Πανεπιστήμιο Πειραιώς
University of Piraeus

Information Gathering via Open Source Intelligence

Dimitrios Lazarakis

A.M: E18089

A thesis submitted in fulfilment of the requirements for the degree of

Bachelor of ICT

2023

School of Information and Communication Technologies

University of Piraeus

Supervisory Panel

Dr. Stefanos Gkritzalis

Professor of Information and Communication Systems Security at the Lab of Systems Security, Dept. of Digital Systems, University of Piraeus, Greece and Director of the Postgraduate Programme "MSc in Law and Information & Communication Technologies".

Abstract

This thesis examines the increasing significance of Open Source Intelligence (OSINT) in the age of freely available information. With the rise of openly accessible information the value of Open Source Intelligence (OSINT) investigations and tools are accentuated. The data sources used in OSINT is data available from a non private source. Its roots can be traced to military intelligence collection techniques, where during the second World War OSINT techniques were utilized for intelligence collection about the enemy. Its recent form, was introduced with the Iranian Protests of 2009. In the 21st century OSINT techniques and sources are mainly used in cybersecurity operations and for keeping tabs on global events. This is largely because of the abundance of information on the internet and how simple it is to obtain and evaluate it. Red teams simulate attacks against a target, find holes and flaws in their defenses, and make recommendations for strengthening company security posture, therefore high-yield information collection is critical. OSINT can assist them in the gathering of data about a target's personnel, corporate operations, and infrastructure. On the other hand, Blue teams can utilize open sources of threat data and analytics to prepare their defenses according to threats. Furthermore, recent events such as the war in Ukraine have resulted in wider utilization of Open Source Intelligence and information collection. The widespread use of social media has resulted in an abundance of publicly available data on people's and business' profiles. Major events taking place, such as protest and wars have a measurable impact online. When conducting an OSINT investigation the topic of ethics should be raised. In order to avoid putting peoples privacy in danger and to produce the best investigation results certain judgement criteria are proposed.

Keywords: OSINT, intelligence, investigation, cybersecurity, journalism

Contents

List of Figures

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Chapter 1

Introduction

OSINT or Open Source Intelligence involves collecting and analyzing data from publicly accessible sources in order to produce valuable intelligence [8]. This specific topic was chosen due to the impact of investigations based on modern open sources. The sources and methods vary according to the needs of the organization or the person performing Open Source Intelligence. For example, National Intelligence Services utilize sources such as foreign news broadcasts. An attorney may search for publicly available state records to draw information for a case. A Red Team operator would search for mentions of the target in publicly available online services during reconnaissance. The invasion of Ukraine highlighted the effectiveness of OSINT [9] in the monitoring of world event. The investigations performed by state intelligence services and journalists helped uncover critical enemy movements and observe the situation from the eyes of the soldiers and civilians.

The objective of this thesis is to introduce OSINT to an audience without previous technical investigation knowledge, while presenting how are investigations used in current events and conflicts. The topics analyzed range from the military roots of Open Source Investigations to free platforms that are used for sharing threat intelligence.

Chapter 2

Military Intelligence

2.1 Defining Intelligence

Intelligence as defined by the *Britannica encyclopedia* is *”Intelligence, in military science, information concerning an enemy or an area. The term is also used for an agency that gathers such information.”* [10]. In essence, intelligence is information about a third party.

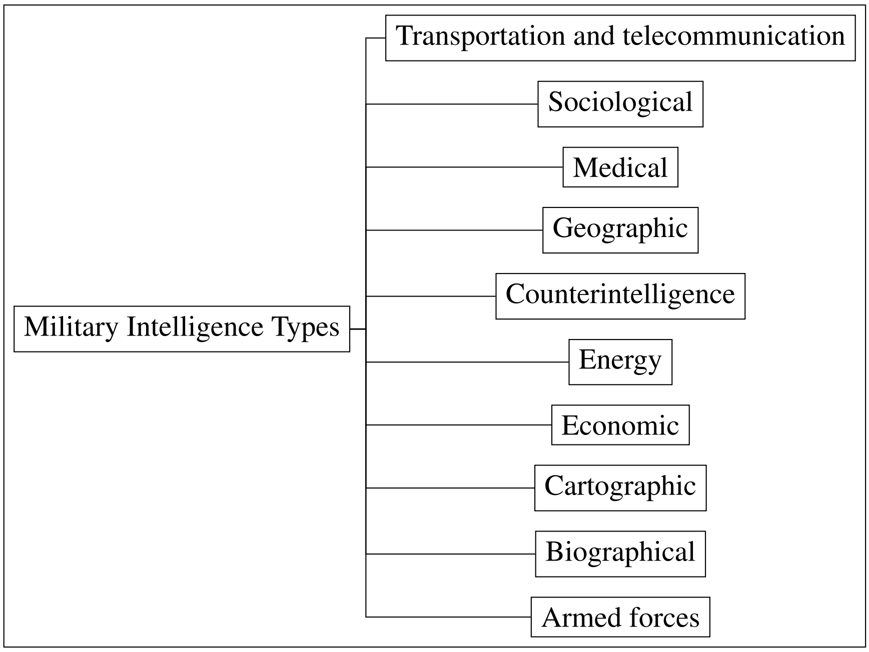


Figure 2.1: Types of military intelligence [1].

The production of usable intelligence involves assessing conflicting pieces of incomplete information, in order to determine the correct items and afterwards combine the accurate items into a document that meets the needs of the operation. The final product is called Intelligence Appraisal or Intelligence Assessment and might contain pieces of incorrect information [10]. Intelligence about the **Armed forces** involves collection of information about the military potential of the enemy in quantifiable and non quantifiable aspects, such as number of personnel, training, capabilities and readiness. Such information could are vital in the search for enemy weak spots, in case of conflict. **Biographical** information involves collecting a variety of characteristics about leaders and high-level individuals. From political views, to habits and professional history such details help analysts uncover motives and intentions of political actions. Knowledge about maps and charts are classified in the **Cartographic** type and can provide information about rough terrain and roads to the combatant troops. **Geographic** intelligence involves gaining information about the natural characteristics of a place, including transportation, population distribution and military locations. The **Economic** type involves information about any type of economic activity inside a nation's borders or internationally. Estimating the economic capabilities of the enemy might correlate with military capacity and could uncover intentions. The **Energy** sector of intelligence consists of data about energy resources, policies, latest energy technologies, power production and use. In the case of military planning, the energy footprint of an military operation must be calculated in advance. Counterintelligence operations planned to counteract espionage, clandestine intelligence activities, sabotage and terrorist attacks are designated in the **Counterintelligence** intelligence type. The **Medical** type focuses on the influence of foreign natural or man made environments on the health of military forces. This information can be utilized to predict enemy medical weaknesses and to adequately protect one's own forces. Social characteristics, beliefs and values are **Sociological** types of information, vital for nations with social, racial or national factions. Finally transference capabilities and communication aptness can by found in the **Transportation and telecommunication** intelligence type [1].

2.2 The Intelligence Cycle

According to the CIA the Intelligence Cycle are the following 5 steps (Figure 2.2):

1. **Planning and Direction**: At first the goals of the intelligence gathering are discussed. The issue and what needs to be found out are decided, as well as the ways that the

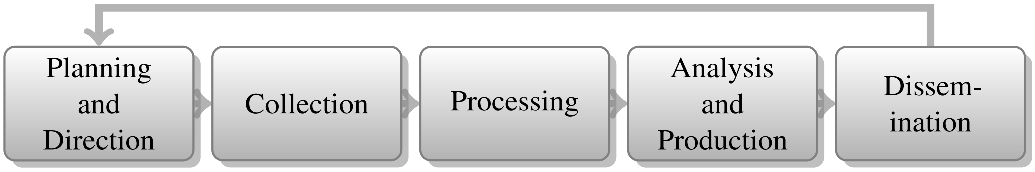


Figure 2.2: The Intelligence Cycle [2].

intelligence will be gathered.

2. **Collection**: The collection steps involves gathering information overtly and covertly. Overtly involves gathering information from Open Sources such as radio broadcasts, magazines and newspapers. Covertly entails secretly monitoring, with the use of technologies such as hidden cameras, listening devices and satellite imagery.

3. **Processing**: All the information collected is recorded and an intelligence report is composed.

4. **Analysis and Production**: The information is analyzed as a whole and the answers of the goal set up at planning phase are answered. Notably, the answers involve the facts discovered, the rationale, future events and how the country of the United States is affected.

5. **Dissemination**: The answers are presented in an analysis format to the policymaker. If more questions are raised or supplementary facts are required, the process restarts [2].

2.3 Intelligence Collection Disciplines

OSINT is a form of collecting intelligence and is a subset of collection disciplines often referred to as ”Intelligence Collection Disciplines” or the ”INTs”. By the U.S. Naval War College, the five main types are:

1. **Human Intelligence (HUMINT)**: Information collection from human sources. The collection may be done openly, with witness or suspect interviews, or it may be done through covert means (espionage).

2. **Signals Intelligence (SIGINT)**: Electronic transmissions that are collected by various means, with the use of ships, ground sites, planes and satellites. Communications Intelligence (COMINT), a type of SIGINT, is the intercepted communication between two parties.

3. **Imagery Intelligence (IMINT)**: Image Intelligence, also known as Photo Intelligence (PHOTINT), has changed over form due to the development of imaging technology. In the American Civil War hot air balloons with cameras were used for observation. In the first and second World Wars aircrafts were equipped with cameras in order to gather intelligence. Nowdays imagery satellites are used due to their high quality photographic capabilities. Geospatial Intelligence (GEOINT) is the analysis and visual rendering of security related activities on the earth produced by combing imagery, imagery intelligence and geospatial information.

4. **Measurement and Signatures Intelligence (MASINT)** : Advanced Information processing from overhead and airborne IMINT and SIGINT systems. Weapon's telemetry can be intercepted - Telemetry Intelligence (TELINT), as well as sensor data from modern weapons and tracking systems - electronic intelligence (ELINT).

5. **Open-Source Intelligence (OSINT)**: Collection of information that is publicly available, mainly from sources covered in Section 3.1 [11].

Chapter 3

Open Source Intelligence

3.1 Data Sources

The Data portal of the European Union classifies the sources of OSINT in six categories:

- **Public media**: Such as television, newspapers and magazines.
- **Internet**: Every online source of information from social media to forums and online publications or blogs.
- **Public government data**: Publicly available documents such as government reports, budgets, court hearings and speeches.
- **Professional and academic publications**: Publicised work i.e. journals, academic papers, theses and conferences.
- **Commercial data**: Any type of data of commercial nature, financial and business assessments and databases.
- **Grey literature**: Harder to find documents such as patents, technical reports, unpublished works and business documents [8].

3.2 Standard Operating Procedure (SOP)

Before starting any data collection a **Standard Operating Procedure** should be decided upon. The **Goal Framework** should be set according to the questions the investigator wants to answer. The massive amount of publicly available information impedes investigations without a clear goal set due to the noise created. Afterwards a **Strategy** will be selected according to the desired level of interaction with the target:

- **Passive**: The investigator will only access publicly available information without interacting with the target or their assets.
- **Semi-Passive**: The investigator will be carefully sending some data to target assets to gather more technical information.
- **Active**: Direct Engagement with systems or persons, resulting to greater data collection. Has increased risk of being noticed and flagged as malicious [12].

3.3 Historic Reference

The history of Open Source Intelligence is difficult to be traced due to its use by military and intelligence services. According to the *Journal of US. Intelligence Studies* OSINT can be traced back to the exploitation of intelligence to aid in decision making of governments. Its methodical practise can be attributed to the founding of the *Foreign Broadcast Monitoring Service (FBMS)*, which was the result of research of the Princeton University. The *FBMS* gained boost after Pearl Harbor and was integrated in the newly established CIA in 1947. Post 9/11, *FMBS* was renamed to *Director of National Intelligence's Open Source Center (OSC)*. During the 2nd World War Britain launched its own journalism monitoring program utilizing the resources of BBC, now known as *BBC Monitoring*, later partnering with US intelligence services [1]. During the Cold War, Open Sources was the leading source of all intelligence according to CIA analyst Stephen Mercado [13].

After the collapse of the Soviet Union, the intelligence services of the West redirected their attention to non-state actors causing conflicts in regions such as Asia or Africa and to potential national threats. Thematic priorities up to the present day include:

- Conflict in expeditionary environments
- Political and religious terrorism
- Weapons of Mass Destruction (WMD)

- Vulnerabilities of computer networks

The term OSINT was officially introduced in the late 1980s. The founding of *Community Open Source Program Office (COSPO)* by the *CIA* and joint efforts with the EU would be done in order to harvest the ever-growing flood of information. Post 9/11, the US agencies tasked with Open Source information gathering would be absorbed into the newly created *Open Source Center (OSC)* which would be named textitOpen Source Enterprise (OSE) in 2015 [13].

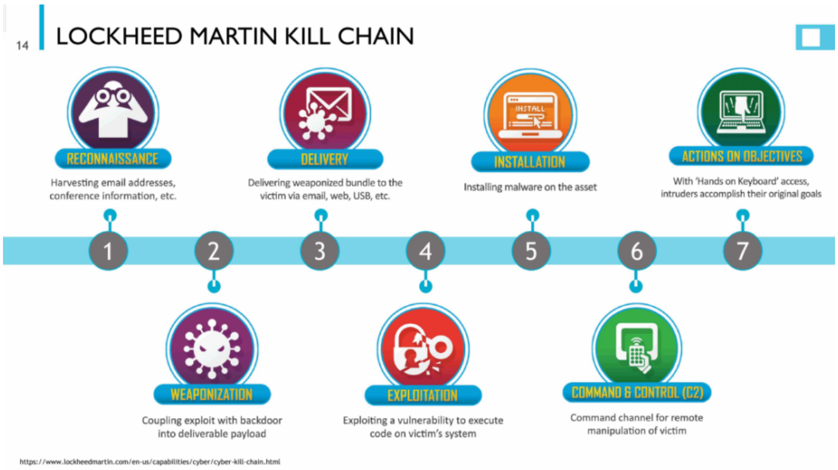
Bellingcat, an worldwide group of people utilizing Open Source investigative methods to find information about pressing issues, from drug trafficking to human rights violations during the invasion of Ukraine, claims that William Donovan, a World War One veteran and International Lawyer laid the founding grounds for the role 'Coordinator of Information' in the US government. Post Pearl Harbor the need for intelligence was highlighted and the Department where Dovovan worked in was renamed to Office of Strategic Services, the precursor of the CIA. A whole branch of the department was dedicated to Open Source Intelligence [14]. He wrote: *"Even a regimented press will again and again betray their nation's interests to a painstaking observer"* [14]. After World War Two the department saw a decline in agents and went into a deep sleep up to Iranian Protests of 2009. Due to increased internet usage in Iran the internet was flooded with citizen data about the political landscape of the country. Mining social media feeds, blog posts, articles and Twitter posts provided insightful analysis and formed the modern methodology of Open Source Intelligence. Other organizations such as the US and the UK took notice and implemented the same methods [13]. The potency of Open Source Investigations can be highlighted in examples such as an US military operation where a bomb making factory was identified in a selfie by the roof structure visible and destroyed in the time span of 24 hours [15].

Chapter 4

Open Source Investigations

4.1 Information Collection in Cyber Security Operations

The definition of a Red Team according to *NIST* is: *A group of people authorized and organized to emulate a potential adversary's attack or exploitation capabilities against an enterprise's security posture. The Red Team's objective is to improve enterprise cybersecurity by demonstrating the impacts of successful attacks and by demonstrating what works for the defenders (i.e., the Blue Team) in an operational environment. Also known as Cyber Red Team.* [16]. The Cyber Kill Chain is the outline of the stages of a Red Team Assessment as seen on figure 4.1.



14 LOCKHEED MARTIN KILL CHAIN

1 RECONNAISSANCE
Harvesting email addresses, conference information, etc.

2 WEAPONIZATION
Coupling exploit with backdoor into deliverable payload

3 DELIVERY
Delivering weaponized bundle to the victim via email, web, USB, etc.

4 EXPLOITATION
Exploiting a vulnerability to execute code on victim's system

5 INSTALLATION
Installing malware on the asset

6 COMMAND & CONTROL (C2)
Command channel for remote manipulation of victim

7 ACTIONS ON OBJECTIVES
With 'Hands on Keyboard' access, intruders accomplish their original goals

https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

Figure 4.1: The Cyber Kill Chain by Lockheed Martin [3].

4.1.1 Reconnaissance Techniques

In order to emulate real adversary attacks the red team needs to perform reconnaissance. Open Source information will be the first domain of reconnaissance since the interaction with the target can be as little as agreed in the Standard Operating Procedure (Chapter 3). A very useful resource for red and blue teams is the *MITRE ATT&CK®* framework [17]. The *ATT&CK* framework is an open knowledge base of adversary techniques and tactics. The reconnaissance category includes the following tactics in the following order:

- **Active Scanning**: Actively scanning for open ports, vulnerabilities or hidden directories on the target. Mass scanning that interacts with the target is noisy and may arouse suspicion.

- **Gathering Victim Host Information**: Gathering information from the exposed surface area of the target. A photograph posted on social media with the operating system of the employees being visible, along with the endpoint management solution may aid an attacker in specializing their exploitation path. Moreover, public relationships with companies that design countermeasure solutions may be used as a hint of target infrastructure. Finally, exposed or leaked configuration of the defense solutions is of great importance since an attacker can further specialize their attack. For example, finding that an Endpoint Detection and Response software is configured to ignore a custom file path in Windows can assist the adversary in landing their malware onto the disk of their victims silently.

- **Gathering Victim Identity Information**: Gathering data about the victim's identity such as personal details, i.e. Names, emails, phone numbers, and sensitive details such as credentials. Leaked credentials present easy entry for an attacker if Multi Factor Authentication (MFA) is not enabled into the infrastructure of the organization as an authenticated user.

- **Gather Victim Network Information**: Finding information about the networking infrastructure of the target. Such data are internal subnets used, topology and trusts. An adversary would target specific networks and avoid detection via Network Based Intrusion Detection Systems (NIDS) with knowledge of the internal setup.

- **Gather Victim Org Information**: Discovering non-infrastructure related information about the target organization such as branches, department names, internal roles and business procedures.

- **Phishing for Information**: Sending phishing emails or messages in order to elicit sensitive information from the target. The target is tricked into submitting sensitive information such as domain credentials, online account credentials to be used later by the adversary.

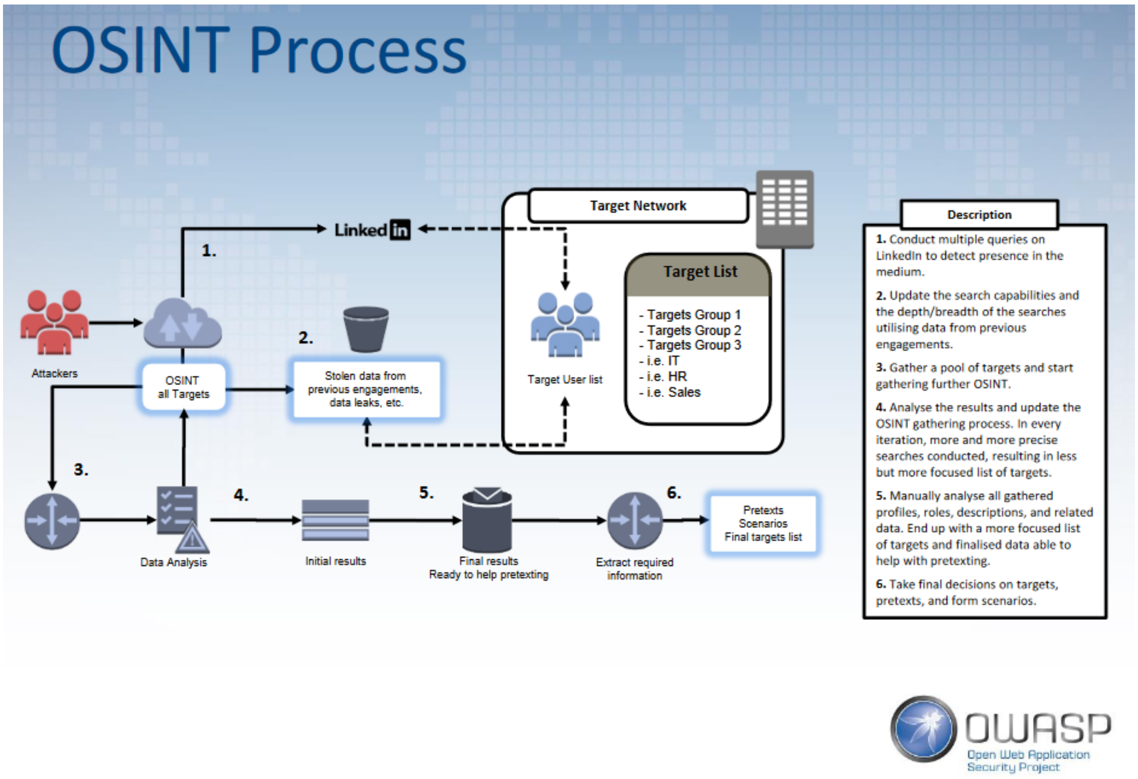
- **Search Closed Sources**: Finding insight about the target from private sources such as private databases containing data about the organization , purchased technical data and Thread Intelligence feeds. These types of data are sold by both reputable and criminal companies for some amount of money. Advanced Persistent Threats can buy leaked databases or exfiltrated data from the organization to gain insight into their target. Due to the demand of closed data sources many markets have emerged on the clear and hidden web.

- **Search Open Technical Databases**: Gathering information from public sources including DNS records, website certificates, WHOIS data and databases containing scans. Many Open Source searching tools focus on this part of reconnaissance since the interaction with the target ranges from none to minimal and is one of the first information gathering phases of an attacker.

- **Searching Open Websites/Domains**: Utilizing data put willingly online by the target on various website not owned by them, including social media account feeds, indexed data by search engines and public code repositories.

- **Search Victim-Owned Websites**: Gathering data from websites owned by the target organization in order to extract insight. For example, such data include physical locations, key information about business procedures, contact info and business relationships with other organizations [17].

As it can be seen from the mapped techniques of *ATT&CK*, information from open sources is in the core of reconnaissance. When a red team is against an organization with modern security measures each piece of insight collected is valuable to reaching the objectives of the assessment. In Figure 4.2 the social media *Linkedin* is used to get information about the target organization. By first finding the company on Linkedin and then indexing the employees, the search for leaked credentials and stolen data becomes increasingly precise and targeted. Potential targets with little cybersecurity and data protection knowledge be discovered and targeted individually utilizing a technique called Spear Phishing [18]. In case the target of the phishing campaign is a high value target, such as an executive member, the technique is called Whaling [18]. The effectiveness of these attacks is multiplied when the attacker already has gained access to an email account belonging to the organization, to a vendor, a customer or a partner since the sender address is trusted by the victim.



OSINT Process

1. Conduct multiple queries on LinkedIn to detect presence in the medium.

2. Update the search capabilities and the depth/breadth of the searches utilising data from previous engagements.

3. Gather a pool of targets and start gathering further OSINT.

4. Analyse the results and update the OSINT gathering process. In every iteration, more and more precise searches conducted, resulting in less but more focused list of targets.

5. Manually analyse all gathered profiles, roles, descriptions, and related data. End up with a more focused list of targets and finalised data able to help with pretexting.

6. Take final decisions on targets, pretexts, and form scenarios.

Figure 4.2: OSINT Process by OWASP [4].

4.1.2 Threat Intelligence

According to Crowdstrike Threat Intelligence is the process of gathering, processing, and analyzing data to comprehend the objectives, targets, and methods of a threat actor [19]. Utilizing Threat Intelligence enables proactive protection as far as digital threats are concerned. The techniques of ransomware groups can be predictable. In order for a malware to execute, it has to bypass most of the digital protections of the organization and even then it may yield limited damage due to modern ransomware protection built in Endpoint Detection and Response applications. Finding such techniques is a product of research, a time consuming task that may not be effective after all. Therefore, most ransomware groups utilize a limited number of TTPs, that may be slightly altered over time, on a large number of victims. With the help of Threat Intelligence platforms organizations of any size are able to get the latest intel in cyber defense, as well as identify potential attacks. The platforms can be divided into public and private ones. Private platforms do not share their intel or evidence publicly and boast a high level of confidentiality. Public platforms may or may not require an account and are open to anyone to sign up. Some of the many noteworthy platforms accessible publicly can be found bellow:

- **OTX** by Alienvault is an excellent platform for Threat Intelligence due to the vast amount of information being uploaded and analyzed. OTX presents a strong tool in Alienvault's line of firewalls due to their excellent integration of cloud threat data and local network analysis [20].

- **TALOS**: TALOS by CISCO offers a plethora of data based on IP reputation and regular reports about APT groups. The platform integrates with many security products in order to perform reputational IP lookups or download and evaluate intel containing data such as newly discovered Command and Control servers [21].

- **Virustotal**: VirusTotal scans objects using more than 70 antivirus scanners and URL/domain blocklisting services in order to extract potentially malicious signals. The platform offers free scans via the online interface, browser extensions, and an API. Static and dynamic analysis is performed and a comment area is provided to add community comments about safety of the file uploaded [22].

- **Abuse.ch**: Abuse.ch offers community-driven threat intelligence. Its primary purpose is defending IT infrastructure against malware. IT security researchers, vendors, and law enforcement organizations poll data from Abuse.ch. Their publicly available platforms are Malware Bazaar, Feodo Tracker, SSL blacklist, URL Haus, Threat Fox and Yara IFY [23]. Each service covers a sector of threat intelligence and defense.

4.2 OSINT for Monitoring World Events - Invasion of Ukraine

Due to the plethora of information uploaded on a daily basis on the internet in the form of text, media or metadata, real world events have an measurable impact online. An observer can utilize a variety of platforms and tools to gather media and information such events. Furthermore, social media have been used as protest tools in order to propel action. Such example is the protest of the housing crisis in Kenya before the elections, where social media played a critical role in spreading the message with the use of hashtags [24]. Unarguably the reach and swift spread of news and events that the internet offers have changed the way information is being distributed and consumed.

Social Media such as Twitter, Instagram, Snapchat and Facebook are used by millions of people every day. The pandemic had limited social interactions, though the need for contact remained the same resulting in seeking such interactions through social media [25]. The usage of a smartphone is synonymous with social networking platforms. With the improvement of camera technology and internet connectivity it is only logical that a person would utilize a highly capable device to capture moments of their life or an event happening in front of them. Spreading information through social media is primarily done through photographs or short form videos. During the invasion of Ukraine Twitter and messaging platforms such as Telegram were flooded with photos took by soldiers and videos from battle or during resting time. Intelligence services closely monitored known channels for communication in order to survey enemy movements and uncover tactics and squad positions. Many of these videos and pictures were shared publicly and reached hundreds of thousands of people through news outlet or social media platforms, offering a view into the conflict from the eyes of the soldiers. The war of Ukraine has undoubtedly brought Open Source Intelligence investigations into mainstream discussion as seen by the following events and persons.

4.2.1 Twitter Reporters

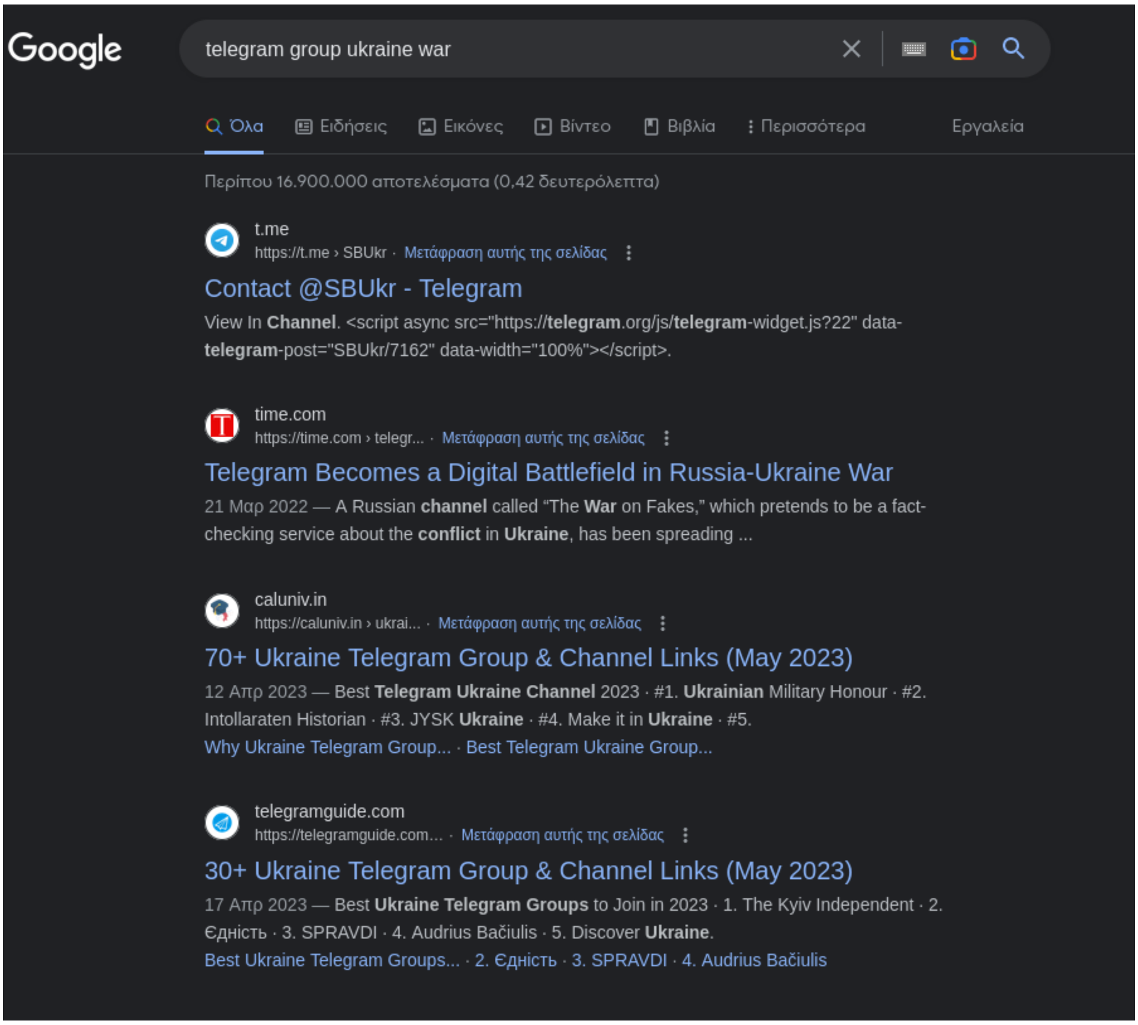
Journalists from all over the world have been particularly interested in the ongoing Ukrainian war and open-source intelligence (OSINT) has been crucial to their reporting.

”Twitter journalists” give audiences insightful perspectives and analysis, while quickly delivering reports of events. Data sources range from social media to satellite photography and other open internet data sources. More specifically Twitter accounts created for monitoring world events mostly use the following source:

- **ADS-B**: An accurate surveillance interface between aircraft, Air Traffic Control and public is made possible by the technology known as Automatic Dependent Surveillance-Broadcast (ADS-B). It integrates an aircraft’s location source, avionics, and ground infrastructure, while being more accurate than a ground radar. ADS-B is a is comprised of two distinct services: ADS-B Out and ADS-B In [26].
- **Satellite imagery**: From NASA FIRMS [27] to commercial satellite imagery by providers such as Maxar Technologies [28] images of earth can be found for free or bought. Satellite imagery has been used for purposes such as verifications of claims by armies, post battle analysis [29], military movements [30] and uncovering human rights abuses [31].
- **Publicly shared information**: Information comprised from any official source. Official sources include the social media profiles of the armies and the announcements made by each country’s ministry of foreign affairs or defense ministry. Such information is biased, though it can be really useful in assessing both sides of the story during a war.
- **Private messaging groups**: Due to the on going war many groups have been set up in favor of either country. The channels present information in a biased way and spread fake news. A researcher has to cross reference events mentioned in such channels in order to assess their validity. At the time of writing a prominent app for groups is Telegram as seen on Figure 4.3.

4.2.2 Secret Document Leak

Shortly after the online leak of classified US documents about strategic Russian bases originating from a discord channel, a U.S. Guardsman was arrested. In regards to this incident, many aspects of OSINT can be discussed. To begin with, the documents had been



Περίπου 16.900.000 αποτελέσματα (0,42 δευτερόλεπτα)

t.me
https://t.me › SBUkr · Μετάφραση αυτής της σελίδας

Contact @SBUkr - Telegram

View In **Channel**. <script async src="https://**telegram**.org/js/**telegram**-widget.js?22" data-**telegram**-post="SBUkr/7162" data-width="100%"></script>.

time.com
https://time.com › telegr... · Μετάφραση αυτής της σελίδας

Telegram Becomes a Digital Battlefield in Russia-Ukraine War

21 Μαρ 2022 — A Russian **channel** called "The **War** on Fakes," which pretends to be a fact-checking service about the **conflict** in **Ukraine**, has been spreading ...

caluniv.in
https://caluniv.in › ukrai... · Μετάφραση αυτής της σελίδας

70+ Ukraine Telegram Group & Channel Links (May 2023)

12 Απρ 2023 — Best **Telegram Ukraine Channel** 2023 · #1. **Ukrainian** Military Honour · #2. Intollaraten Historian · #3. JYSK **Ukraine** · #4. Make it in **Ukraine** · #5.
Why Ukraine Telegram Group... · Best Telegram Ukraine Group...

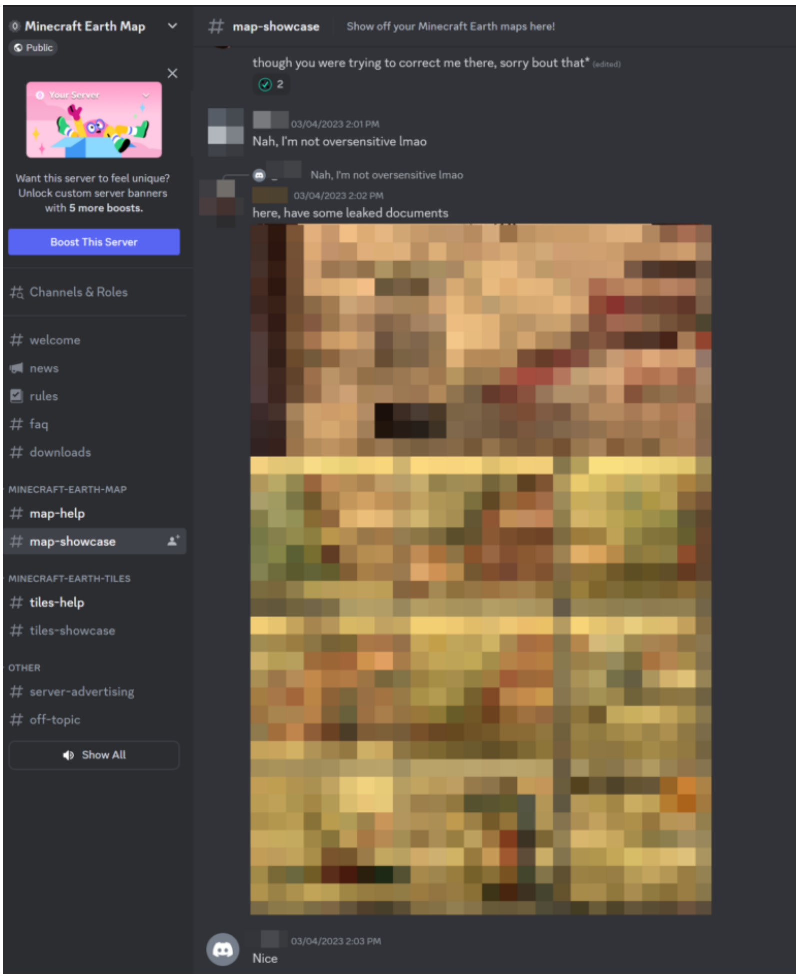
telegramguide.com
https://telegramguide.com... · Μετάφραση αυτής της σελίδας

30+ Ukraine Telegram Group & Channel Links (May 2023)

17 Απρ 2023 — Best **Ukraine Telegram Groups** to Join in 2023 · 1. The Kyiv Independent · 2. Єдність · 3. SPRAVDI · 4. Audrius Bačiulis · 5. Discover **Ukraine**.
Best Ukraine Telegram Groups... · 2. Єдність · 3. SPRAVDI · 4. Audrius Bačiulis

Figure 4.3: Google search for Telegram groups about the war in Ukraine.

leaked to many sites, by both pro-ukrainian and pro-russian groups. The interesting fact is that when comparing the leaked map found on a forum messaging board called 4chan (Figure 4.5) to the leaks shared on the Telegram channel ”Donbass Devushka” containing the same map, some details are different. According to Bellingcat the Telegram channel map mentions more Ukrainian casualties and less Russian casualties. The Telegram map was discovered to be edited in a bad manner, indicated by bad spacing between numbers, letters and unmatched fonts. However, both of these images were then later discovered to not be the original leak. Before the public publication of the documents, they are believed to have circumvented around various Discord channels. One of the Discord posts predating the 4chan leak can be seen in the following screenshot provided by Bellingcat of a private Discord Minecraft group in Figure 4.4 [5] [32]. The original source of the leak is still debated as of the 14th of April 2023. The impression of this event for an OSINT researcher can be summed up to the following: Be mindful of disinformation campaigns, cross reference material from different sources and check the evidence for inconsistencies.



though you were trying to correct me there, sorry bout that* (edited)

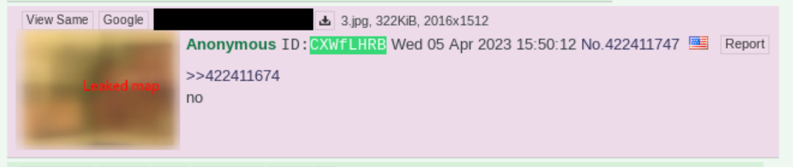
2

03/04/2023 2:01 PM
Nah, I'm not oversensitive lmao

Nah, I'm not oversensitive lmao
03/04/2023 2:02 PM
here, have some leaked documents

03/04/2023 2:03 PM
Nice

Figure 4.4: Bellingcat's screenshot of the Discord map leaks [5]



[redacted] 3.jpg, 322KiB, 2016x1512

Anonymous ID:CXWfLHRB Wed 05 Apr 2023 15:50:12 No.422411747

>>422411674
no

Figure 4.5: The 4chan leak post mentioned in Bellingcat's article (Subsection 4.2.2).

4.2.3 ”Eyes on Russia” Map

CIR or the Centre for Information Resilience is an non-profit autonomous organization. Their goal is to expose human rights violations, battle misinformation, and prevent abusive behavior toward women and minorities. In order to gather and authenticate any videos, pictures, satellite images, or other media resulted from Russia's invasion of Ukraine, CIR developed the "Eyes on Russia" project [6]. The pieces of media uploaded are verified and geo-located. The purpose of this map is to make reliable information accessible to the general public, politicians, NGOs and journalists. A snapshot of the map during the map of April can be found on image 4.6.

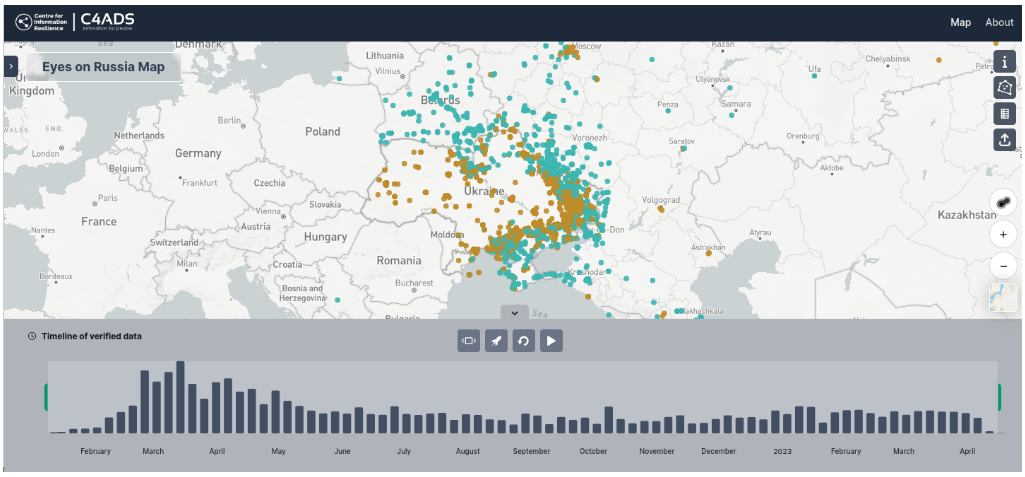


Figure 4.6: "Eyes on Russia" - April 2023. [6]

Chapter 5

Gathering Intelligence

5.1 Red Team Toolset

As discussed on Chapter 4 Red Teams could extract a great amount of data about their target using OSINT. Special tools have been created that automate repeated processes and tasks.

5.1.1 Tools

Amass

Amass is a Golang based tool created by OWASP which utilizes passive and active reconnaissance to map the attack surface and discover public facing assets of an organization. The techniques followed range from interacting with external APIs such as Shodan's [33], to looking up TLS certificate and IP info. The tool is widely trusted and used by red teams due to the automation it enables.

gitleaks

Gitleaks is a Golang based tool used to search git directories for hardcoded secrets such as credentials and API keys [34]. Leaked secrets have resulted in unauthorized access in company property such as the case of Toyota where attackers got access to customer data through a leaked access key [35]. Moreover, notable is that fact that gitleaks offers a Github Action where each code commit of a repository would be vetted for leaked secrets.

SpiderFoot

Spiderfoor is an Open Source automation tool. It offers several great features with its 200 integrated modules, from domain enumeration to Threat Intelligence. The power of Spiderfoot lies in its simplicity, where only a single input such as a username, an email address or a domain can extract a lot of information about a person or company. The tool presents the pieces of information discovered in a connection oriented graph fashion in order to demonstrate the links between data [36].

LinkedtInt

LinkedInt is an Open Source scraper for the social network Linkedin. It indexes the employees of a company based on their Linkedin profiles and discovers company email addresses based on naming conventions [37]. Discovering the employees of a company is very useful for a red team since several pieces of information, such as departments and personnel roles, can be discovered. Moreover, employees can be selected for targeted phishing in order to infiltrate the organization.

5.1.2 Internet Device Search Engines

Due to the number of internet connected devices, special search engines have been created in order to catalog exposed services. The entire public IP range is scanned for open ports and the hosted services are enumerated. An attacker could utilize the data provided by the following search engines in order to discover exposed services without interacting with the targets. An organization can opt out of scans by blocking the IP addresses associated with each the scanning service.

Shodan

Shodan is a search engine for internet connected devices. Internet wide port and service scans are performed and the results are uploaded to their website. When performing a query the familiar layout of a search engine is presented, with filtering options such as country, service or IP owner [33]. A user can quickly go through massive amounts of indexed open services with the help of filters. An advantage of Shodan is that information about services are saved. For example, in case the port 3389 is found open (Remote Desktop Protocol) a screenshot of the login screen is saved. Therefore, an attacker could get a high level overview of the services exposed in an IP range or an organization with communicating with the hosts. Many modules about Shodan's API have been written in a variety of programming languages, enabling automation and mass enumeration.

Censys

Censys is a platform comparable to Shodan. Internet connected devices are routinely scanned and their services are indexed. With Censys' search engine a user has access to the scan data, as well as to filters to aid investigators narrow down their results [38].

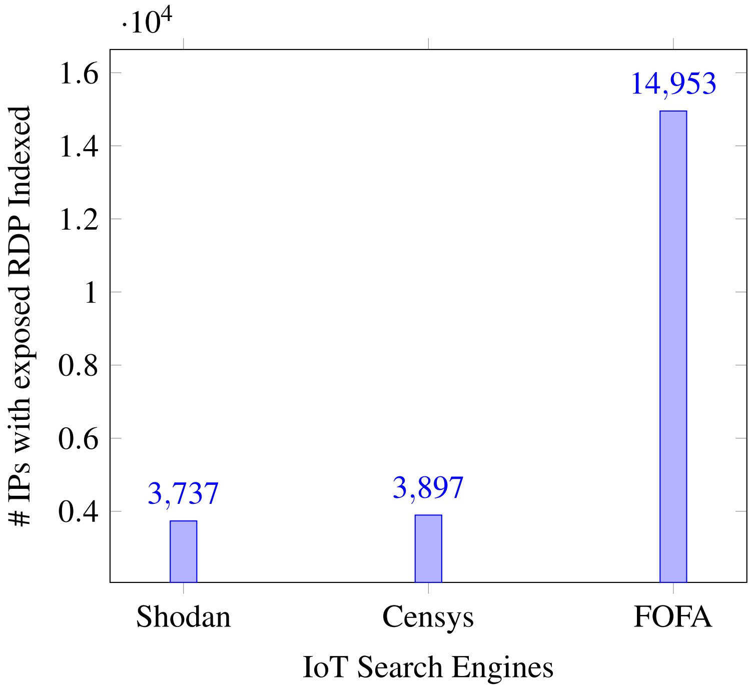
FOFA

FOFA is a Chinese internet connected device search engine that indexes the publicly services and devices [39]. FOFA seems to more aggressively enumerate the open services compared to the other scanners.

LeakIX

LeakIX is comparable to the other platforms, though the the search and indexing is optimized for finding bad configurations and leaked info through website [40].

5.1.3 Service Discovery Comparison



In order to compare the effectiveness of the IoT search engines the common query of instances of port 3389 (RDP - Remote Desktop Protocol port) in Greece was performed. RDP was selected due to its attack surface. RDP is used to remotely control an OS and an attacker could access the organization's private network with a set of compromised user credentials. The proposed way to use RDP from another network is through a VPN or a third party solution such as Citrix. As seen in the Graph 5.1.3 the results yielded from the FOFA search are a lot more than its counterparts. This could be attributed to a combination of returning historic data - Shodan and Censys do not return historic data in a query and of the obscurity of the IPs of FOFA's scanners. Simply put, while an administrator can disallow inbound scanner access to the IP ranges owned by Shodan and Censys, FOFA provides no opt-out method.

5.2 Utilizing Social Media

Due to the widespread use of social media a great deal of information exists on personal or business social media pages.

SnapChat

Snapchat is a popular multimedia messaging app developed by Snap Inc. It allows users to send photos, videos, and messages to other users. It also includes features such as Stories and Discover, which allows users to access content from various media outlets. One of the most useful tools for OSINT researchers is Snapchat Map, a feature of Snapchat that allows anyone to view other people's locations on a map where a video or a picture has been uploaded [41]. The map updates in real-time and provides location-based filters, which can be used to enhance snaps taken in certain locations. The map features a heatmap where major world events can be easily discovered since the more pictures or videos are uploaded from a certain location the more noticeable the point becomes on the map. The researcher can then click on the map and get an idea of what is happening from the uploaded pictures and videos

Instagram

Instagram can be used for OSINT in various ways, such as gathering information about trends, people, locations, and more. OSINT on Instagram can be used to identify patterns and trends in user behavior, which is used for marketing purposes such as understanding customer habits. Additionally, OSINT on Instagram can be used to uncover relationships between people, in the form of follows, and can provide valuable insight about a particular person or location. The Story feature on Instagram is disappearing pictures or short videos often accompanied by the location. Public profiles that upload stories or posts tagged with a specific location are indexed and can be discovered by anyone searching for that specific location.

Twitter

Twitter can be a useful tool for monitoring world events and gathering open source intelligence. Because Twitter is a widely used platform, many people, including journalists, government officials, and other sources of information, use it to share news and information about current events. This makes it a valuable source of information for anyone interested in gathering open source intelligence. The instant sharing of information achievable through Twitter can be utilized for OSINT purposes with the following methods:

- Following relevant accounts: Accounts and Hashtags related to the topic of choice that share news and updates. For example during the invasion of Ukraine journalists created accounts that gather and fact check information and footage from the war. Such intel is then used to analyze enemy movements or to spread awareness.

- Using Alerts: Twitter has an advanced search feature that allows the set up of alerts. The user will then be notified when new content is posted related to the search parameter of choice.

- Analytics: Twitter analytics are tools provided by Twitter to businesses to mainly gauge the engagement with followers. Open Source tools such as birdwatcher use the Twitter API to download and then analyze the tweets locally [42].

Chapter 6

Ethics

While OSINT is unquestionably a useful tool for acquiring intelligence and making decisions, some significant ethical questions are posed. The notions of privacy are being brought up first. One important ethical factor for OSINT is privacy, which has been one of the most hotly debated subjects in recent years. Individuals' sensitive information and personal information may unintentionally be exposed during the gathering and analysis of publicly available data. OSINT practitioners must take precautions to preserve people's privacy in order to address this issue, including redacting identifying information and making sure that any data obtained is only utilized for the reason for which it was collected.

6.1 Doxing and Swatting

Doxing or ”dropping documents” is the malicious act of gathering and making public personal data about a person. Those data can be the person's name, address, phone number, email address, social media profiles, and other sensitive information. Such practises are used by malicious individuals or groups with ill intent such as cyberbullies, hackers, and trolls. Vigilante justice is a typical reason for doxing and is frequently used to harass, coerce, or embarrass a victim into compliance [43]. Hacktivism is another reason, used for social or political advantage. Journalists covering far right scandals have been targeted for their work after their details were leaked online [44]. Furthermore, doxing can have detrimental effects on a person. From personal and professional life blemishes due to reputational harm, job loss, and other unfavorable outcomes, to even deaths as a result of doxing. More specifically, there was a case where, a home address was shared to a Discord channel and someone from that channel ”swatted” the address, resulting in the victim having a heard attack. ”Swatting” is derived from the United States' *Special Weapons and Tactics (SWAT)* team that is is used to handle aggressive and dangerous situation [45]. The malicious person would call emergency services and pretend to report a potentially dangerous situation to the emergency operator. Techniques used in the past are that someone is about to commit suicide, a family is in danger from an intruder, or perhaps they have seen someone with a weapon. The police then, shows up armed and ready to neutralize a threat that does not exist. Other deaths have been noted as a result of ”swatting”, such as a 20 year old man being shot on sight after a hostage 911 call [46].

6.2 Biases

A major ethical issue found in gathering information and concluding is the presence of bias. Bias has many forms. Data found online or on open sources can be inaccurate, due to human error or malicious alteration, in order to promote certain beliefs or to spread fake news. The researcher should closely examine each piece of data for incompleteness and inaccuracy by cross referencing sources or finding information that backs the original data up. Personal bias can affect the accuracy of an investigation, therefore researchers must be aware of their own biases and make an effort to convey information in an unbiased and objective fashion. They must also take precautions to ensure the accuracy of the data they gather and be mindful of any potential biases in the sources they use [47].

6.3 Transparency

Another crucial ethical issue in OSINT is transparency. The methods and sources that OSINT practitioners utilize to gather and evaluate information must be openly disclosed. This includes outlining any conflicts of interest, prejudices, or restrictions on the data that was gathered. By being transparent about the sources and methods used, OSINT practitioners can demonstrate the reliability and credibility of their work, particularly in investigations where accuracy and trustworthiness are critical. Moreover, transparency makes it possible for others to validate the data gathered. Other researchers should be able to duplicate the research and confirm the validity of the findings. Additionally, if during an OSINT research information about specific people are gathered, the researchers should be open and honest about the sources they rely on, the goals behind the data gathering, as well that no needless to the investigation information was processed.

6.4 The Gray Area of Open Sources

The definition of an ”Open Source” is debatable topic. Most of the time sources such as officially public documents, personal social media accounts and documents accidentally made public or leaked are regarded ”Open”. Some times leaked documents do not contain sensitive information, though the case may be different for documents marked as ”Internal Use Only” or ”Private/Confidential”. A researcher downloading such files could be exposed to private information such as patient information, passwords or financial data, not meant to be disclosed to the public. Moreover, social media expose a lot of details about a person’s identity and potentially forgotten information saved in the personal profiles. It can be difficult for a person to estimate the amount of information shared online and what they reveal about themselves. OSINT professionals can examine typical social media behavior of a user, such as posts, comments, likes, and shares to learn more about their behavior, preferences, interests, and relationships. Additionally, they can use the user’s social media accounts to follow events, monitor patterns, and find their political beliefs. Therefore, in order to respect the individual’s privacy, ethical guidelines must be followed [7].

6.5 Judgement Criteria

The investigator behind the website ”OSINT Essentials” [48], Eoghan Sweeney proposed 3 criteria in order to decide whether to investigate a specific lead in case of doubt [7]:

- **Need**: How crucial is the information to the verification of the evidence? Might the data be obtained in another, more ethical manner?
- **Value**: How significant is the knowledge that will stem from this? Is this issue an of public importance? For example, exposing corruption of a minister or public money

wasted.

- **Consequences**: What possible consequences could this method of information acquisition have? Even if lawful, is there a risk that this method would undermine the integrity or or the trust of the journalist or the news organization?

The writer presented a figure (Figure 6.1) on his Medium Article as a means to illustrate his criteria [7].

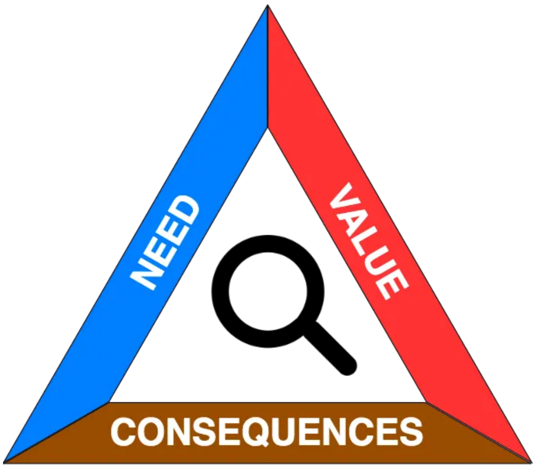


Figure 6.1: The 3 criteria proposed [7]

Chapter 7

Conclusion

In conclusion, this thesis has shed light on the increasing importance and potential of OSINT in various domains. It is clear that Open Source Intelligence has come a long way in recent years. The information available on the internet is actively being used for cyber security purposes, for information collection and for monitoring current events unfolding. OSINT is helping journalists gather data about conflicts, human rights violations and uncover information that would be hidden from the public eye. The public greatly benefits from Open Source research, though some concerns should be raised. Every investigator should consider the ethical aspect of their research, as mentioned on Chapter 6. First of all, the personal privacy of people not involved in an investigation should not be compromised in any way. Furthermore, OSINT investigations targeting certain people should only be performed when an issue of public importance is raised. Privacy campaigns should be launched that help people what information should and should not be shared in order to avoid privacy compromising. Especially non technically inclined social media users may reveal facts about their identity or personal preferences that they may not want online. Advertising organization are creating revenue by selling personal information inadvertently provided by the user.

Open Source information analysis will certainly be on the center of discussions in the following years. The more Virtual Reality (VR) and Augmented Reality (AR) technologies are utilized, the more data are uploaded online. Social Networking will probably closely follow and integrate the newly introduced VR and AR technologies as indicated by the recent investments of Meta [49]. Therefore, a new stream of personal data will undoubtedly be uploaded to such platforms. Finally it could be argued that with the digitization of governmental documents and services, the cyber attack surface of legislative organizations will increase and worsen the effects of an infiltration and leak. Practising privacy and having a proper cyber security posture is vital for persons, organizations and administrative bodies.

Bibliography

[1] N. W. College, “Libguides: Intelligence studies.” [Online]. Available: https://usnwc.libguides.com/intelligence

[2] “The intelligence cycle.” [Online]. Available: https://www.cia.gov/spy-kids/parents-teachers/docs/Briefing-intelligence-cycle.pdf

[3] “Cyber kill chain®.” [Online]. Available: https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

[4] M. Kandias, “Red teaming – osint – phishing.” [Online]. Available: https://owasp.org/www-chapter-dorset/assets/presentations/2020-04/RT_OSINT_Phishing.pdf

[5] A. Toler, “From discord to 4chan: The improbable journey of a us intelligence leak,” Apr 2023. [Online]. Available: https://www.bellingcat.com/news/2023/04/09/from-discord-to-4chan-the-improbable-journey-of-a-us-defence-leak/

[6] “Eyes on russia map.” [Online]. Available: https://eyesonrussia.org/

[7] S. Eoghan, “Yes we can… but should we?” Oct 2020. [Online]. Available: https://osintessentials.medium.com/osint-investigation-online-verification-some-thoughts-on-ethics-267a84418895

[8] “Open source intelligence,” May 2022. [Online]. Available: https://data.europa.eu/en/datastories/open-source-intelligence

[9] “How has open-source intelligence influenced the war in ukraine?” [Online]. Available: https://www.economist.com/ukraine-osint-pod

[10] B. W. Watson, “”intelligence”,” Apr 2023. [Online]. Available: https://www.britannica.com/topic/intelligence-military

[11] “Libguides: Intelligence studies: Types of intelligence collection.” [Online]. Available: https://usnwc.libguides.com/c.php?g=494120&amp;p=3381426

[12] “Hc3: Analyst note,” Aug 2022. [Online]. Available: https://www.hhs.gov/sites/default/files/osint-how-to-analyst-note-tlpwhite.pdf

[13] J. Störger and F. Schaurer, “The evolution of open source intelligence (osint).” [Online]. Available: https://www.afio.com/publications/Schauer_Storger_Evo_of_OSINT_WINTERSPRING2013.pdf

[14] C. Colquhoun, “A brief history of open source intelligence,” Jul 2020. [Online]. Available: https://www.bellingcat.com/resources/articles/2016/07/14/a-brief-history-of-open-source-intelligence/

[15] Person, “Report: ’selfie’ helps air force find and destroy isis hq,” Jun 2015. [Online]. Available: https://www.boston25news.com/news/report-selfie-helps-air-force-find-and-destroy-isis-hq/8816369/

[16] C. C. Editor, “Red team - glossary: Csrc.” [Online]. Available: https://csrc.nist.gov/glossary/term/red_team

[17] “Mitre att&ck®.” [Online]. Available: https://attack.mitre.org/

[18] Dec 2019. [Online]. Available: https://www.imperva.com/learn/application-security/spear-phishing/

[19] “What is cyber threat intelligence? [beginner’s guide],” Mar 2023. [Online]. Available: https://www.crowdstrike.com/cybersecurity-101/threat-intelligence/

[20] Alienvault, “Open threat exchange.” [Online]. Available: https://otx.alienvault.com/

[21] “Comprehensive threat intelligence.” [Online]. Available: https://talosintelligence.com/

[22] “Virustotal.” [Online]. Available: https://www.virustotal.com/

[23] “Fighting malware and botnets.” [Online]. Available: https://abuse.ch/#about

[24] C. Kimeu, “On the street and online: Social media becomes key to protest in kenya,” Jul 2022. [Online]. Available: https://www.theguardian.com/global-development/2022/jul/12/on-the-street-and-online-social-media-becomes-key-to-protest-in-kenya

[25] P. Jessica McBride, “Finding social support through social media during covid lockdowns,” Jun 2022. [Online]. Available: https://today.uconn.edu/2022/06/finding-social-support-through-social-media-during-covid-lockdowns/#

[26] “Automatic dependent surveillance - broadcast (ads-b).” [Online]. Available: https://www.faa.gov/about/office_org/headquarters_offices/avs/offices/afx/afs/afs400/afs410/ads-b

[27] “Firms: Fire information for resource management system.” [Online]. Available: https://firms.modaps.eosdis.nasa.gov/map/

[28] “About us - maxar.” [Online]. Available: https://www.maxar.com/about

[29] A. France-Presse, “Before-and-after satellite imagery will track ukraine cultural damage, un says,” Oct 2022. [Online]. Available: https://www.theguardian.com/world/2022/oct/27/before-and-after-satellite-imagery-will-track-ukraine-cultural-damage-un-says

[30] M. Eckel, “Latest satellite imagery from maxar of russian troop deployments/movements,” Feb 2022. [Online]. Available: https://twitter.com/Mike_Eckel/status/1496356475702235141

[31] D. Murray, Y. McDermott, and K. A. Koenig, “Mapping the Use of Open Source Research in UN Human Rights Investigations,” *Journal of Human Rights Practice*, vol. 14, no. 2, pp. 554–581, 04 2022. [Online]. Available: https://doi.org/10.1093/jhuman/huab059

[32] H. Willis, T. Gibbons-neff, A. Toler, C. Triebert, J. E. Barnes, and M. Browne, “F.b.i. arrests national guardsman in leak of classified documents,”

Apr 2023. [Online]. Available: https://www.nytimes.com/2023/04/13/world/documents-leak-leaker-identity.html

[33] "What is shodan?" [Online]. Available: https://help.shodan.io/the-basics/what-is-shodan

[34] Gitleaks, "gitleaks/gitleaks: Protect and discover secrets using gitleaks." [Online]. Available: https://github.com/gitleaks/gitleaks

[35] B. Toulas, "Toyota discloses data leak after access key exposed on github," Oct 2022. [Online]. Available: https://www.bleepingcomputer.com/news/security/toyota-discloses-data-leak-after-access-key-exposed-on-github/

[36] smicallef, "smicallef/spiderfoot: Spiderfoot automates osint for threat intelligence and mapping your attack surface." [Online]. Available: https://github.com/smicallef/spiderfoot

[37] vysecurity, "vysecurity/linkedint: Linkedin recon tool." [Online]. Available: https://github.com/vysecurity/LinkedInt

[38] "Censys." [Online]. Available: https://about.censys.io/

[39] "Fofa search engine." [Online]. Available: https://en.fofa.info/about/en

[40] "Leakix - about." [Online]. Available: https://leakix.net/about

[41] "Snapchat - snap map." [Online]. Available: https://map.snapchat.com/

[42] michenriksen, "michenriksen/birdwatcher: Data analysis and osint framework for twitter." [Online]. Available: https://github.com/michenriksen/birdwatcher

[43] C. Boyd, "5 years for swatter who caused a man's death for a twitter handle." [Online]. Available: https://www.malwarebytes.com/blog/news/2021/07/5-years-for-swatter-who-caused-a-mans-death-for-a-twitter-handle

[44] J. Wilson, "Doxxing, assault, death threats: The new dangers facing us journalists covering extremism," Jun 2018. [Online]. Available: https://www.theguardian.com/world/2018/jun/14/doxxing-assault-death-threats-the-new-dangers-facing-us-journalists-covering-extremism

[45] M. Cramer, “A grandfather died in “swatting” over his twitter handle, officials say,” Jul 2021. [Online]. Available: https://www.nytimes.com/2021/07/24/us/mark-herring-swatting-tennessee.html

[46] M. Brice-Saddler, A. Selk, and E. Rosenberg, “Prankster sentenced to 20 years for fake 911 call that led police to kill an innocent man,” Mar 2019. [Online]. Available: https://www.washingtonpost.com/nation/2019/03/29/prankster-sentenced-years-fake-call-that-led-police-kill-an-innocent-man/

[47] Oct 2022. [Online]. Available: https://www.liferaftinc.com/blog/5-cognitive-biases-that-could-affect-your-osint-investigations

[48] E. Sweeney, “About - osint essentials.” [Online]. Available: https://www.osintessentials.com/about

[49] T. Bezmalinovic, “This is how much meta is investing in vr, ar and horizon,” Nov 2022. [Online]. Available: https://mixed-news.com/en/this-is-how-much-meta-is-investing-in-vr-ar-and-horizon/