

OSINT-Based Email Investigation

Samrudha Mhatre¹, Franziska Schwarz¹, Klaus Schwarz^{1,3}, Reiner Creutzburg^{1,2}

¹SRH Berlin University of Applied Sciences, Berlin School of Technology, Ernst-Reuter-Platz 10, D-10587 Berlin, Germany
Email: samrudha1995@gmail.com, mail@franziskaschwarz.net, mail@klaussschwarz.net, reiner.creutzburg@srh.de

²Technische Hochschule Brandenburg, Department of Informatics and Media, IT- and Media Forensics Lab, Magdeburger Str. 50, D-14770 Brandenburg, Germany, Email: creutzburg@th-brandenburg.de

³University of Granada, Faculty of Economics and Business, P.^o de Cartuja, 7, ES-18011 Granada, Spain

Keywords: Open Source Intelligence, OSINT, SOCMINT, Cybersecurity, OSINT Investigation, Cybersecurity Training

Abstract

Open-source technologies (OSINT) and Social Media Intelligence (SOCMINT) are becoming increasingly popular with investigative and government agencies, intelligence services, media companies, and corporations - but also for cybercriminals in email phishing. The amount of public and private data available is rising rapidly.

OSINT and SOCMINT technologies use sophisticated techniques and special tools to analyze the continually growing sources of information efficiently. This work aims to find descriptive information using the OSINT tools available online. The target will be achieved with the help of dummy accounts that would help understand the tools and evaluate further different tools. Also, find out what tools are commonly used and what improvements can be made to make them more descriptive for analysts.

Introduction

Millions of users worldwide are sharing, making connections, and exchanging daily life on earth, and our world has seen change quickly over the past 20 years. The amount of public data available is usually free and plentiful. This time can be defined as the "information age." However, people may argue that years of knowledge have helped society evolve into a digital age with its problems. The digital age has come to a particular danger to society. In contrast, digital methods are used to create ease of use for consumers, but the same thing happens to crime, terrorism, and other forms of cruel actors. It is a common misconception that the tactics used to fight crime are divided into categories, which must come from confidential sources. The fact is that today, organizations have collapsed more on open-source software. Information to use the Internet to create new strategies and methods for investigating crime, collecting information, and creating a relationship between data.

Background

OSINT can define a wide range of data collection, from the attack phase of penetration to data analysis for marketing purposes. However, action measures are the same for each model. Open-source intelligence (OSINT) involves collecting, processing, and associating public information from public data sources such as social media, social media, forums, and blogs, which are

publicly available government data, publications, or commercial data that can be accessed for a fee platform (Pastor-Galindo et al., 2020). OSINT is a performance-based approach to the target, and that target can be a person, an organization, or a group of people. Offers information about the target and the method of use of the advanced collection and analysis strategies, OSINT simultaneously enhances the knowledge of the target. The information obtained feeds the collection process so you can get closer to the goal. Over the past few years, social media has seen a record increase in non-active members of their media sharing posts and uploading daily activities.

Where we find the word SOCMINT is a combination of symbols OSINT and Web Mining strategies for various types of social media data to identify and understand the position of social media by expressing the behavior of people on the platform, which can lead to privacy troubles and consciously making rational choices to change the domain of social media to the favored state can be a threat to countrywide safety as we have seen in recent years in the form of Cambridge Analytica scandal. With the amount of information and data available in the public domain and collection techniques, things changed over time. While OSINT was based on newspaper collection data in the past, public talks and discussions are just a few examples. In contrast, today's data is open to the Internet, and data collection methods are becoming more and more advanced and accessible to all through the growth of open-source software development.

Methodology

The research will aim to find descriptive information using the online OSINT tools. The target will be achieved with the help of dummy accounts that would help understand the tools and evaluate all the different tools. Also, find out what tools are commonly used and what improvements can be made to make them more descriptive for analysts.

Research Strategy

A wide range of literature was reviewed. Despite its relevance to other research, this is an original study. The study is based on a thorough investigation of OSINT email addresses that can be used as a basic framework. As part of the investigation, investigative tools/software were applied to evaluate OSINT on

Email Addresses using real-life demonstration accounts/groups as appropriate.

Research Approach

This study began with concrete knowledge of OSINT and an observation of how email addresses can curb security challenges or enhance their other positive uses. In doing so, the context of OSINT was applied to real Email accounts.

Literature review

The section outlines the basic theory of OSINT, its intelligence cycle, related tools, and aspects of OSINT, focusing on Email and its benefits. Additionally, it examines intelligence in general and open-source intelligence (OSINT) in other categories.

Pastor-Galindo & Nespoli describe OSINT as a contemporary phenomenon. They examine it comprehensively, focusing primarily on the services and techniques enhancing cybersecurity. Firstly, the paper strives to analyze this system's strong points and suggest various ways it can be used in cybersecurity. On the other hand, it tends to cover the restrictions once adopted. Since there is a lot to be explored within this extensive field, there can be some open challenges that have to be addressed in the future. Moreover, it is also concerned with the role of OSINT within the public sphere, which represents an ideal environment to utilize open knowledge.

This paper by Michael Glassmana & JuKang introduces the concept of Open-Source Intelligence (OSINT) as an essential component of understanding human problem-solving in the 21st century. As a result of the emergence of the Internet and the growing dominance of the World Wide Web in daily life, many aspects of OSINT reflect the altered relationship between humans and information. This paper discusses changes in intelligence conceptions brought about by the Internet and the Web. This paper aims to explore how the Internet can be used to enhance and extend the use of fluid intelligence. Our paper proposes open-source processes and ethos as a model for a new form of intelligence.

Throughout this paper, Adel & Cusack discuss why OSINT is vital for intelligent forensic investigations and why the quality of retrieved information is crucial for conducting digital forensic analysis, which can provide crucial evidence against organized crime, fraud, and murders and even help trace terrorist activities. Since there is a focus on retrieving high-quality data from tools, determining which tools provide significant advantages in data analyses is more critical than ever. It can provide links to other case-related and unrelated databases.

What is open-source intelligence (OSINT)?

Open-source intelligence (OSINT) is a concept that refers to any publicly available data used to fulfill a specific need for information. Open-source intelligence (OSINT) collects and reviews freely accessible data from web media. Open-source intelligence (OSINT) is information from public sources like the Internet. However, the term isn't strictly limited to the Internet; it means all publicly available sources. The Conclusion from all the possible definitions is: In contrast to closed or private data, OSINT is extracted from publicly available information sources. It creates valuable insight and knowledge about the topics that Open-source intelligence addresses.

OSINT information Gathering Types

OSINT resources can be collected using Active, Passive, and Semi-Passive methods.

- **Active Collection**
The method of information gathering in which you are directly in contact with the target. Among the features of this process are harvesting technical data about the target IT infrastructure through open ports, vulnerability scanning through unpatched Windows systems, scanning server applications, etc. This technique can be risky since the target may be aware, as the system may leave traces that can be detected by Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS). The activated collection also includes going somewhere physically and talking to someone (social engineering attack on the target).
- **Passive Collection**
Gathering data with the quiet observation of data generated by the target is one of the most common forms of data collection. A few examples of this include shoulder surfing, eavesdropping, and studying maps. Many OSINT data are collected passively using publicly accessible resources and can be conducted remotely. This process can be performed anonymously through virtual machines, VPNs, and the Dark Web (TOR). When passively collecting data, finding the most reliable information is difficult. This results in a lack of deeper analysis.
- **Semi-Passive Collection**
The collection process that falls between active and passive is in this category. The target servers are sent a small amount of traffic to gain a general understanding of them through this data collection method. The traffic is designed to look like regular Internet traffic in order not to draw attention to your reconnaissance operation. Rather than profoundly analyzing the target's web resources, you conduct a light review without alarming the target.

OSINT can be categorized into different categories based on where the public data is located

- Internet:
- Academic Publications
- Geospatial data
- Corporate paper
- Media channels

Tool comparison for Email OSINT analysis

This section introduces and demonstrates different tools for email-related intelligence analysis from the platform using open-source information. Various workflows and usage examples will be shown for each of the tools. Platform tools range from simple queries, usually free of charge, to more extensive infrastructure solutions that allow multiple queries on a large data set.

Using machine learning and advanced filtering tactics, these solutions optimize both the process and the outcome. A substantial budget and team are needed to implement many solutions corporations and governments use. Due to this, access to this kind of solution is limited. There are, however, many excellent open-source tools that help users conduct powerful searches. This paper aims to identify the set of tools based on

the use cases and the principal utilities. Each tool shown in this thesis is freely available via the Internet.

List of used tools:

- Hunter
- Emailable
- Phantom Buster
- LeadFuze
- Email Harvester
- Simple Email Reputation
- Email Header Analyzer
- Google Admin Toolbox Message header

Hunter

Hunter is a web application looking for email finder and verification help running email campaigns. Using Hunter’s services, professionals can connect with people who matter. The application founders are François Grante and Antoine Fink.

Demonstration:

- Domain’s search
Domain search allows the user to look for relevant verified mail, and the pattern can be determined with the mail available in the public domain. This search provides a personal and generic view as well.
- Type filter
Show only personal or role-based email addresses.
- Email pattern
The most common email format used in your organization can be identified by selecting from dozens of combinations.
- Find someone
Enter the person’s name, and the user will be given their email address.
- Score & verification
Obtain a confidence score or list of verified email addresses.
- Save a lead
Any email the user receives from your leads can be exported or imported straight into your favorite CRM.
- Sources
Almost every email address has public sources that the user can check and the last discovery date.

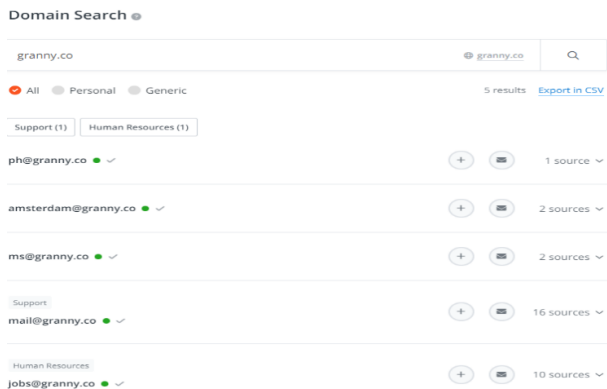


Figure 1. Screenshot of Hunter.io

- Finder
The Finder has two options to look for:
 1. Email Finder
 2. Author Finder

With the Email Finder, you can find a professional email address based on a name and a domain name. Using the data, we have about the given domain name, we can guess the email address from our database. Eventually, we verify the address or return the information with a confidence score.

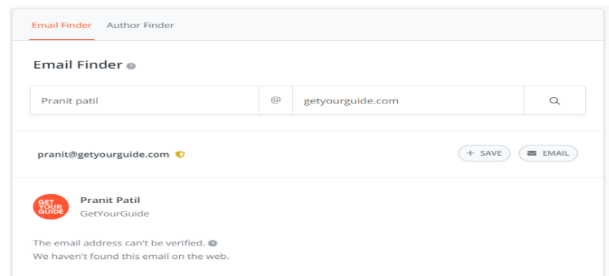


Figure 2. Screenshot of Hunter.io

- Author finder:
With the Author Finder, the user can find out who wrote an article and their professional email address. Based on the application’s information about the domain name of the given article, we can guess the email address or search for it in our base. Based on the result, the application provides confidence that it is a genuine email address.

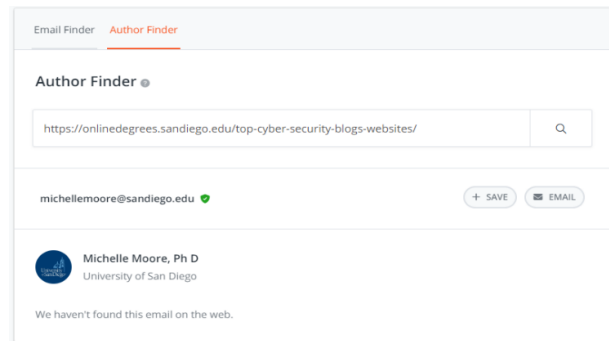


Figure 3. Screenshot of Hunter.io

- Verify
Using Email Verification, the application can check a recipient’s deliverability without emailing. It verifies the format, domain information, and responses from the mail servers to verify that an email address can be used. The example below is about a current employee and an old employee. The older employee is Invalid as it is disabled, while the current employee’s mail is valid. This section in the application provides information about the format, type, server status, and email status.
- Bulks
The bulks option in this application allows the user to perform various tasks simultaneously. The task includes domain search, email Finder, author Finder, and email verification.

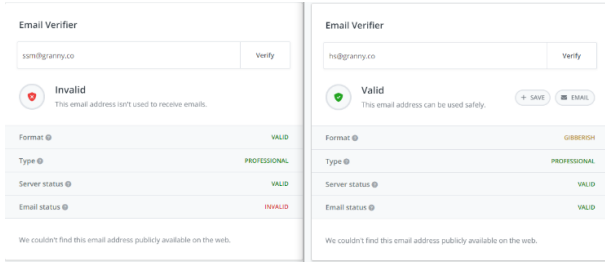


Figure 4. Screenshot of Hunter.io

- Campaign

The campaign features allow the application user to run a campaign for multiple mail IDs with options or follow-up and provide the performance chart of the whole campaign. The campaign feature allows the user to set up the content with an option of follow-up emails. The next step is setting up the audience or the target for the campaign. After the campaign starts, all the emails are within a period. The statistics show various stats about the number of emails sent, the number of emails opened, the number of clicks on the mail link, and the number of replies. The activity section shows notifications about the campaign, such as the update of mail opened, clicked, or replied to.

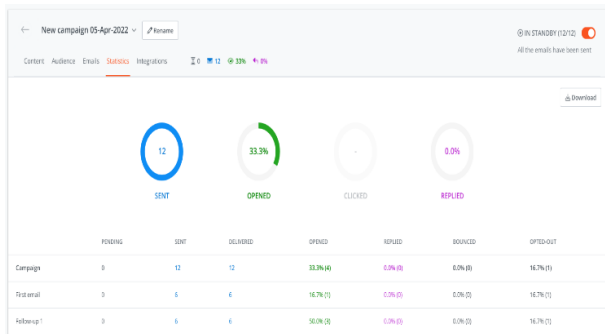


Figure 5. Screenshot of Hunter.io

- Add-ons

Hunter has add-on features like Chrome extension, Google Sheets add-on, mail tracker, and templates. All these features are easy to use. Chrome extension allows the user to get the email address of the website it is currently browsing. Steps to add Chrome extension:

1. Search Hunter’s Chrome Extension page.
2. Then click + Add to Chrome” button.
3. Click the Add extension button to confirm.

After the extension is installed, it allows one to find email addresses with a simple click while you are on the website.

1. Go to the website you want to get the email address for.
2. Click on the extension icon in your browser.

Using that method, you can find all the email addresses with the same domain (i.e., finishing with Talon.One) we found on the web, along with the URL where each address was found.

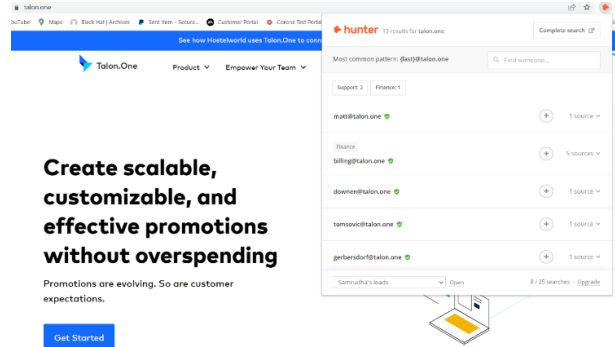


Figure 6. Screenshot of Hunter.io

The extension also provides information about the source of the mail address. The extension provides the most common pattern: {last}@talon.one. When available, Hunter’s Chrome extension will also return other information about the mail, such as:

1. Full name
2. Job Title
3. Telephone number
4. LinkedIn profile
5. Twitter profile

When the user uses this add-on feature for the 1st time, the add-on will ask you for your API key before connecting to your Hunter account.

The next is MailTracker by Hunter. This extension lets the user know whether the recipient has opened the mail. To add the email tracker, the user must add the extension to the Chrome browser and then sign in with the user’s Gmail. Currently, the feature works with Gmail only.



Figure 7. Screenshot of gmail.com

The feature also shows the number of times the recipient has opened the mail.-

- Limitation

You can perform several email verifications per month based on your monthly verification quota. A limit is set per domain: you can verify 200 email addresses per 24 hours from a single domain name. Bulk email verification allows for a maximum of 10,000 emails to be verified.

Emailable

Emailable is a Fully equipped email verification solution. Sending emails is not sufficient. They must be delivered. Deliverability and ROI for your email marketing campaigns will increase with one of the most affordable and reliable mail-checking services.

Demonstration:

The emailable web application has various categories: bulk, monitor, single, and API. The bulk category allows for the creation of a list of targeted emails. The list can be added to the bulk

by the computer using the copy-paste method or by connecting two third-party applications. Once the mail IDs have been added to the list, can we proceed with verification and other information?

The application provides a view of mail that is deliverable and undeliverable. The reason for mail being decided as undeliverable would be a non-existent email address, suspension of mail, or other reasons such as termination of an employee's mail ID. The other outputs from this application would be risk, duplication, and other unknown reasons. The reasons for the risky email category would be low quality or low deliverability. The unknown reasons would be no content, time out, unexpected error, and SMTP unavailability.

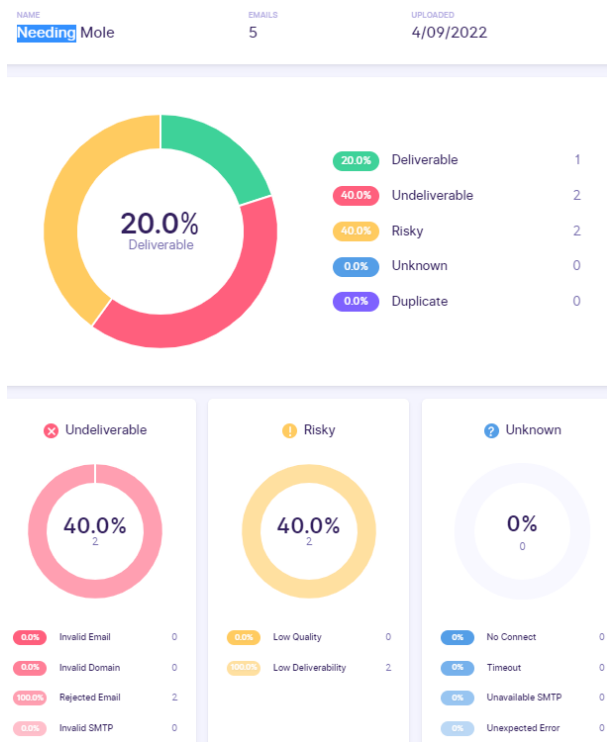


Figure 8. Screenshot of Emailable

The email IDs tested on this application were from various organizations. Two employees are currently working in the organization, while two employees have left the organization and a startup-based mail ID. The application also has an Email Verifier, which provides an excellent visual view of the verified mail. The application also allows the user to set up the monitored list, automatically allowing seamless email verification. Take responsibility for cleaning the email list. The monitoring list can be set up with the help of a simple integration using APIs for various kinds of applications like HubSpot, Intercom, Campaign Monitor, Active Campaign, Shopify, and various other applications.

The application also allows the setup of an API using JSON.

Limitation

The application shows a lot of false-positive results. The use of the application in its free version is limited. Most of the

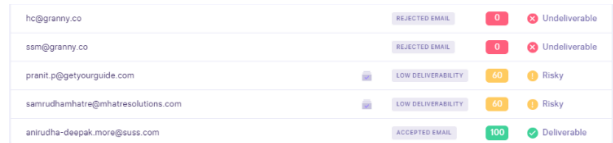


Figure 9. Screenshot of Emailable

features need to be enabled after the payment is done.

Phantom Buster

Phantom Busters is a data extraction solution designed to help marketing and sales teams across businesses of all sizes collect information on LinkedIn, Twitter, Instagram, Facebook, and other forums for convenient customer relationship management. The forum also enables administrators to automatically schedule and execute actions such as following profiles, liking posts, posting customized messages, receiving requests, and more to share hopes for increasing the visibility across the web and turn any webpage you know into the source of information. Phantom Buster will visit the webpage on your behalf and gather the information for you. Phantom can do any action on the web. It is efficient and works 24/7. The main goal of Phantom Buster is to quickly set up automatic and essential growth strategies for non-technical users and be creative.

Demonstration:

Creating a spreadsheet is the base for users of any Phantom Buster categories. The spreadsheet setup is essential.

In column A, enter a list of full names, one person per row. Having first names in one column and last names in another is also possible. After that, enter the names of the corresponding companies (or websites) for each person.

A column titled "name" should be used for the full name, while a column titled "company" should be used for company names.

Make this spreadsheet public so Phantom Buster can access it.

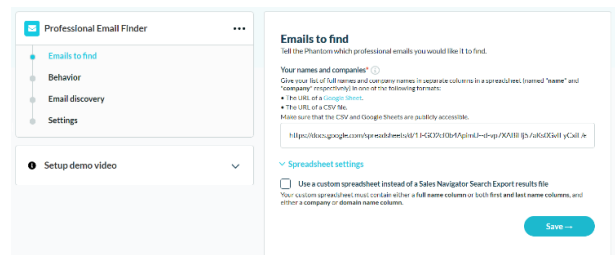


Figure 10. Screenshot of Phantom Buster

After saving the Emails to find. The following task behavior. Here, the user will enter the number of spreadsheet rows to process per launch, name your results file, and Fields to keep in the CSV file. The next step is Email discovery. There are various options for email discovery services from various services, such as Phantom Buster, Dropcontact.io, Hunter.io, and Snov.io. I have used the Phantom Buster service and clicked on save. In the settings tab, I used Launch Manually, and in notifications, I

selected in case of error option to get notified in case of error and clicked on save. After the launch process is complete, download the CSV file that contains the email address, qualification, phone number, job profile, etc. The output CSV contains the email address, qualifications, phone number, job profile, etc, that can be used further.

Limitation:

The web application has limited access during the trial period. The execution time is just 2 hours, while only five searches were possible.

LeadFuze

LeadFuze is a Lead Generation Software that provides advanced communication data. It uses Artificial Intelligence to find specific clues in specific fields and industries. They are used by sales, hiring teams, and marketing organizations. Key features offered by LeadFuze software include automated listing, reliable email access, and a focus on exciting prospects.

Get the contact details of any business professional. Get contact details and social media profile information such as Facebook, LinkedIn, Instagram, etc. Search all market segments or specific people or accounts. They are used by retailers, employers, and advertisers in marketing agencies, employees, IT, and start-ups.

Demonstration:

LeadFuze consists of different searches, such as market-based or account-based. Market search helps search the organization by accepting various filters like the industry in which the organization lies, Location, Number of employees, and year in which the organization was founded. Other factors include the monthly budget, the technology used, posts for which the organization is hiring, and News.

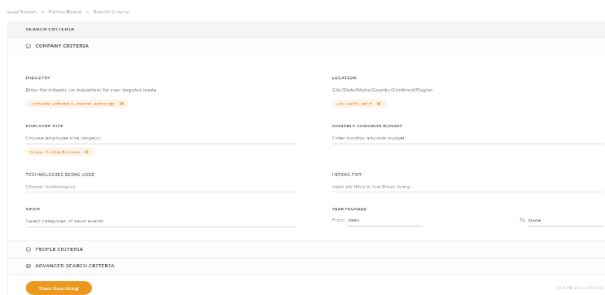


Figure 11. Screenshot of LeadFuze

A market-based search allows users to look for an extensive search area. They can choose from small businesses to large companies.

The output shows a range of people from the Computer Software and Internet Technology Industry in Berlin, Germany. The employee sizes defined are 500-1000, founded in 1995.

The application provides information about the person the user selects. Personal information like LinkedIn profile, years of experience work, skills, estimated salary, job history, time in the latest role, and interests. Company details like year found industry, employee size, the technology used, and information about the

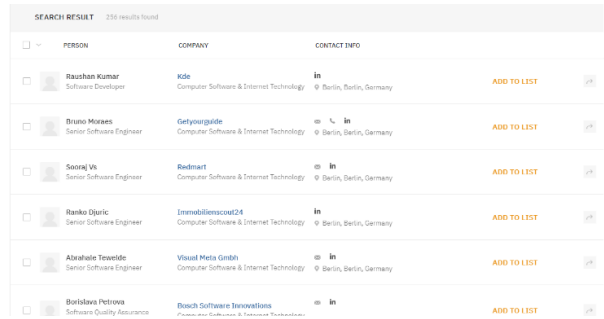


Figure 12. Screenshot of LeadFuze

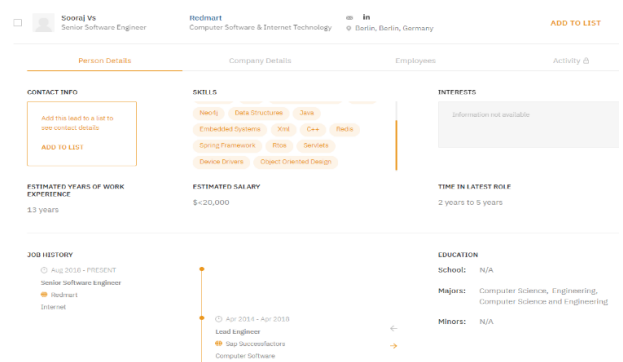


Figure 13. Screenshot of LeadFuze

role for which currently hiring. The employee’s section shows the current employee members of the organization.

The following important part is account-based search. The account-based search allows the user to look for people using their first name, last name, company name, and role. When searching using an account-based search, click on account-based search, add the first name, last name, company name, the role of the person, etc., and click on start searching.

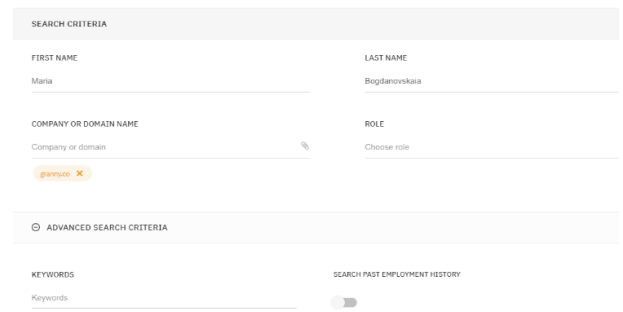


Figure 14. Screenshot of LeadFuze

It also offers filters such as required mail, contact number, mail address, etc., and one can check the employment status. Further, we can add the results to the lists to get detailed information.

Limitation:

LeadFuze is a little challenging to use initially. It is hard to

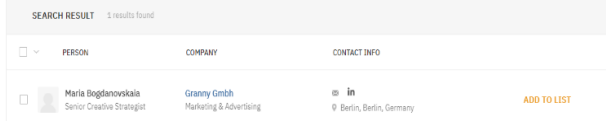


Figure 15. Screenshot of LeadFuze

find enough double-verified leads in the categories needed. When people leave jobs or change positions, it's not always updated. Some things are difficult to see, but getting answers quickly with live chat.

Email Harvester

In this package, you will find Email Harvester, a tool to retrieve Domain email addresses from search engines.

Email harvesting or scraping can obtain email address lists through various methods.

Demonstration

The EmailHarvester helps retrieve domain email addresses from search engines. It is a GitHub-based application and runs smoothly on Kali Linux and Parrot OS. This is a straightforward but very effective one in the early stages of a penetration test or when trying to determine the visibility of a company online and getting Started with Email Harvester.

The user needs to clone the application from GitHub. Before cloning the application, the essential requirement needs to be fulfilled. The basic requirements are Python 3.x, termcolor, colorama, requests, and validators. Most of the requirements are covered in Kali Linux. Most importantly, Python must be above version 3 after covering all the basic requirements. The next step is cloning the application from GitHub.

Cloning involves two basic steps:

```
git clone https://github.com/maldevel/EmailHarvester and
pip install -r requirements.txt
```



Figure 16. Screenshot of emailharvester

When the installation is done, the user can start using the tools with a simple command like the one below:

```
Search in Google
emailharvester -d getyourguide.com -e googleplus
Search the site using Search engines:
emailharvester -d getyourguide.com -e linkedin
emailharvester -d getyourguide.com -e twitter
emailharvester -d getyourguide.com -e googleplus
Search across all engines/sites:
emailharvester -d granny.co -e all
```

Output:

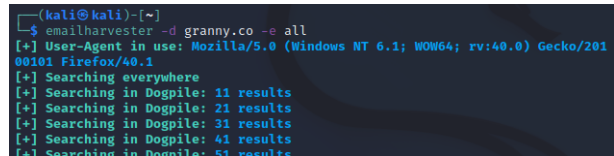


Figure 17. Screenshot of emailharvester

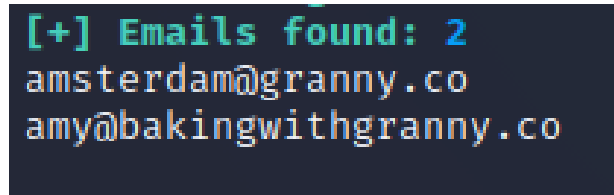


Figure 18. Screenshot of emailharvester

Search in all engines/sites but exclude some:
 emailharvester -d example.com -e all -r twitter,ask
 Limit results

```
emailharvester -d example.com -e all -l 200
```

Export emails

```
emailharvester -d example.com -e all -l 200 -s emails.txt
```

Proxy Server

```
emailharvester -d example.com -e all -x
http://127.0.0.1:8080
```

Limitation:

The result is not decisive, and the result varies. The output provided is limited using emailharvester.

Simple Email Reputation

EmailRep is an arrangement of crawlers, scanners, and advancement benefits that gather information on email locations, areas, and web personas. EmailRep utilizes many elements from online entertainment profiles, proficient systems administration destinations, dull web certification spills, information breaks, phishing units, phishing messages, spam records, open mail transfers, space age and notoriety, and deliverability, and more to foresee the gamble of an email address. There are hundreds of factors considered by EmailRep, such as the domain age, traffic rank, presence on social media, professional social networking sites, personal connections, public records, deliverability, dark web credentials leaks, phishing emails, and emails sent by threat actors.

Demonstration:

The free version of Simple Email Reputation allows users to test 250 monthly queries and up to 10 queries/day. There is a paid version, which is commercial and enterprise. Commercial allows users to email and have 1000 queries per month and no daily limit with Email.

Input: pranita.p@getyourguide.com

Output: High Reputation

Not suspicious. We have not sent any direct reference to this Email, but the sender's domain is highly reputable. The Email is

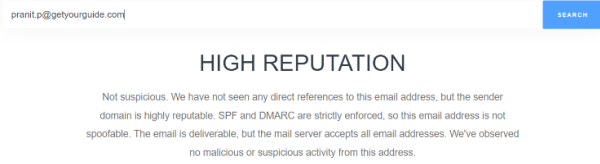


Figure 19. Screenshot of Simple Email Reputation

deliverable, but the mail server accepts all the Email addresses. No malicious or suspicious activities are found from this address.

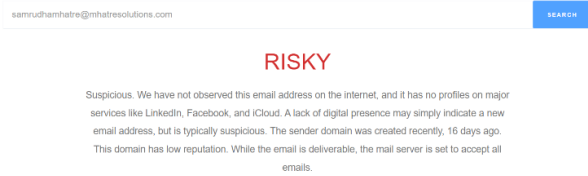


Figure 20. Screenshot of Simple Email Reputation

Input: samrudhamhatre@mhatresolutions.com

Output: Risky

Suspicious. The email address is not found on the Internet, and there is no primary profile on LinkedIn, Facebook, or other social media accounts. A lack of digital presence may indicate that this is new and can be suspicious mail. The domain has a low reputation. The email address is deliverable, and the mail server accepts all the emails.

Limitation:

The application does not provide a proper output and can be a false positive output.

E-Mail Header Analyzer

An email header can be checked and analyzed with this tool. Received lines are displayed separately, and the data is displayed. The tool is all about providing information extracted from the header of any mail. Information like Time Overview, Description, Received Details, Public IP Addresses, and Header Description (Recipient hostname, Sender hostname). The tool is available for free. The owner of the tool is Gaijin.

To use this tool, the user needs to get the header of the particular mail about which the user needs more information. The header needs to be pasted in the box and searched.

Demonstration:

Once the header is copied, paste it into the box and click enter.

After clicking enter, the output will have a systematically arranged header, time overview of the mail, description, and other information.

The time overview will provide information about the sender, receiver, and the cycle from where the mail has traveled to the end recipient. It will also show IP addresses.

The description will show a proper timeline for the mail.

The receiver details more information about the sender until it reaches its target. Here, a warning will be raised regarding any

Time Overview

The Received lines are sorted in reverse order (in the order they were entered).

Source	Sent from	Received by	Received time	Duration
Date:	Fri, 29 Apr 2022 06:00:30 +0000 (UTC)		2022-04-29 06:00:30 UTC	
Received:	[127.0.0.1] (unknown)	geopod-lsmtpd-4-0	2022-04-29 06:00:29 UTC	- 1s
Received:		filterrecv-599676dbf5-jrgmt	1970-01-01 00:00:00 UTC	
Received:	o1678914x67.outbound-mail.sendgrid.net. (167.89.14.67)	mx.google.com	2022-04-29 06:00:30 UTC	
Received:			2022-04-29 06:00:30 UTC	0s

Description

On 2022-04-29 at 06:00:30 UTC, the message was sent by the sender's mail program (Nodemailer (0.3.42-8; +http://www.nodemailer.com/)).
 On 2022-04-29 at 06:00:29 UTC, the message was sent from the computer "[127.0.0.1] (unknown)" to the mail server "".
 On 1970-01-01 at 00:00:00 UTC, the message was sent from the computer "" to the mail server "".
 On 2022-04-29 at 06:00:30 UTC, the message was sent from the computer "o1678914x67.outbound-mail.sendgrid.net." (167.89.14.67) to the mail server "mx.google.com".
 On 2022-04-29 at 06:00:30 UTC, the message was sent from the computer "" to the mail server "".
 The mail server of the recipient was "mx.google.com". This usually belongs to the recipient's provider or company.

Figure 21. Screenshot of Email Header Analyzer

suspicious sender.

The next important thing that can be useful is IP addresses.

The application also provides information regarding the header. In this section, the output of each header description includes both a description and a formatted and decoded header, if available.

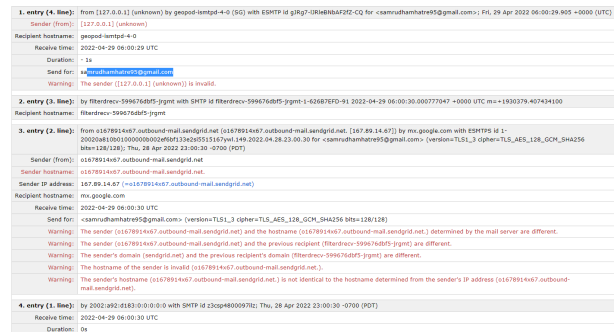


Figure 22. Screenshot of Email Header Analyzer

Information like delivered to Return-Path, X-Google-SMTP-Source, Authentication-Results, message-ID, DKIM-Signature with an explanation for each information shown.

This tool helps to get a person's IP address. This can be a helpful thing in case of fraudulent activities. This application can help police keep track of information regarding any suspicious activities.

Limitation:

The application is free, and there is no control over who will use this service. IP addresses failing into the wrong hands can be dangerous. That login page and security are essential for this service to be used.

Google Admin Toolbox Message header

This is another Email Header checker. These tools provide more accurate information regarding the mail using its header. The Gmail headers are crucial in revealing sensitive information about the sender and other aspects of your network. Thus, one will likely find sensitive information if one carefully analyses Gmail headers. The Gmail header contains the following elements:

Delivered-to: This field indicates the email address of the recipient. It usually contains the same email ID used to analyze

Gmail headers.

The email header indicates which SMTP server the message was received from as indicated by the "Received By" element:

1. Server's IP address
2. SMTP id of the server visited
3. Time and date at which SMTP received a message

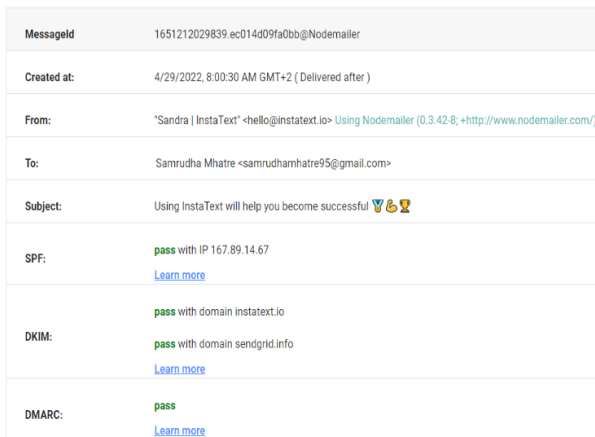
Email addresses include a header field called X-Received to indicate the presence of non-standard headers. Among its contents are:

1. In the case of a message received by a server, the server's IP address,
2. contains the server's SMTP email,
3. contains the time and date when the Email was received.

Signature DKIM: The DKIM signature header contains a file email containing the digital signature embedded in the Email. The mail server maintains another authentication key to allow data sharing with secure encryption.

Demonstration:

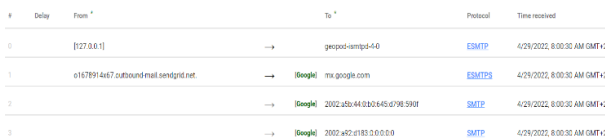
The Google admin toolbox helps to get information from the header. Paste the header acquired from through Gmail.



MessageId	1651212029839.ec014d09fa0bb@Nodemailer
Created at	4/29/2022, 8:00:30 AM GMT+2 (Delivered after)
From	"Sandra InstaText" <hello@instatext.io> Using Nodemailer (0.3.42.8; +http://www.nodemailer.com/)
To	Samrudha Mhatre <samrudhamhatre95@gmail.com>
Subject	Using InstaText will help you become successful 🙌🏆
SPF	pass with IP 167.89.14.67 Learn more
DKIM	pass with domain instatext.io pass with domain sendgrid.info Learn more
DMARC	pass Learn more

Figure 23. Screenshot of Google Admin Toolbox Message header

The output in the figure below will have information about the message held, created at, from, to, subject, and DKIM.



#	Delay	From	To	Protocol	Time received
0	[027.0.0.1]		google-remq4-4-0	ESMTP	4/29/2022, 8:00:30 AM GMT+2
1	01878914817.outboundmail.sendgrid.net		[Google] mx.google.com	ESMTPS	4/29/2022, 8:00:30 AM GMT+2
2			[Google] 2002.450:44:0:0:0:0:0:0	SMTP	4/29/2022, 8:00:30 AM GMT+2
3			[Google] 2002.492:0183:0:0:0:0	SMTP	4/29/2022, 8:00:30 AM GMT+2

Figure 24. Screenshot of Google Admin Toolbox Message header

The other part of the output shows the entire timeline from the destination to the recipient. The timestamp is from when it is received and the protocol.

Findings

Based on statistical analysis, we found that each tool produces a different result, has different usability, and, most importantly, has a different benefit. The following information has been summarized in the following table after carefully comparing the tools, how they operate, and how widely they are used.

The Findings from the analysis of these tools are attached in the table below.

Use Cases

The focus of the study in this section is to design the use case for some tools when it comes to Email addresses, which allowed the study to select specific tools to achieve results for different scenarios in the study. Intelligence can be found by emailing sentiment, emails, and sentiment analysis. As Email is one of the best platforms for searching for people and gaining that person's information to emails form it into intelligence, the information that Email can provide can be more accurate since most of the target organization's email addresses are available. Most ransomware attacks take place through mail when targeted towards a specific target. In this case study, we will try to get information about suspicious/Ransomware mail. Therefore, to perform a thorough analysis on which further investigation can be based, it is necessary to collect relevant data for the analysis, which creates a problem in understanding the means and scope of the research; it is also necessary to establish a foundation of facts.

Problem description:

The use cases are an example of Blackmailing mail and Lottery/Fund Mail. There is an email received on the target mail addresses. The mail has blackmailed the person to reveal the secret of the business accessed by the sender. The sender has demanded a Bitcoin to solve this issue. Another example is a person trying to send money to the target by making an emotional and nationality-based connection. The solution here would provide how an investigator can use your email analyzer and other tools to get information about all the relevant parties, like the organization whose server has been used. All this analysis will be done using the email header. Email headers serve as passports for your messages. Each email server it encounters inserts entries into the header along the way. Therefore, the longer the header, the more servers route the Email. The ransomware threat is not possible using Google Mail.

•Blackmailing mails

A lot of information is contained in email headers. Email readers will only see the subject, the sender's Email, and other information.

Here is an email that can help the investigator find the sender's information, like IP address, server information, etc. The investigator will use the mail header to collect more information about the sender. The header can be accessed by clicking on the show original option in the Gmail account.

The user will now copy the header from the mail. This header will be pasted into the Email Header Analyzer or Gmail Email header tool. Now, the header will be pasted into the analyzing tools.

The header will then be analyzed using the Google Gmail header tools. To investigate suspicious mail, it is essential to know

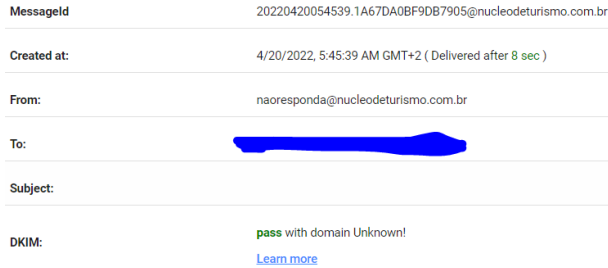


Figure 25. Screenshot of Google Admin Toolbox Message header (Use case)

who, what, and when it is. The analyzer provides this information. The output will provide a timeline of the mail travel from the sender server to the recipient. The image (Figure 25) shows information about the server that starts the mail until it reaches the recipient. This Email is created when the first email server receives the Email from the sender's computer. This entry includes information about the server hosting the email web application if the client is web-based.

#	Delay	From *	To *	Protocol	Time received
1	4 sec	vps-74188.fhnet.fr	srv1.primesaas.com.br		4/20/2022, 5:45:35 AM GMT+2
2	8 sec	ec2-18-228-133-124.sa-east-1.compute.amazonaws.com	c1051.mx.srv.dfn.de	ESMTPS	4/20/2022, 5:45:43 AM GMT+2
3	1 sec	c1051.mx.srv.dfn.de	localhost	ESMTP	4/20/2022, 5:45:44 AM GMT+2
4	1 sec	localhost	c1051.mx.srv.dfn.de	ESMTP	4/20/2022, 5:45:45 AM GMT+2
5	2 sec	c1051.mx.srv.dfn.de	tiu.fh-brandenburg.de	ESMTPS	4/20/2022, 5:45:47 AM GMT+2
6		tiu.fh-brandenburg.de	z52-rz.fh-brandenburg.de	ESMTPS	4/20/2022, 5:45:47 AM GMT+2
7		z52-rz.fh-brandenburg.de	localhost	ESMTP	4/20/2022, 5:45:47 AM GMT+2
8		localhost	z52-rz.fh-brandenburg.de	ESMTPS	4/20/2022, 5:45:47 AM GMT+2
9		z52-rz.fh-brandenburg.de	z52-rz.fh-brandenburg.de	ESMTPS	4/20/2022, 5:45:47 AM GMT+2

Figure 26. Screenshot of Google Admin Toolbox Message header (Use case)

The sender or the attacker when is sending spam messages to the target. The attacker will send multiple spam messages. All this will be done from a rented server. When the server is rented, the attacker requires a provider already in the network. For this very reason, the attacker scams the provider. For this use case, the sender or attacker scammed the software-as-a-service provider (srv1.primesaas.com.br). The attacker will use a fake credit card or phishing for the admin Gmail and mail account.

The important thing here is that the attacker has rented the server, and this server is essential. In this use case, the server is rented from France. Later, they hacked the service provider from Brazil (srv1.primesaas.com.br). Before starting the technical analysis, it is essential to inform the service provider (nucleodeturismo.com.br) that their mail server was hacked. Here, the investigator can contact the service provider by visiting the domain and looking for ways to contact them. This is to inform them about their mail account being hacked so that they take all the preventive measures.

For more information, the investigator can use the Email Header Analyzer, which will help them gain information like the IP addresses and timeline of the mail.

The primary investigation starts here; the attacker here is vps-74188.fhnet.fr. VPS is a virtual private server that is used

Time Overview

The Received lines are sorted in reverse order (in the order they were entered).

Source	Sent from	Received by	Received time	Duration
Date:	Date: 20 Apr 2022 05:45:39 +0200		2022-04-20 03:45:39 UTC	- 4s
Received:	vps-74188.fhnet.fr ([185.13.37.254]:63194)	srv1.primesaas.com.br	2022-04-20 03:45:35 UTC	8s
Received:	ec2-18-228-133-124.sa-east-1.compute.amazonaws.com (18.228.133.124)	c1051.mx.srv.dfn.de	2022-04-20 03:45:43 UTC	1s
Received:	c1051.mx.srv.dfn.de (127.0.0.1)	localhost	2022-04-20 03:45:44 UTC	2s
Received:	localhost (127.0.0.1)	c1051.mx.srv.dfn.de	2022-04-20 03:45:45 UTC	5s
Received:	c1051.mx.srv.dfn.de (194.95.238.38)	tiu.fh-brandenburg.de	2022-04-20 03:45:47 UTC	0s
Received:	tiu.fh-brandenburg.de (195.37.0.8)	z52-rz.fh-brandenburg.de	2022-04-20 03:45:47 UTC	0s
Received:	z52-rz.fh-brandenburg.de (127.0.0.1)	localhost	2022-04-20 03:45:47 UTC	0s
Received:	localhost (127.0.0.1)	z52-rz.fh-brandenburg.de	2022-04-20 03:45:47 UTC	0s
Received:	localhost (127.0.0.1)	z52-rz.fh-brandenburg.de	2022-04-20 03:45:47 UTC	0s
Received:	LHL0.z52-rz.fh-brandenburg.de	z52-rz.fh-brandenburg.de	2022-04-20 03:45:47 UTC	0s

Figure 27. Screenshot of Email Header Analyzer (Use case)

for this suspicious activity. For more information regarding the server, the investigator can use Passive Total's website.

Passive Total gathers data from the entire web, extracts intelligence to identify threats and attacker infrastructure, and uses machine learning to scale threat hunting and response. Using Passive Total, you get contextual information on who is attacking you, their tools, and their systems, as well as indicators of compromise from inside and outside the firewall. The application provides features like Rapid Threat Investigation and Scale Threat Hunting Automate Response.

Here, the investigator needs to paste the acquired address. The output received provides information like a heatmap (Informing the duration of time the server has been used)

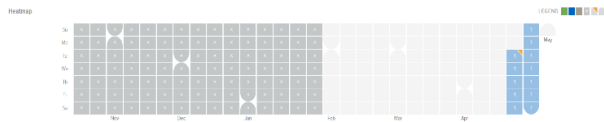


Figure 28. Screenshot of Passive Total (Use case)

The website provides resolutions, WHOIS, certificates, sub-domains, trackers, components, and other information.

Resolution ID	Resolution	Location	Network	ASN	First	Last	Source	Type
1	185.13.37.254	FR	185.13.37.0/24	197932	2022-04-19	2022-04-30	envy,ipjoy	Dynamic

Figure 29. Screenshot of Passive Total (Use case)

The resolution shows the IP addresses, location, network, and first and last seen.

As the last seen is April 30, 2022, the addresses still share spam mail. This also means that the attacker has not used any hacked credit card. This is because the provider will recognize the hacked credit card.

The next part is getting to know who owned the domain. The WHOIS section here will help to get information regarding the provider. Here, the information will help get a handle on someone responsible for the whole ransomware issue.

The WHOIS section provides information like Registrar and Email. Status of the domain, name (Not visible to the average user but will be visible to police), organization, and other information. Now, the police or the investigator can contact the domain owner using the information gained through a mail address like tech@ovh.net so that they can take the required measures against the possible misuse of their domain.

Record Updated 2022-01-11 : Last Scanned 2022-04-29
Checked by RiskIQ | Expires in 8 months | Created 14 years ago | [Hide Diff](#) | [Hide Raw Record](#)

Attribute	Value
WHOIS Server	whois.nic.fr
Registrar	OVH
Domain Status	ACTIVE
Email	tcxt9gk4jwkut5wx4wvx@i.o-w-o.info (admin) tech@ovh.net (tech)
Name	Ano Nymous (registrant)
Organization	Techcrea Solutions SARL (admin) OVH NET (tech)
Street	-
City	-
State	-
Postal Code	-
Country	-
Phone	33 8 99 70 17 61 (tech) 33972125539 (admin)
NameServers	dns1.firsttheberg.com dns2.firsttheberg.com

Figure 30. Screenshot of Passive Total (Use case)

The name here is Anonymous, as due to GDPR in the EU, data privacy/ data protection is critical. It will always be anonymous for average users, but for police, the name will be visible.

The police can now use the IP address and contact the domain provider to get information about the payment method or details to get hold of the attacker involved in the phishing or ransomware.

•Lottery Mails

Here is a mail about someone sending money to the target using the nation link. The Next use case is about a user receiving mail about money stored in the United States of America. The scamster here is looking for a target who is greedy by offering 3500000 dollars.

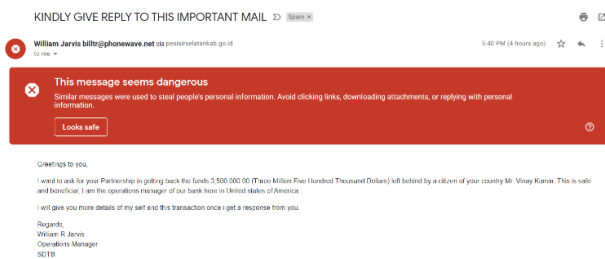


Figure 31. Screenshot of Gmail Inbox (Use case)

Now, use the same tools as the Email Header Analyzer or Google Email Header tool to get more information about the scamster. Using the Gmail Email header tool, we get to know the domain as previously demonstrated. Now, use the header extracted from the source in the Gmail account. We use the header in the Google email header analyzer.

Here, we found SPF showing a soft fail with IP Unknown.

MessageId	20220504154045.476D6EE05E2F45D5@phonewave.net
Created at:	5/4/2022, 5:40:45 PM GMT+2 (Delivered after 3 sec)
From:	William Jarvis <billtr@phonewave.net>
To:	[REDACTED]
Subject:	KINDLY GIVE REPLY TO THIS IMPORTANT MAIL
SPF:	softfail with IP Unknown! Learn more
DKIM:	pass with domain pesisirselatankab.go.id Learn more

Figure 32. Screenshot of Google Admin Toolbox Message header (Use case)

An IP address may or may not be allowed to be sent from that domain if a soft fail occurs. Despite being marked as suspicious, the mailbox provider will nonetheless accept the message. Mailbox providers use other data points to make a filtering decision, so a soft fail is not always the source of deliverability problems.

#	Delay	From	To	Protocol	Time received
1	0:00:20.39210	ps.pesisirselatankab.go.id	ps.pesisirselatankab.go.id	SMTP	5/5/2022, 5:40:45 PM GMT+2
2	3:sec	vps.pesisirselatankab.go.id	smtp.google.com	SMTP	5/4/2022, 5:40:48 PM GMT+2
3			smtp.google.com	SMTP	5/4/2022, 5:40:48 PM GMT+2
4			smtp.google.com	SMTP	5/4/2022, 5:40:48 PM GMT+2

Figure 33. Screenshot of Google Admin Toolbox Message header (Use case)

The next task is to look for the person responsible for the whole mail. Using the Google Email header tool, we found vps.pesisirselatankab.go.id as the real culprit. The common thing in both use cases is VPS, a virtual private server.

The domain here belongs to Indonesia, which is suspicious for a person in the United States of America. Now, using Passive Total, we can find more information.

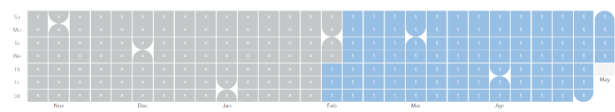


Figure 34. Screenshot of Passive Total (Use case)

The heat map we got from the Passive Total shows the domain has been online since February 3 and is still working. The domain is still being used to send spam messages.

Resolve	Location	Network	ASN	Port	Last	Status	Tags
103.143.153.50	ID	103.143.153.0/24	139402	2025.132.06	2022-05-04	HTTP, HTTPS	Personal, Anonymous, Possible, Sensitive, Private

Figure 35. Screenshot of Passive Total (Use case)

In the resolution session, we found the network the domain belongs to with the IP addresses. The next important thing is the WHOIS section. Here, the information gathered is much less than in the previous example. This shows that data gathered using various tools depend on various factors like the location of the target and the scamster with the domain location.

Attribute	Value
WHOIS Server	rdap.pandi.id
Registrar	pandi.id
Domain Status	client transfer prohibited server transfer prohibited
Email	-
Name	-
Organization	-
Street	-
City	-
State	-
Postal Code	-
Country	-
Phone	-
NameServers	ns1.pesisirselatankab.go.id ns2.pesisirselatankab.go.id

Figure 36. Screenshot of Passive Total (Use case)

The previous use case had direct information about Name, Organization, Domain status, and Email. In this use case, the domain status is prohibited, and the domain Email is shared.

In the previous case study, we found the Email and the organization's name. The Who Is section has another part where we can find the location as Jakarta, which is shown.

The information gathered here is mail addresses with domain pandi.id and organization name, the Ministry of Communication and Information Technology in Indonesia. As we find the domain name, we can inform the domain provider about their server being misused for scamming people. The domain found here is a domain provider in Indonesia. Thus, the investigator can inform the domain handler that their domain has been compromised and used to scam people.

•Mail for Money transfer

The subsequent use case is about identifying similar phishing mail. The mail received is about sharing the wealth of a minister in Guinea Bissau. The mail uses a respectable and well-known person to make it look legitimate. The mail here is about sharing 30% of \$60 Million. This is a phishing, where the user needs to request funds. Making the user share the bank details as the user falls into the trap.

Now we will try to investigate this mail. Using Google Email Using the Gmail Email header tool, we get to know the domain as previously demonstrated. Now, use the header extracted from the source in the Gmail account. We use the header in the Google

```

REGISTRAR
VCard:
  Organization: Kementerian Komunikasi dan Informatika
  TYPE: work
  TYPE: voice
  Address:
    Extended: Jl. Medan Merdeka Barat No. 9
    Street:
    Locale: Jakarta Pusat
    Region: Jakarta
    Postal Code: 10110
    Country: ID
    Email: hostmaster@pandi.id
    TYPE: work
Remark:
  Title: Whois Service
  Description: The port-43 whois service for this TLD is whois.id
Remark:
  Title: Remark
  Description: Registration information: https://pandi.id/?lang=en
Remark:
  Title: Source
  Description: PANDI
Link:
  Rel: related
  Value: https://about.rdap.org
  Type: About RDAP
  Href: https://about.rdap.org

```

Figure 37. Screenshot of Passive Total (Use case)

```

Dear Sir/Madam,
my name is Joseph Camarah Vieira, i am from Guinea Bissau, my late father was the former minis
ter of minee in my country Guinea Bissau, he was short dead by the rebels in my country, before
his death he deposited $60 million Dollars with Global Trust Security Company Accra Ghana, i wa
nt you to help me receive this money in your country for investment in your country i will give
you 30% of the total sum when the funds arrive your country.

Regards,
Mr Joseph Camarah Vieira
00233 244 617 863
my email:carrr444@yahoo.com

```

Figure 38. Screenshot of Gmail Inbox (Use case)

email header analyzer.

```

MessageId      201301172333.r0HNKZS028539@mail.shako.com.tw
Created at:    1/18/2013, 12:46:07 AM GMT+1 ( Delivered after -11 mins )
From:         JOSEPH CAMARAH VIEIRA <vieira@aol.com> Using Microsoft Outlook Express 6.00.2600.0000
To:          [REDACTED]
Subject:      [Spam-Mail] Dear Sir/Madam. (This message should be blocked: ctdos35128)

```

Figure 39. Screenshot of Google Admin Toolbox Message header (Use case)

The header will then be analyzed using the Google Gmail header tools. There is an odd 11-minute delay at the beginning, which may indicate an overloaded spam-sending server. There is a possibility of false positives due to time differences between servers. To investigate suspicious mail, it is essential to know who, what, and when it is. The analyzer provides this information.

The output will provide a timeline of the mail travel from the sender server to the recipient. Let's take a closer look at this entry. In the Emails entry, the from part indicates the source of the Email for this leg: User. After the email origin, you're taken to mail.shako.com.tw. Whenever an email server encounters this

Delay	From *	To *	Protocol	Time received
+12 mins	User	→ mail.shako.com.tw		1/18/2013, 12:33:38 AM GMT+1
34 sec	59-125-100-112.HNET.FINET.net	→ bf.shako.com.tw		1/18/2013, 12:34:12 AM GMT+1
76 sec	bf.shako.com.tw	→ TKCE5H4HS007.bigfish.com		1/18/2013, 12:35:28 AM GMT+1
3 sec	unirovn	→ mail240x42.bigfish.com	ESMTP	1/18/2013, 12:35:31 AM GMT+1
2 sec	localhost	→ mail240x42.bigfish.com	ESMTP	1/18/2013, 12:35:33 AM GMT+1

Figure 40. Screenshot of Google Admin Toolbox Message header (Use case)

header entry, it adds another one below it. There is a high likelihood that this email server is under the control of a malicious sender. This information should not be trusted. It is still worth investigating. We should try to determine the location of the email server. After seeing the delay and the domain provider, the sender's mail address veria@aol.com is fake. Thus, the investigator can inform aol.com about a possible breach in their domain.

For more information regarding the server, the investigator can use Passive Total's website. The heat map shows that the domain mail.shako.com.tw has been live since February 1, 2022, and is still alive.

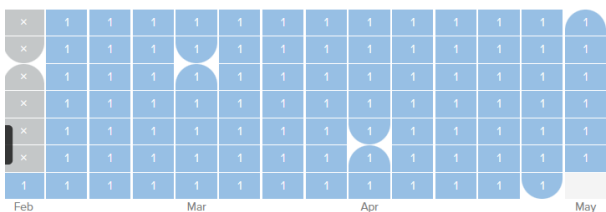


Figure 41. Screenshot of Passive Total (Use case)

The investigator can look into the resolution and WHOIS sections to get more information about the organization's IP, Email, and name.

Attribute	Value
WHOIS Server	whois.taiwan.net
Registrar	
Domain Status	ok
Email	betty@shako.com.tw (pg@shakt, admin, tech)
Name	SHAKO CO., LTD. Betty Lee (pg@shakt, admin, tech)
Organization	
Street	
City	
State	
Postal Code	
Country	
Phone	8869000000 (pg@shakt, admin, tech)
NameServers	dns1.manufacture.com.tw dns2.manufacture.com.tw

Figure 42. Screenshot of Passive Total (Use case)

Using the tool on the domain mail.shako.com.tw, we found that the server is based in Taiwan. A genuine person requesting help would instead use a legitimate server from the country in which they are located. The person won't use a server in Taiwan to get help.

The information collected until now is the location of the server and the delay time in sending the mail. This factor is

enough to understand that the mail is phishing and thus can be reported. Using this tool, companies can set up filters accordingly and thus block or mark the mail as spam to prevent companies' employees from falling into the trap.

Result Analysis, Conclusion, and Future Scope

Analysis

Every OSINT tool behaves and approaches situations differently depending on the scenario, which changes the tools and settings every time. Section V offers a complete breakdown of each OSINT tool and its primary findings. Each tool is explained and demonstrated as the limitation of these tools. Depending on the use cases associated with intelligence gathering and analysis, these tools can be helpful individually and together. An author has created a table to find a solution to this problem, where different tools are shown without undermining the effectiveness of the other tools. Based on their effectiveness, the ten tools listed in Appendix A have been selected from 115.

Each tool provides different types of information. Typically, OSINT tools for Email have the primary function of obtaining the Email address of a targeted individual from an organization. Table VI illustrates that many tools have different features, all of which are considered while analyzing tools. The best tool selection relies on which works in most cases when analyzing large data sets. The author found the tool hunter to have the most required features to work with different methods and with more accuracy than others. These tools have many good features like Domain Search, Email Finder, and Verifier. There is also an option for Bulk tasks where the user can run or perform all the features with multiple inputs. The Email Analyzer is an effective tool that police can use to investigate fraud mail to catch such culprits or for companies to set up filters to prevent employees from getting phished.

Conclusion and limitations

This study suggested that OSINT tools could be a good source of information about mail addresses from any organization during a demonstration and analysis of the OSINT tool for email addresses. People tend to keep a lot of data in their email addresses that can be analyzed in many ways and based on different data models, which are significant resources in data collection, mining, etc. The data can be the location of the person, the history of the person, and other things. When used as digital footprints, data can be analyzed simultaneously across various targets with fewer resources than traditional data gathering, processing, and analysis methods. Automating the analysis of big data sets with a set of tools is very important. The analyst selects the toolset from the list of use cases and scenarios, even though the author has found a list of tools with these features. In analyzing the toolset, researchers found that users can access only public data via these tools, regardless of the location from which they access the data. If the researcher has significant funds to spare for the task, it is possible to consider premium plans available after the basic plan. Most tools offer data export functions as well. Once an Email address is found or identified, various hacking techniques like Phishing (Email spoofing), spear phishing, and dictionary attacks are used. Thus, the account or mail address is subject to a

possible attack. These tools can be used for phishing campaigns as part of security awareness training. It is not difficult to identify a phishing email. Emails can be automatically rejected or quarantined by SPF, DKIM & DMARC. Before enabling any form of email security, organizations should think deeper.

Future work

An improvement in OSINT-based investigation of email addresses requires further research. When email addresses are inputted, optimizing and developing current tools is essential to get more information from them. Additionally, a tool with higher accuracy is required. However, there are a limited number of tools available now. Thus, more tools are required. There should also be tools that can break apart closed groups. As part of the intention to expand the tool collection beyond the current collection, it may be possible to analyze and demonstrate other tools in the future. Comparing the premium features of open-source tools in this study would be very interesting, primarily since the study focuses on open-source tools. Several tools were demonstrated in this study with premium features and capabilities that could be extremely valuable for users who need a set of tools of the highest quality.

Acknowledgments

This work was supported partially by the European Union in the framework of ERASMUS MUNDUS, Project CyberMACS (Project #101082683) (<https://cybermacs.eu>).

References

[1] Abdallah Qusef and Hamzeh Alkilani, "The effect of ISO/IEC 27001 standard over open-source intelligence." (2022), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8771761/>

[2] ESTEBAN BORGES, "What is OSINT? How can I make use of it?" (2021), <https://securitytrails.com/blog/what-is-osint-how-can-i-make-use-of-it>

[3] INTELLIGENCE COLLECTION ACTIVITIES AND DISCIPLINES <https://irp.fas.org/nsa/ioss/threat96/part02.htm#:~:text=These%20disciplines%20include%20human%20intelligence,United%20States%20to%20some%20degree>

[4] KRIPA THAPA, "Open Source Intelligence Gathering (OSINT)", <https://medium.com/infosec/open-source-intelligence-gatheringosint-f170973ec000>

[5] Nihad Hassan, "An Introduction To Open-Source Intelligence (OSINT) Gathering", <https://www.secjuice.com/introduction-to-open-source-intelligence-osint/>,

[6] "Intelligence Cycle and Process", <https://www.e-education.psu.edu/sgam/node/15>,

[7] Alec Smith & Steve Cook, "A Guide To Open Source Intelligence (OSINT)", <https://www.strategicstudyindia.com/2023/05/a-guide-to-open-source-intelligence.html#:~:text=OSINT%20is%20the%20collection%20and,%2C%20academic%20journals%2C%20public%20events.>

[8] Kali Linux, <https://www.kali.org/tools/emailharvester/>

[9] Hatice Ozsahan, "Top 15 Email Finder Tools: Pros and Cons + Reviews for 2022", <https://popupsmart.com/blog/email-finder-tools>,

[10] A. Adel, B. Cusack, "INVESTIGATIONS: OPEN-SOURCE INTELLIGENCE INVESTIGATION ANALYSIS" <https://www.semanticscholar.org/paper/INVESTIGATIONS3A-OPEN-SOURCE-INTELLIGENCE-ANALYSIS-Adel-Cusack/64e9d103a1d1a3a53b840e3f54d23d77982f08f1>

[11] Javier Pastor-Galindo, Pantaleone Nespoli, Félix Gómez Mármol, Gregorio Martínez Pérez, "The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends," <https://ieeexplore.ieee.org/document/8954668>,

[12] Michael Glassmana, Min JuKang, "Intelligence in the Internet age: The emergence and evolution of Open Source Intelligence (OSINT)", (2012), <https://www.sciencedirect.com/science/article/abs/pii/S0747563211002585>,

[13] Schwarz, Klaus; Franziska Schwarz, Reiner Creutzburg: "Conception and implementation of professional laboratory exercises in the field of open source intelligence (OSINT)". Proceed. Electronic Imaging Symposium 2020 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2020, <https://doi.org/10.2352/ISSN.2470-1173.2020.3.MOBMU-278>, (Last access: Nov. 22, 2022).

[14] Schwarz, Klaus; Reiner Creutzburg: "Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools - Part 1: RiskIQ PassiveTotal". Proceed. Electronic Imaging Symposium 2021 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2021, <https://doi.org/10.2352/ISSN.2470-1173.2021.3.MOBMU-043>, Last access: Nov. 22, 2022).

[15] Schwarz, Klaus; Reiner Creutzburg: "Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools - Part 2: Censys". Proceed. Electronic Imaging Symposium 2021 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2021, <https://doi.org/10.2352/ISSN.2470-1173.2021.3.MOBMU-044>, Last access: Nov. 22, 2022.

[16] Schwarz, Klaus; Reiner Creutzburg: "Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools - Part 3: Maltgo". Proceed. Electronic Imaging Symposium 2021 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2021, <https://doi.org/10.2352/ISSN.2470-1173.2021.3.MOBMU-045>, Last access: Nov. 22, 2022.

[17] Kant, Daniel; Reiner Creutzburg: "Investigation of risks for Critical Infrastructures due to the exposure of SCADA

systems and industrial controls on the Internet based on the search engine Shodan”. Proceed. Electronic Imaging Symposium 2020 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2020 <https://doi.org/10.2352/ISSN.2470-1173.2020.3.MOBMU-253>, Last access: Nov. 22, 2022.

- [18] M. S. Wong, N. Hideki and N. Yasuyuki, “The Incorporation of Social Media in an Emergency Supply and Demand Framework in Disaster Response,” 2018 IEEE Intl. Conf. on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications <https://ieeexplore.ieee.org/document/8672243>, (Last access: Nov. 20, 2022).
- [19] T. Sakaki et al., “The possibility of social media analysis for disaster management,” 2013 IEEE Region 10 Humanitarian Technology Conference, 2013, pp. 238-243, <https://www.scopus.com/record/display.uri?eid=2-s2.0-84893406250&origin=inward&txGid=7adf7d88a2a5fe170927ab1110f2009f>, (Last access: Nov. 20, 2022).

Author Biography

Samrudha Mhatre received his Master’s in Computer Science, focusing on Cyber Security in 2022. His research interests include computer security and OSINT technologies and applications.

Franziska Schwarz received her M.Sc. in Computer Science from Technische Hochschule Brandenburg (Germany) in 2022. Since 2021, she has worked in cyber security consulting with clients in the public and private sectors. Her research focuses on Cybersecurity and Management, Data Protection, IoT, and Smart Home Security.

Klaus Schwarz received his B.Sc. and M.Sc. in Computer Science from Brandenburg University of Applied Sciences (Germany) in 2017 and 2020, respectively. Klaus is working in technology consulting as an AI specialist for clients in the public and private sectors. Furthermore, he is a Ph.D. student at the University of Granada, Spain. His research interests include IoT and smart home security, OSINT, mechatronics, additive manufacturing, embedded systems, artificial intelligence, and cloud security. As an SRH Berlin University of Applied Sciences faculty member, he developed a graduate program in Applied Mechatronic Systems focusing on Embedded Systems at SRH Berlin University of Applied Sciences.

Reiner Creutzburg is a retired Professor of Applied Computer Science at the Technische Hochschule Brandenburg in Brandenburg, Germany. Since 2019, he has been a Professor of IT Security at the SRH Berlin University of Applied Sciences, Berlin School of Technology. He has been a member of the IEEE and SPIE and chairman of the Multimedia on Mobile Devices (MOBMU) Conference at the Electronic Imaging conferences since 2005. In 2019, he was elected a member of the Leibniz Society of Sciences to Berlin e.V. His research interest is focused on Cybersecurity, Digital Forensics, Open Source Intelligence (OSINT), Multimedia Signal Processing, eLearning, Parallel Memory Architectures, and Modern Digital Media and Imaging Applica-

tions.

Tool	Installation	Operating Platform	Free Trial	Format of Results	Benefits	Limitations	Data Collection Methods	Pricing
Hunter	easy	Web-based	25 searches and 50 monthly verifications	text	account option for teams; API option; search features including domain search, bulk domain search, ...; verifying email addresses, tracking email campaigns, and sending email drip campaigns	credits are deducted from Hunter for generating emails, but Hunter also takes away credits for validating those emails	most recent, real-time	25 searches and 50 verifications per month; free (paid version with more features)
Emailable	easy	Web-based	Limited to credit	graphical	helps marketing teams plan campaigns by sorting email addresses into multiple categories; provides open API, developers can connect to common programming languages	veracity of information	not known	limited to credit; free (paid version with more features)
Phantom Buster	easy	Web-based	14-day free trial	text	helps to find professional email quite easily; reduces the complexity of automating search with various platforms like LinkedIn and Twitter	pricing is quite high; social media platforms are heavily emphasized in software	real-time	14-day free trial; free (paid version with more features)
LeadFuze	easy	Web-based	25 leads free	text	quick set-up; high-quality detailed data	confusing to use; export limits; automation limits; pricing is high	real-time	25 leads free; free (paid version with more features)
Email Harvester	moderate	runs on virtual machine	unlimited	text	output is specific to the command	output is not real-time; needs virtual machine	stored data is reveal	unlimited; free
Simple Email Reputation	easy	Web-based	250 queries per month, up to 10 queries/day	text	easy to use; provides good information	accuracy of the information	real-time	250 queries per month, up to 10 queries/day; free (paid version to use more features)
SpiderFoot HX	easy	Web-based	5 scans available and 1 target per scan	text	easy to use; good detailed information	paid version needed to enable many features; limited time for the scans in the free version	Unknown	5 scans available and 1 target per scan; free (paid version to use more features)
AeroLeads Email Finder	moderate	Web-based	limited to 5 credits	text	pricing is relatively good; enriches the contact with other details	improve on their ease of use; accuracy needs to be improved	not known	5 credits; free (paid version to use more features)

E-Mail Header Analyzer	easy	Web-based	unlimited	text	easy to use; good results (information is produced)	accuracy needs to be improved	most recent; real-time	free
Google Admin Toolbox Message Header	easy	Web-based	unlimited	text	easy to use; good results (information is produced)	accuracy needs to be improved	most recent; real-time	free

Table 1: List of tested and recommended OSINT-based Email investigation tools