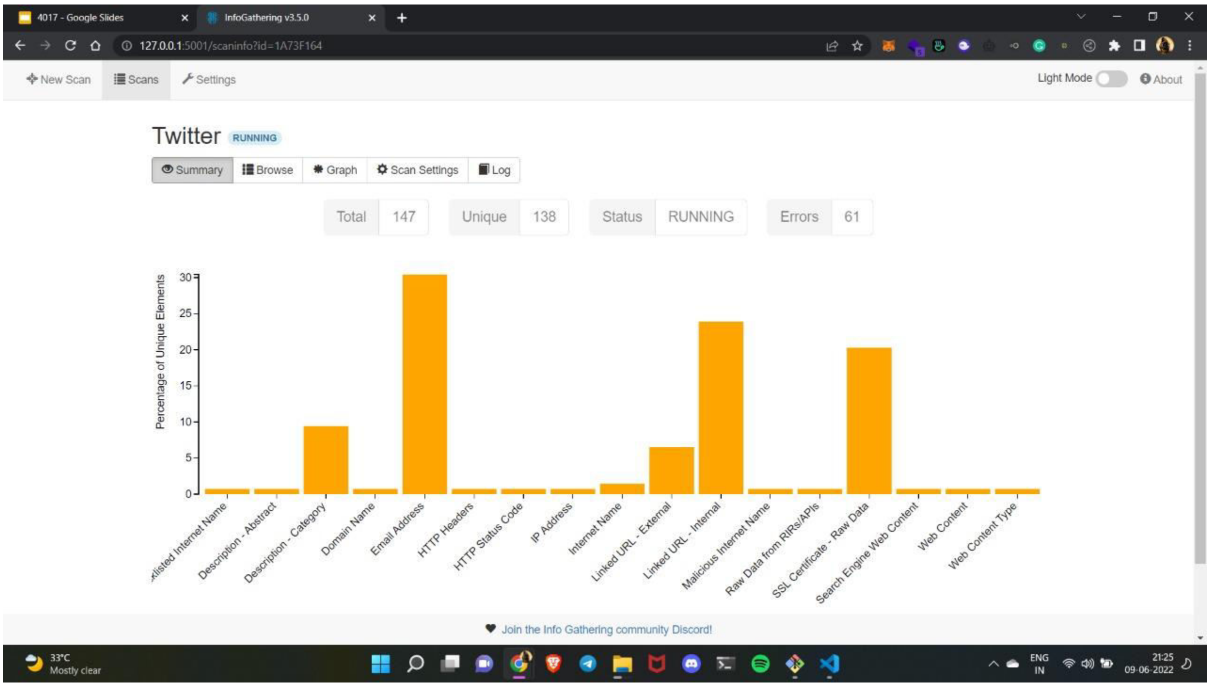
– Harvesting – Collecting the relevant data from the identified resources.

– Data Processing – Processing the acquired data and get the meaningful Information.

Analyzing – Combining the data acquired from multiple sources.

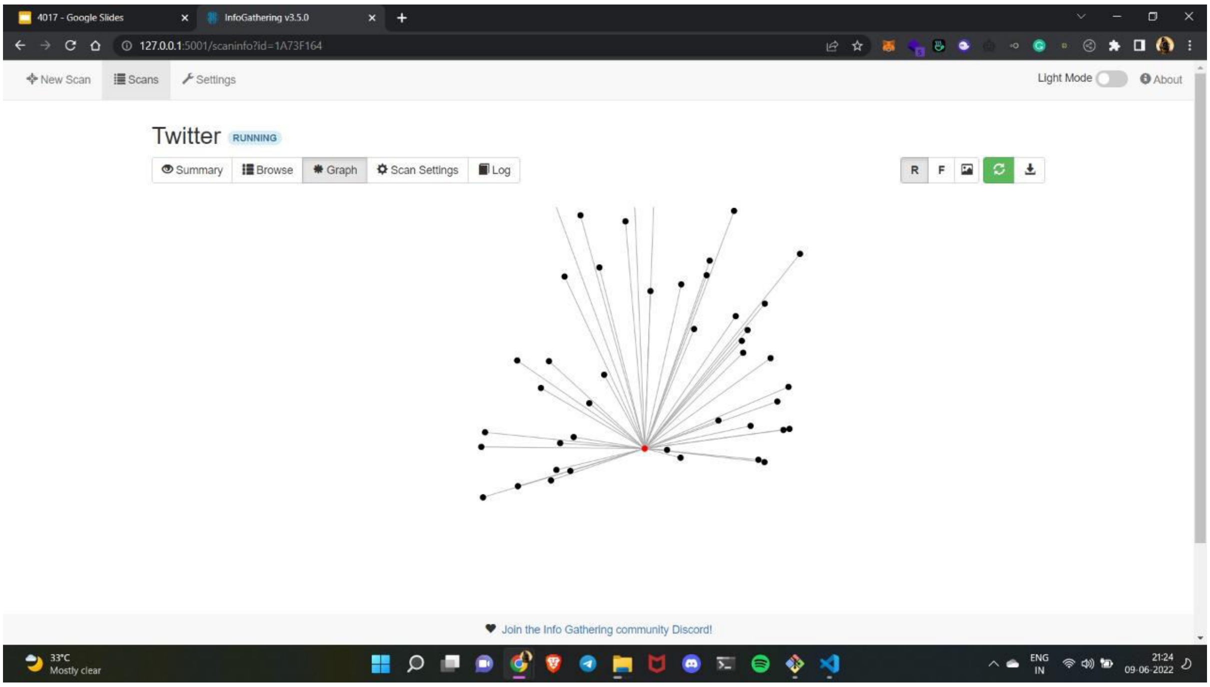
– Reporting Creating the final report.

The Output for the OSINT framework for penetration testing of Twitter Information



Total 147 Unique 138 Status RUNNING Errors 61

The Output for the Twitter graph was collected using OSINT automation techniques



Conclusion:

The OSINT architecture, as you can see, comprises a vast network of themes, connections, and tools. There are a variety of ways to gather information about the target; you must select the ideal one for you, and you will be surprised. The world of open source intelligence will not remain static; advancement in other technologies will pose a challenge to OSINT practise due to changes in the nature of data and how it is gathered; and the same advancement in technology will improve OSINT practice's ability to deal with such challenges effectively. Since the US military first implemented OSINT in the late 1980s, a lot has changed as technology has progressed.

Future Enhancement:

- OSINT gathers information about IP addresses, domain names, names, and more by automatically searching numerous public data sources using an email address.
- In comparison to the past module, we improve the process by using Accuracy and Multi-Target Scanning to collect information faster and easier.

Author Biography

Mr. B. ARUNMOZHI M.E., is an Assistant Professor in the Department of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu. He has completed his M.E, CSE under Anna University Affiliation College in the year 2011. He has done his B.E, CSE under Anna University Affiliation College in the year 2007. Mr. B. ARUNMOZHI has 11 years of teaching experience and has 12 publications in International Journals and Conferences. His area of interests includes Network Security, Computer Networks, Data Science and Machine Learning. He is an active member of CSI and IEANG. He has organized various International Conferences, workshops, and Seminars in the area of Computer Networks, Cloud Computing & Machine Learning respectively.

Mr Vinoth Kumar V B.E., Student of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu. I had attended many International Conference, Workshops, Hackathons and Seminars in the area of Data Science Organizations, Python ,Machine Learning And Deep learning Respectively. I got Placed in some respected companies like Drawlead and HCL Technologies.

Mr Krishna Prakash S B.E., Student of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu. I had attended National programme on technology enhanced learning in data science and obtained a course completion certificate and also participated in many Workshops and Seminars in the area of Data Science, Python. I got Placed in some respected companies like TCS and Dell.