




Article

# Mapping Tools for Open Source Intelligence with Cyber Kill Chain for Adversarial Aware Security

Muhammad Mudassar Yamin <sup>1</sup>, Mohib Ullah <sup>1,\*</sup>, Habib Ullah <sup>2</sup>, Basel Katt <sup>1</sup> , Mohammad Hijji <sup>3</sup>   
and Khan Muhammad <sup>4,\*</sup> 

<sup>1</sup> Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, 2815 Gjøvik, Norway; muhammad.m.yamin@ntnu.no (M.M.Y.); basel.katt@ntnu.no (B.K.)

<sup>2</sup> Faculty of Science and Technology, Norwegian University of Life Sciences, 1430 Ås, Norway; habib.ullah@nmbu.no

<sup>3</sup> Industrial Innovation and Robotic Center (IIRC), University of Tabuk, Tabuk 47711, Saudi Arabia; m.a.hijji@gmail.com

<sup>4</sup> Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), Department of Applied Artificial Intelligence, School of Convergence, College of Computing and Informatics, Sungkyunkwan University, Seoul 03063, Korea

\* Correspondence: mohib.ullah@ntnu.no (M.U.); khan.muhammad@ieee.org (K.M.)

**Abstract:** Open-source intelligence (OSINT) tools are used for gathering information using different publicly available sources. With the rapid advancement in information technology and excessive use of social media in our daily lives, more public information sources are available than ever before. The access to public information from different sources can be used for unlawful purposes. Extracting relevant information from pools of massive public information sources is a large task. Multiple tools and techniques have been developed for this task, which can be used to identify people, aircraft, ships, satellites, and more. In this paper, we identify the tools used for extracting the OSINT information and their effectiveness concerning each other in different test cases. We mapped the identified tools with Cyber Kill Chain and used them in realistic cybersecurity scenarios to check their effectiveness in gathering OSINT.

**Keywords:** OSINT; cybersecurity; AI; Cyber Kill Chain; public information misuse

**MSC:** 68M25



**Citation:** Yamin, M.M.; Ullah, M.; Ullah, H.; Katt, B.; Hijji, M.; Muhammad, K. Mapping Tools for Open Source Intelligence with Cyber Kill Chain for Adversarial Aware Security. *Mathematics* **2022**, *10*, 2054. <https://doi.org/10.3390/math10122054>

Academic Editor: Todor Tagarev

Received: 4 April 2022

Accepted: 18 May 2022

Published: 14 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

San Tzu [1] narrated that, "If you know the enemy and know yourself, you need not fear the result of a hundred battles". In the past, battles were fought by swords and arrows; however, humans found a more efficient way of killing each other with the advancement of civilisations. Guns and missiles replaced swords and arrows. However, one thing remained the same in ever-evolving weapons technology; all the adversaries compete for resources.

The battlespace for these adversaries is now changed from surface and subsurface warfare to cyberspace warfare. In any warfare, access to information plays a vital role. The information helps the war planners to make intelligent decisions against their adversaries to obtain an advantage over them.

Due to the digital transformation of society, warfare is transformed from the physical space into a hybrid space. This hybrid space warfare contains the cyber and physical aspects of warfare. Actions in cyberspace now have direct or indirect consequences in the physical space. Thus, dominance over cyberspace is now contested among multiple entities who attempt to dominate each other. For establishing this dominance, the adversaries attempt to gather as much information about each other as possible.

This information gathering may involve active measures in which adversaries interact with each other in the form of scanning and exploitation. The adversaries can use passive

means using OSINT tools to identify an adversary weaknesses from an attacker preservative. These tools can also be used to identify weaknesses in one's defence capabilities, leading to vulnerabilities and exploitation in a cyber engagement.

Researchers are focusing on such OSINT capabilities and conducting studies to better understand the topic. One study in 2022 [2] attempted to explain different concepts present within OSINT domain and attempted to explain its importance with other intelligence techniques. The study continued with the usage of OSINT in cyber space and the challenges and opportunities it presents.

According to researchers [2], OSINT provides a massive amount of information, and processing it is a difficult task; moreover, the data sources of OSINT need to be validated before making a conclusion. Another study in 2022 [3] focused on the military application of OSINT to detect information leakages. The researchers performed an experiment with different tools and searched for keywords related to the military for the identification of relevant information.

This kind of study focuses on the practical usage of OSINT for specific purposes; however, there is a need for mapping such tools in the overall execution of an attack in cyber domain. In this work, we analyse multiple state-of-the-art OSINT tools from an adversarial perspective. We analyse the tools both from the perspective of an attacker and defender simultaneously.

In order to identify the usage of the OSINT tools at different stages of cyber attacks and defences, the Lockheed Martin [4] Cyber Kill Chain is used to map the tools. The Cyber Kill Chain is used to indicate different stages of an attack that can be leveraged to develop an action plan to stop an ongoing attack.

The rest of the paper is organized in the following sections. We present the related works about the usage of the OSINT tools in the cyber domain in Section 2. We present the methodology of our work in Section 3. The usage of different OSINT tools with respect to seven stages of the Lockheed Martin [4] Cyber Kill Chain is described in Section 4. We also mention some of the cyber incidents that are attributed to the usage of the OSINT tools in Section 5. We highlight some of the countermeasures against the OSINT tools in Section 6. We discuss recent works based on artificial intelligence and machine learning in Section 7. We then discuss and conclude our work in Section 8.

## 2. Related Works

Many surveys and review works have been published in the past. Therefore, this section provides a brief overview of the works related to OSINT tools and other investigations in the field. Tabatabaei et al. [5] conducted a detailed survey about the OSINT tools in the context of cybersecurity. The survey focused on the tools and methods that are used for cybercrime investigation. They first categorised different cybercrimes, such as financial crimes, cyber vandalism, cyberstalking, etc. Then, they identified their investigation methods using the OSINT techniques. The investigation methods they identified were data mining, text mining, social network analysis, and optimisation methods (based on game theory). They identified the tools that implement those methods for collecting, storing, and classifying open-source data. They also shared open-source datasets where they are applicable.

Revell et al. [6] conducted a survey on tools for OSINT-based investigations. They focused on applications, websites, and services that the OSINT practitioners are using. They identified different OSINT tools and used their developed assessment framework for security, reliability, and legality for cyber investigations. Their assessment framework consisted of document information, supplier assessment, external assessment, and practitioner's assessment. In the document information, they assessed the tool's usefulness and traceable characteristics. The Supplier Assessment assessed the claims, legal terms, and policies usage. In the external assessments, the reviews about the tools from the external users, support, maintenance, vulnerabilities are assessed.

The Practitioner's Assessment presented the assessment in terms of reliability, anti-virus response, and data content of tools. Moreover, to address the detection of offensive

cyber operations using known public sources Tagarev et al. [7] defined five fundamental rules for the management of cybersecurity systems. The rules assist in formulating cybersecurity policy. Khanna et al. [8] analysed tools used for the doxing attacks [9]. In the doxing attacks, attackers use publicly available information from social media websites to threaten, harass, or embarrass the victims.

They proposed methods to protect organisations against such attacks. They focused on the Maltego tool [10] to identify an organisation structure and the social media footprint of its employees. They identified the limitations of the Maltego tool from the perspective of an attacker. For example, what type of data it can retrieve from social media and what type of data is restricted. They also introduced a method to deal with the doxing attacks, which included proactively doxing yourself, implementing in-depth defence principles, establishing demilitarised zones for organisational networks, using password managers and communicating using secure protocols, such as TOR [11].

Similarly, Tagarev et al. [12] identified four important dimensions for developing and characterizing cyber attack scenarios. Among them, the most important are the vulnerabilities exploitation and the capabilities of the malicious actors. He et al. [13] developed a cybersecurity framework for Connected and the Autonomous Vehicles (CAVs). The framework is based on unified modeling language (UML) and assist in classifying the vulnerabilities in the CAV system. For classification, the authors used classical machine learning classifiers, such as naive Bayes and decision trees. Jang et al. [14] worked on malware classification for cybersecurity and proposed an image-based malware classification algorithm that leverages local feature visualization techniques. For the local features, opcodes and API functions names that are extracted from the malware were used.

Hayes et al. [15] studied the usage of the OSINT tools for risk assessment. They used OSINT tools to analyse an organisation working in the critical infrastructure domain. They identified key individuals working in the organisation with respect to their job titles and roles. They identified the employees family, friends, and political views on social media, to assess the risk associated with insider threats. They also identified the majority of the network infrastructure using the OSINT tools and established that the Windows operating system is the dominant operating system in the organisation network. They identified the associated vulnerability with the operating system and recommended security patches. They concluded that organisations should be proactive in identifying their OSINT footprint to reduce the risk associated with unintentional information leakage.

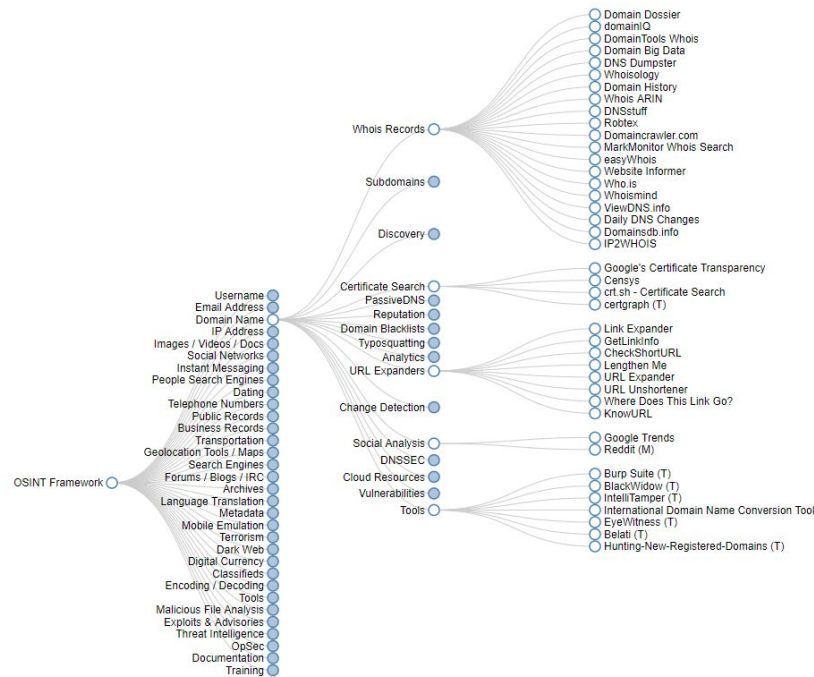
Xu et al. [16] studied network interconnection and the potential of cybersecurity breaches. They come up with a mathematical model that assist in allocating the budget constraints. Factors, such as the system interconnection, and the size of information sets are considered in the constraints. Justin Nordine [17] investigated the freely available OSINT tools. He categorised them into 32 main categories, including considering user names, email addresses, domain names, and IP addresses, to name a few. He then presented different tools with respect to each category. He further classified the tools that required registration or were downloaded locally on a machine to run.

The OSINT framework is presented in Figure 1. Most of the tools presented are web-based that contain data or metadata related to different IT-based services. These services contain information ranging from cyber vulnerability search engines to leaked databases. It provides cyber attackers with the opportunity to achieve their objectives in a very passive manner. On the other hand, for cyber defenders, these tools provide the opportunity to map their organisation attack surface and identify the organisational data available in public data breaches to secure their organisation proactively.

Quse et al. [18] analysed the effect of OSINT on 27001. The researcher attempted to demonstrate the effectiveness of OSINT by utilizing different tools and gathering massive amount of data for identifying key insights that can be utilized in a cyber operation. According to the researcher, ISO 27001 had an impact on the security of the system.

However, there are many sources that are available containing public information that make the overall security of the system weaker, which is not ideal. Therefore, the researcher

proposed their own tools for detecting such information and mitigating such weaknesses before a cyber attack. Similarly, Pieterse et al. [19] focused on situational awareness using OSINT tools. The researchers employed OSINT tools to gather information from media sources and analysed them to obtain an overall picture of an incident. The researchers stated that finding relevant information from a massive data source is difficult and there is a great deal of automation needed to make intelligent decision making based on the gathered information.



**Figure 1.** The OSINT security framework [17]. It represents a collection of OSINT tools to make intel and data collection tasks easier. The framework can be exploited by the information security researchers and testers for digital footprinting and intelligence information gathering.

These works lack much insight into the OSINT field and they focus on particular phase of Cyber Kill Chain. On the contrary, in our work, we cover different phases and aspects of Cyber-Kill chain to analyse the OSINT field in-depth. This will provide us better understanding of attacker and defender capabilities to ensure implementation of defence in depth paradigm. Such in depth analysis will make organisation and individuals more resilient against cyber attacks and ensure overall improvement of the cybersecurity ecosystem. An overview of different surveys conducted on OSINT and their difference is presented in Table 1.

**Table 1.** OSINT survey paper overview.

Reference Survey	Year	Theme	Remarks
Mapping tools for open source intelligence with Cyber Kill Chain for adversarial aware security (Our)	2022	Understanding OISNT tools utilization for attackers and defenders with respect to Cyber Kill Chain	Identification of OSINT tools with different stages of Cyber Kill Chain. Experiments for demonstrating their capabilities, highlighting challenges and future opportunities
Current Status and Security Trend of OSINT [2]	2022	General overview of OISNT in cybersecurity	OSINT overview, OSINT definition, Current challenges, Future direction

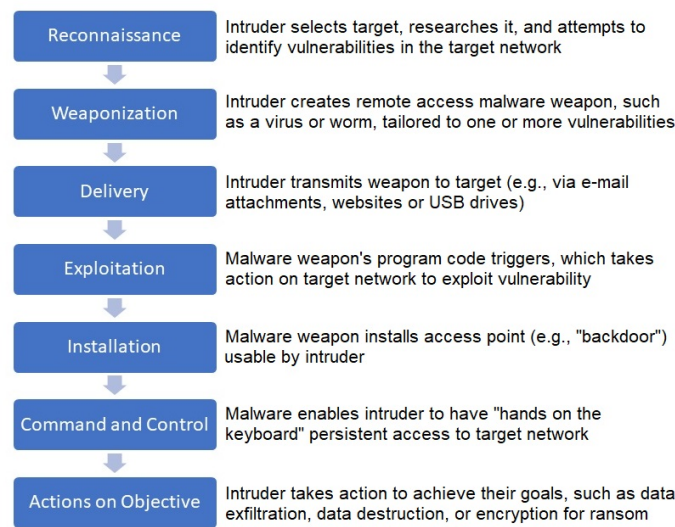
Table 1. Cont.

Reference Survey	Year	Theme	Remarks
How national CSIRTs leverage public data, OSINT and free tools in operational practices: An empirical study [20]	2022	OSINT data usage in cybersecurity incident response process	Qualitative interviews from incident response professionals, Empirical case studies, Challenges with OSINT tools and techniques
A survey exploring open source intelligence for smarter password cracking [21]	2020	OSINT data usage for forensic analysis and its legal implications	Identification of OSINT tools for password cracking. Legal and ethical considerations for using such tools.
Cyber Intelligence and OSINT: Developing Mitigation Techniques Against Cybercrime Threats on Social Media [22]	2018	Analysing social media data using OSINT to address criminal activity	Social media analysis, Physical crime activity investigation, Future challenges in using such techniques and data
OSINT in the Context of Cyber-Security [5]	2016	General overview of OSINT in cybersecurity	OSINT overview, Cyber threats terminology and classification, Detection and prevention of cyber threats
Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT) [23]	2012	General overview of OSINT	The evolution of OSINT, Emerging trends and suggestions for OSINT data utilization
Sailing the Sea of OSINT in the Information Age [24]	2009	General overview of OSINT	Opportunities OSINT provides, challenges in OSINT adoption, Potential use cases in intelligence collection

### 3. Methodology of Our Work

In order to identify the relevant literature in the field, we used a keyword-based search introduced by Jesson et al. [25]. We used Google scholar database as it contains research articles from indexed publication channels, including but not limited to IEEE and ACM publishers. We used the keywords OSINT, Information Security, and Tools. The search scope was limited to papers published between 2016 and 2021. The search resulted in 365 articles published in different venues. We considered only peer-reviewed articles in the research to ensure the quality of results. We only included those articles about OSINT tools in the context of information security.

We categorized the tools with respect to the domain from which they extract information, such as social media, vulnerability scanners, location-based services, etc. We then explored the Lockheed Martin [4] Cyber Kill Chain to map the usage of identified tools in the context of the cyber domain. The Cyber Kill Chain is an intelligence-driven framework employed to identify the intrusion activity in a cyber operation and defines the steps an attacker should take to achieve the attack objectives. The model defines seven steps, which are presented in Figure 2. The cybersecurity incidents in which the tools were applied are identified using credible news sources and presented in the paper for corresponding steps in the Cyber Kill Chain [4].



**Figure 2.** Cyber Kill Chain is a cybersecurity model that discovers the phases of a cyber attack, recognizes vulnerabilities, and supports security people to encounter the attacks at every step of the chain.

#### 4. OSINT Tools with Respect to Cyber Kill Chain

##### 4.1. Reconnaissance

###### Digital Footprints

Digital footprints contain information about organizations and individuals publicly available over the Internet. This information is collected by various methods, such as job posts, social media activity, and accidental leaked information. In a cyber attack, attackers first need to perform reconnaissance on the target, and active and passive manners can achieve this. In an active reconnaissance, an activity attacker needs to interact with organizational infrastructure and workforce using port scanners and social engineering tactics. In passive reconnaissance, an attacker uses digital footprinting techniques on an organization to identify organizational infrastructure and workforce information, which can be used for the following stages of a cyber attack.

- **IT Infrastructure**  
IT infrastructure contains the details of domains and server information on which an organization hosts its services, such as websites and application interfaces. Attackers attempt to perform reconnaissance on this infrastructure to identify weak points that they can use for launching an attack. These weak points contain information related to domain registry, hosting servers, etc. Different OSINT tools are developed to retrieve this information, which are presented in Table 2.

**Table 2.** Different tools used to obtain the OSINT information affiliated with the IT infrastructure of an organization.

ID	Name	Description	Reference
1	DNS Dumpster	Web-based tool used identify IT Infrastructure from DNS Records.	[26]
2	domainsdb	Web-based registry of domain names.	[27]
3	ip2whois	Web-based tool for WHOIS look up from given IP address.	[28]
4	recon-ng	Python-based tool used for comprehensive cyber reconnaissance purposes.	[29]

- **Archives**  
Organizations often upload sensitive data over the internet with human negligence or by mistake. They remove the data as soon they realize the sensitive nature of their data. However, services over the internet periodically archive the information present on different publicly available web applications and store them in publicly available databases. Cyber attackers use the information from those publicly available databases to identify important information and changes in organizational infrastructure for devising attack strategies. Different OSINT tools are developed to retrieve this information, which is presented in Table 3.

**Table 3.** Tools used to obtain the OSINT information from archived data on publicly available sources.

ID	Name	Description	Reference
1	WayBackMachine	Web-application-based website archiving service.	[30]
2	weleakinfo	Web application based service that contains information of various data breaches.	[31]
3	cryptome	Web-application-based service that is used to archive important government documents.	[32]

- **Social Media**  
Employees working in an organization mostly use some kind of or another type of social media, depending upon their personal preferences. Their social media footprints can be used to identify the organization's day-to-day activities or can be used to influence employees' behaviour within the organization through social engineering to achieve cyber-attack objectives. There are multiple services used to gather data of employees working in an organization for the advantage of the attacker, which are presented in Table 4.

**Table 4.** Tools used for OSINT on social media.

ID	Name	Description	Reference
1	Raven	A Python-based application that is used to identify information of an organizational employees using linkedin.	[33]
2	Social Searcher	A web application that can search for all public activity of a user in different social media websites.	[34]
3	Social media Monitoring Wikki	A list of paid and free social media monitoring solutions.	[35]

#### 4.2. Weaponization

##### Cyber Vulnerabilities Information

After identifying the organisation footprints, attackers attempt to find vulnerabilities in its infrastructure and workforce for exploitation. These vulnerabilities can be identified by actively engaging with organisational infrastructure using vulnerability scanners or can be identified by passive means using the OSINT techniques. Actively engaging the organisation infrastructure with vulnerability scanners leaves footprints and alerts the organisation that somebody is attempting to identify vulnerabilities in its infrastructure. The OSINT tools provide the opportunity to gather vulnerability information passively. This is achieved by gathering organisational information from an online vulnerability search engine and then identifying working exploit based upon the gathered information [36].

- Vulnerability Search Engines**  
 Currently, multiple organisations actively scan the internet to identify the live device and its open ports. Some organisations use this information to provide products, such as vulnerability search engines. These vulnerability search engines contain information, including the operating systems and running services version, which is openly available. Cyber attackers can use this information to identify vulnerable operating systems and services running on an organisation network for exploitation. Different OSINT tools are developed to retrieve this information presented in Table 5.

**Table 5.** Tools used for OSINT on vulnerability search engines.

ID	Name	Description	Reference
1	Shodan	Shodan is a mass vulnerability scanner that scan the internet for connected devices and their vulnerabilities.	[37]
2	Censys	Similar to Shodan a mass vulnerability scanner.	[38]
3	GreyNoise	It collects data from multiple mass scanners and correlate the information with known botnets and malware for identification of vulnerable and exploited devices.	[39]

- Vulnerability Databases**  
 After the identification of open ports and services, cyber attackers attempt to identify whether those ports and services are vulnerable or not. Many government and private organizations maintain the list of vulnerable software and services in order to provide vulnerability information to unprotected organizations. Cyber attackers use this information for their advantage and map the identified ports and service information from vulnerability search engines with the information they retrieve from vulnerability databases. Different OSINT tools are developed to retrieve this information, which are presented in Table 6.

**Table 6.** Tools used for OSINT on vulnerability databases.

ID	Name	Description	Reference
1	VULN DB	Community driven vulnerability database, which contains vulnerabilities from 1970s to now.	[40]
2	NVD	United states government national vulnerability database.	[41]
3	CNNVD	Chinese government national vulnerability database.	[42]
4	FS TEC	Russian government national vulnerability database.	[43]

- Exploit Databases**  
 Identifying a vulnerability in one thing and successfully exploiting that vulnerability is an other thing. Security researchers often publicly post working exploits for vulnerable applications and services. The purpose of sharing such information is to provide for proof of working exploit for a vulnerability. This information is used by the attackers for their advantage. The publicly posted exploits are curated by multiple organizations over the internet, which are mostly freely available and are presented in Table 7.



**Table 7.** Tools used for OSINT on exploit databases.

ID	Name	Description	Reference
1	Exploit DB	Free open source exploit publicly available on a web application.	[44]
2	Rapid7 Exploit DB	Free and paid exploit publicly available on a web application.	[45]
3	0 day Today	Paid 0 day exploit publicly available on a web application.	[46]

#### 4.3. Delivery

##### Usernames and Email Addresses

Phishing is one of the most widely used method to deliver malware and exploit an organization via email and other communication methods. The OSINT provides the attackers the tools and capability to enumerate organizational email addresses and identify the individuals working in an organization. After identifying employees email addresses and social media accounts, attackers use deceptive techniques in order to deliver the exploit to the organization. Organizations require security awareness to tackle such kind of attacks. However, in most cases, the employees make mistakes, which results in the successful delivery of the exploit.

- **Email Addresses**  
Information related to organizational email addresses is important. Attackers use this information to launch spear phishing attack for the delivery of malware and the achievement of there objectives. Spear phishing reacquires detail knowledge of organizational hierarchy, which OSINT tools and techniques provide. With OSINT, cyber attackers are able to identify internal hierarchy of organization that is who is reporting to whom. Therefore, the cyber attackers can launch a targeted attack on a vulnerable victim. Many tools are available to identify, verify, and retrieve organizational email addresses from public web and data breaches, which are presented in Table 8.

**Table 8.** Tools used for OSINT on email addressees.

ID	Name	Description	Reference
1	hunter	Online web application that can be used to identify corporate email addresses.	[47]
2	haveibeenpwned	Online web application that can be used to identify email addresses that are compromised during different data breaches.	[48]

- **Usernames**  
Similar to email addresses, Usernames of organizational employees are also considered important information for attackers. Usernames can reveal organizational employees activity in technical forums where they discuss technical issue, which may be important for cyber attackers. Moreover, those usernames also disclose the employees activities on social media and other digital media, which may assist cyber attackers in achieving their objectives. The OSINT tools provide the capability to cyber attackers to identify the usernames of different organizational employees on social media and web platforms. Some of them are presented in Table 9.

**Table 9.** Tools used for OSINT on usernames.

ID	Name	Description	Reference
1	namechk	A web and desktop utility that can search for given user name on different social media.	[49]
2	thatsthem	A web application that can be used to identify people details based upon their name in United States.	[50]
3	usersearch	A web-application-based user search engine.	[51]

- **Default Passwords**

Most of the IT equipment is shipped with default configurations. The default configurations contain default access and management accounts with default passwords. Due to large size of organizations and human negligence, some IT equipment within the organization operates with these default configurations. Attackers exploit this issue for their advantage and use default passwords to access organizational equipment in order to install malware or achieve their objectives. There are many online services that are used to check the default configurations of IT equipment. However, cyber attackers use that information for their advantage using the OSINT methods. Some of the tools that use to retrieve this information are presented in Table 10.

**Table 10.** Tools used for OSINT on default passwords.

ID	Name	Description	Reference
1	Default Password Db	Web application that list the default password of many applications and devices.	[52]
2	Default-password	Similar to Default Password Db but application and devices are sorted in alphabetical order.	[53]
3	Default Password Lookup	Similar to Default Password Db but default password are sorted with respect to companies and organizations.	[54]
4	routerpasswords	Web application that contains the default password of many network routers.	[55]

#### 4.4. Exploitation

##### File Analysis

Cyber attackers mostly use malware to exploit organizations and achieve their objectives. These malware exploit known vulnerabilities and have an identifiable execution behaviour and signatures. Cyber defenders use the information of their signatures in order to proactively defend their organizations. In order to avoid detection by cyber defenders, cyber attackers attempt to develop new malware that has unknown behaviours and signatures. To test the developed malware, attackers use freely available online services, which have the capability to analyse the malware infected files using different industry leading antivirus and anti-malware software and solutions.

- **Office file analysis tools**

Microsoft Office files are commonly used by cyber attackers to deliver the malicious exploitation payload to the target organization. The malicious payload usually contains macro or vulnerability exploits within Microsoft office software, which can lead to target system exploitation when executed. Therefore, Office files from suspicious sources need to be analysed before they are opened in a system. This analysis can be performed by various online tools, which can identify the exploit signature from code and behaviour. These tools open the Office files in a sandbox environment for the identification of malicious activity within a test system and then give a rating on

how malicious the Office file is. Some of the tools that use to retrieve this information are presented in Table 11.

**Table 11.** Tools used for OSINT on Microsoft Office file analysis.

ID	Name	Description	Reference
1	reconstructor	Different desktop based tools that are used to specially analyse Office files of different formats, such as doc, docx, rtf, and ppt.	[56]
2	Office File Analysis Tool	Microsoft Office file analysis tool help researchers to analyse Office files locally at a binary level.	[57]

- **Malware Analysis Tools**

Similar to Microsoft Office files, other executable files that come from suspicious sources are needed to be analysed in a secure manner. Freely available online tools provide this surface in order to gather signature of new malware and exploits. Cyber defenders can upload the suspicious looking executable files on these tools to scan the files using multiple antivirus and anti-malware solutions in order to identify known malware signatures and behaviours in those files. Details of some of the malware file analysis services are given below.

#### 4.5. Installation and Execution Environment

After the development of malware, the malware should be tested in a similar execution environment in which it is planned to be executed. After collecting information of organization IT infrastructure environment, emulation of parts of the target environment is required to fine-tune the working of the delivered malware. Multiple open source and free tools are available for cyber attackers disposal for emulating organisation infrastructure. Cyber attackers first develop the malware and then test and verify the functionality of that malware in the emulated environment, before the final delivery to the target organization.

- **Virtualization Platforms**

Different virtual platforms are available to test the functionality of various malware, exploits, and vulnerabilities. The virtual platforms provide the capability to emulate different operating systems ranging from desktop to Internet of Things (IoT). They also provide environment for testing of different applications for identification and exploitation of vulnerabilities. Some of these platforms cost money, however, free alternative solutions are also available. For testing of different software applications, multiple paid and free software libraries are available, from where sample of software can be download for testing. The details of virtualization platforms and software libraries are presented in Table 12.

**Table 12.** Tools used for the creation of a virtualization environment.

ID	Name	Description	Reference
1	Vmware	Commercial VM creator for different operating systems with free trial.	[58]
2	Virtual Box	Free VM creator for different operating systems.	[59]
3	EdgeHTML	Browser emulation tool from Microsoft.	[60]

- **Mobile applications and services emulation**

Smartphones have become the primary source of product and services consumption by majority of people. Therefore, cyber attackers are more active to exploit vulnerabilities in smartphone and mobile applications than ever before. The major smartphone

operating systems that dominate the market are IOS and Android. For testing vulnerabilities and developing exploits for IOS, attackers need to have actual devices. The IOS development environment XCODE provides limited application emulation capabilities. On the other hand, multiple third party solutions are available to emulate Android operating systems and applications. Most of them are freely available and are presented in Table 13.

**Table 13.** Tools used for mobile applications and services emulation.

ID	Name	Description	Reference
1	Genymotion	desktop based Android OS emulator.	[61]
2	Blustacks	desktop based Android OS emulator.	[62]
3	Andyroid	desktop based Android OS emulator.	[63]
4	NoxPlayer	desktop based Android OS emulator.	[64]

#### 4.6. Command and Control

##### Compromise Notification Services

After successfully exploitation, cyber attackers set up command and control mechanisms to persistently control the victim machine in the target organization. These command and control setups often use communication channels that are considered benign in order to avoid detection. With use of the OSINT techniques, it is possible to identify the command and control channels that are employed by cyber attackers. There are services that allow notification about compromised systems anonymously and publicly. These services can be used by cyber defenders in order to secure the compromised systems or by the cyber attackers for launching further attacks on the target organization.

- Security compromise notification  
Cybersecurity researchers or cyber criminal can voluntarily notify about comprised systems in publicly accessible forums and web platforms in an anonymous manner as seen in Figure 3. They do it in order to boost their reputation within the cybersecurity community. In term of the OSINT perspectives, this information can represent a high value because of the identification of vulnerable systems that are exploitable. With this information, it is even possible to identify the scale of cyber attacks by correlating the information of compromised system with respect to the notifier in order to obtain information about mass compromise or a single isolated cybersecurity compromise event. The details of compromise notification websites and services are presented in Table 14.

**Table 14.** Tools used for OSINT on security compromise notification.

ID	Name	Description	Reference
1	Zone-H	Online web application that is used to anonymously disclose compromised websites.	[65]

- Malware-tracking services  
Some OSINT services provide malware tracking by sharing information of malware hashes and infected system details. They also curate the domains from which the malware is spreading. This is collected by information shared by malware researchers who identify the behaviour and communication channels of the malware. Utilizing this information in an evolving cyber situation is crucial. In fact, spreading malware can affect organizational systems within a certain amount of time. Therefore, identifying and blocking malware signature and communication channels in a proactive manner should be set as a priority. The details of the applications and services that provide this information are presented in Table 15.

**Table 15.** Tools used for OSINT for malware tracking.

ID	Name	Description	Reference
1	Ransomware tracker	Online website contains list of ransomware distribution domains.	[66]
2	Malware domain black list	Online website contains black listed domains that are spreading malware.	[67]
3	Malware patrol	Online platforms where thread data is shared, which is obtained by multiple public and propriety resources.	[68]

[ENABLE FILTERS]

Total notifications: 1,133 of which 104 single ip and 1,029 mass defacements

Legend:  
 H - Homepage defacement  
 M - Mass defacement (click to view all defacements of this IP)  
 R - Redefacement (click to view all defacements of this site)  
 L - IP address location  
 ★ - Special defacement (special defacements are important websites)

Time	Notifier	H	M	R	L	★ Domain	OS	View
14:07	/MrAxxxCT-					www.davdav.fr/maxxct.htm	Linux	mirror
13:55	xdawelax	H	M			settled.asia	Unknown	mirror
13:53	Family Attack Cyber		M			forkom-msk.ru/olala.htm	Linux	mirror
13:53	Family Attack Cyber		M			pkc-format.ru/olala.htm	Linux	mirror
13:53	Family Attack Cyber		M			forkom-spb.ru/olala.htm	Linux	mirror
13:23	3thical_b0y	H	M			bangkokpeop.com	Unknown	mirror
13:19	3thical_b0y	H	M			republicvietnam.com	Unknown	mirror
13:18	3thical_b0y	H	M			notciasdobrasil.com	Unknown	mirror
13:17	3thical_b0y	H				nagornokarabakhexchange.com	Unknown	mirror
13:08	SilentClown					como-Official.services/breaksh...	Unknown	mirror
12:50	ifactoryx		M			www.ivyeducation-us.com/rx.html	Win 2016	mirror
12:41	ifactoryx		M			crcy.org.mx/rx.html	Win 2016	mirror
12:41	ifactoryx		M			aba.world/rx.html	Unknown	mirror
12:41	ifactoryx		M			arte-lcg.tk/rx.html	Win 2016	mirror
12:41	ifactoryx		M			alvaraq-pub.com/rx.html	Win 2016	mirror
12:40	ifactoryx		M			1stunitedcargo.com/rx.html	Win 2016	mirror
12:40	ifactoryx		M			audiovisualesvale.com.ve/rx.html	Win 2016	mirror
12:40	ifactoryx		M			arte-lcg.com.mx/rx.html	Win 2016	mirror
12:40	ifactoryx		M			anaboliclabs.com/rx.html	Win 2016	mirror
12:40	ifactoryx		M			aplicaciones-moviles.com/rx.html	Win 2016	mirror
12:40	ifactoryx		M			bankstoreegypt.com/rx.html	Win 2016	mirror
12:40	ifactoryx		M			asaanpaisainvest.com/rx.html	Win 2016	mirror
12:40	ifactoryx		M			aventurayecoturismo.com/rx.html	Win 2016	mirror
12:40	ifactoryx		M			bombasyelectricidad.com/rx.html	Win 2016	mirror
12:40	ifactoryx		M			asaanpaisa.com/rx.html	Win 2016	mirror

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

**DISCLAIMER:** all the information contained in Zone-H's cybercrime archive were either collected online from public sources or directly notified **anonymously** to us. Zone-H is neither responsible for the reported computer crimes nor it is directly or indirectly involved with them. You might find some offensive contents in the mirrored defacements. Zone-H didn't produce them so we cannot be responsible for such contents. [Read more](#)

Home News Events Archive Archive ★ Onhold Notify Stats Register Login Disclaimer Contact

Attribution-NonCommercial-NoDerivs 3.0 Unported License

**Figure 3.** Website compromise disclosure [65].

#### 4.7. Actions on Objective

##### Threat Intelligence

Threat intelligence provides the organizations the capability to actively identify cybersecurity threats that are affecting other organizations. Whenever there is a cybersecurity incident, the organizations share the information of the incident with other organizations so that they can prepare themselves for similar incidents. This kind of corporation can use selective information sharing or open-source threat intelligence information sharing [69]. In selective-information-sharing schemes, only the organizations that are under legal obligation share the information with relevant bodies who decide weather to share the information or not. As with sharing information with a national-level computer emergency and response team (CERT) who then coordinates the disclosure of information with relevant stake holders.

In contrast to this, selective threat information sharing in open-source threat intelligence information sharing organization the information of incidents as they happen in public threat feeds. This is beneficial for organizations who want to address the cybersecurity threats as soon as possible. However, it also provides the opportunity to attackers to identify new vulnerabilities and exploits.

- **Indicator of compromise tools**  
Indicator of compromise (IOC) is a combination of virus, malware signatures and domain name, and IP addresses that are used to control the command and control channel of a botnet network. The IOC are usually identified by incident responders and computer forensics investigators during an ongoing cyber incident. Cyber defenders within an organization can use the IOC information to develop new intrusion detection and prevention rules in order to proactively defend their organization. Details of some of the IOC channels publicly available are presented in Table 16.

**Table 16.** Tools used for OSINT on indicator of compromise (IOC).

ID	Name	Description	Reference
1	IOC parser	Open-source Python-based tool that is used to extract IOC information from different local and web resources.	[70]
2	IOC extract	Open-source Python-based tool similar to IOC parser.	[71]

- **Tactics techniques and procedures**  
Tactics techniques and procedures (TTP) is a term used to identify the behaviour of an advanced persistent threat (APT). The APT threats are considered as state-sponsored cyber threat actors that use relatively advanced and unique cyber tactics in order to compromise the cybersecurity of an organization or a country. Multiple organizations monitor and share the activities of different APTs by different states. The organizations share the monitored information in regular reports in which they share the APT tactics techniques and procedures, and also provides the countermeasure against them. Detail of the OSINT information sources related to the TTP are presented in Table 17.

**Table 17.** Tools used for OSINT on adversary tactics techniques and procedures.

ID	Name	Description	Reference
1	Fire eye	fire eye threat research reports for APT, available free online.	[72]
2	MITRE	enterprise attack techniques and procedures.	[73]

- **Open-source threat feeds**  
Threat intelligence sharing among different organizations is finding traction now. This is due to the increase in cyber incidents. Cybersecurity is now considered as a collective responsibility. Therefore, increasingly organizations are sharing the cyber threat intelligence information. It contains the details of previous and ongoing cyber incidents. Different tools and platforms are developed in order to assist this information sharing process. These tools and platforms collect data from the organizations, anonymize the organization information and then share the information in publicly available threat feeds. The details of such tools and services are presented in Table 18.

**Table 18.** Tools used for OSINT for open-source threat feeds.

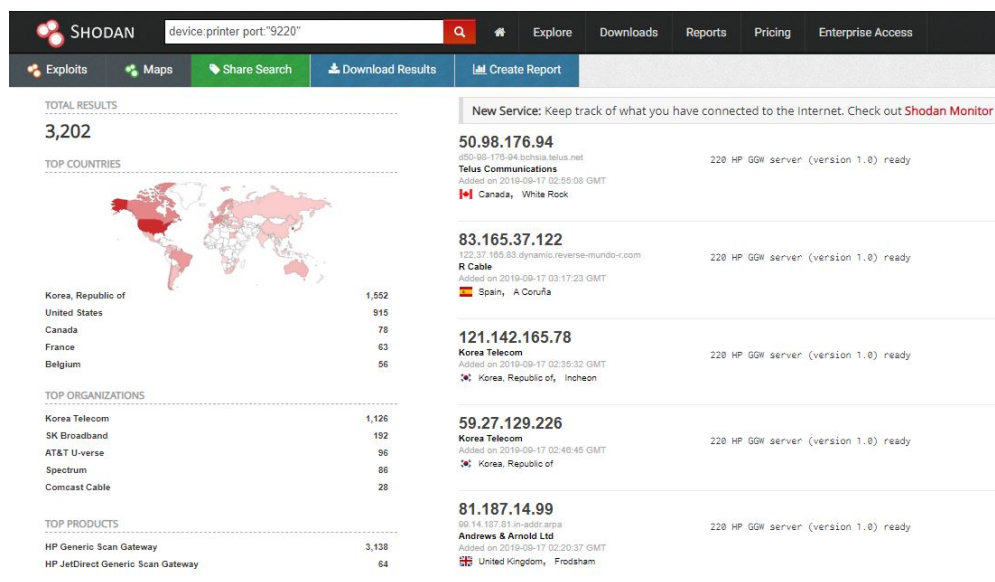
ID	Name	Description	Reference
1	XForce exchange	Free and pain threat feeds available by IBM.	[74]
2	MISP	Malware information-sharing platform, the information is shared in a structured format.	[75]
3	Threat exchange network	Enterprise-level threat-exchange network designed to assist during incident response.	[76]

### 5. Use of OSINT Tools in the Wild

Multiple cybersecurity incidents took place in the past by using different OSINT tools. To discuss the usage of the OSINT tools, some of the cybersecurity incidents are shared here. As many cybersecurity incidents happens every day, only those incidents are shared that are directly executed by the usage of freely available OSINT tools.

#### 5.1. Printers for Propaganda

In early December 2018, there was an intense competition between a Swedish Youtuber namely PewDiePie and Indian music company named T-Series for first place on Youtube in terms of subscribers count. During this competition, a hacker who was the supporter of PewDiePie was able to compromise 50,000 printers connected over the internet. The hacker printed propaganda to motivate people to subscribe to PewDiePie [77]. The hacker was able to achieve this by using a famous OSINT tool known as Shodan [78]. This tool can scan the internet for the identification of vulnerable connected devices. The attacker identified the IP addresses of the vulnerable printers, and then used a publicly known exploit [79,80] to print the propaganda material. The attack received a great deal of media attention and exploited the effectiveness of OSINT tools. In Figure 4, identifying vulnerable printers using Shodan is presented.



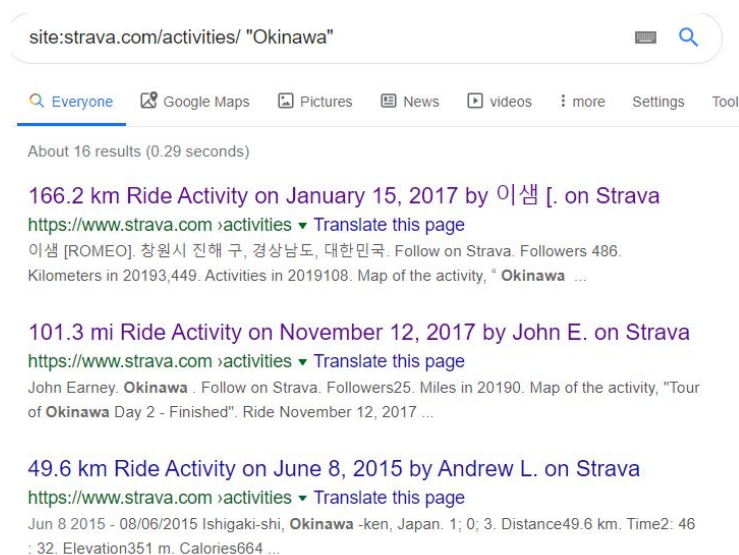
**Figure 4.** Identifying vulnerable printers on Shodan.

#### 5.2. Starva for Secret Locations

Starva is an activity-tracking application that is used to record exercise activities, such as running, cycling, and walking etc. It saves the activity data online presenting a map that indicates the places where the activities are performed. The application was popular in the United States military personal who use it for recording their exercise around different military bases.

In January 2018, it was reported [81] that the application was leaking the information of secret army bases in its activity maps. This happened because Starva is uploading the activity maps on publicly available URL addresses. These addresses were cached in the Google search engine. Anybody with sufficient knowledge of Google dorking can search the internet for identification of those publicly available URL addresses for retrieval of activity maps.

If we dig deeper into this information leakage, we can identify that the skill level requirement for finding this information is barely minimum. From a cyber attacker perspective, the attacker only had to utilize a technique known as Google dorking [44]. Using this technique, the attackers use specially crafted search queries for the identification resources with relevant information. In case of Starva, the public activities of military personals were indexed in the Google search engine and were accessible using simple Google dork techniques. To demonstrate the concept, the researchers used the Google dork technique to fetch available Starva activities on Island of Okinawa. This is the island where a large United States military base is located. The fetched data is presented in Figure 5.



**Figure 5.** Accessing activity maps on Starva.

### 5.3. Snapchat for Ship Tracking

Snapchat is an instant messaging application that uses photos as a primary mode of sharing information. It has a feature to upload photo and video status, which geotags the status on an interactive world map. Due to security and privacy concern, British navy banned its usage on its ships in early 2017 [82]. In a recent maritime cybersecurity exercise, it was identified that the location of maritime vessels under NATO command can be identified using social media services specially snapchat [83]. The issues with snapchat were already known by the military service. Therefore, this information leakage could be attributed to human negligence and lack of policy enforcement on critical navel assets.

To demonstrate the effectiveness of tracking military movement using Snapchat, we conducted their own experiment on Snapchat map feature. The experiment aimed to identify military assets using only Snapchat. The geotag of Snapchat status are publicly available over the internet in a browser. A cyber attacker can access them without even installing Snapchat application on a smartphone. We focused again on the geographical area of Japan where United States military is present in high numbers. The researcher identified multiple Snapchat status related to military movement. A sample of identified military movement is presented in Figure 6. The details of military equipment and number of personal involved is clearly visible. Such information is consider as high value information for intelligence reasons as it reveals operational status of a military in a particular region.



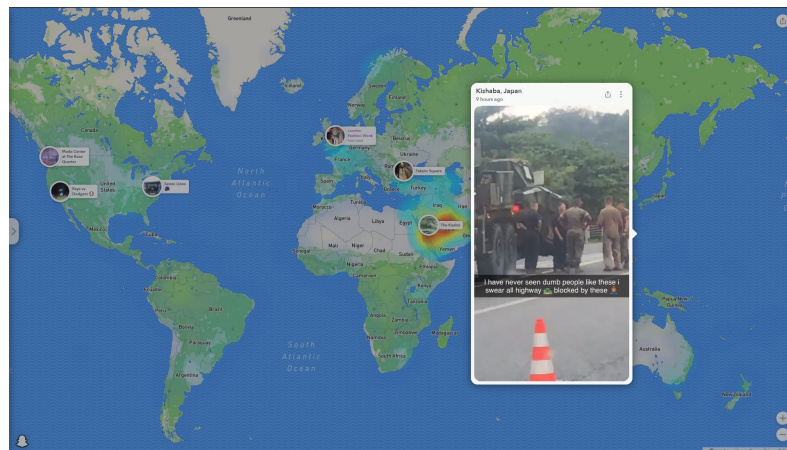


Figure 6. Accessing real-time military movement on Snapchat.

### 5.4. Generalized Scenario

Above examples are specific scenarios. However, in general we can model the OSINT scenarios with respect to the Cyber Kill Chain. The information flow in the Cyber Kill Chain can be used to represent the perspective of both an attacker and defender at the same time. A generalized attack and defence scenario is presented in Figure 7.

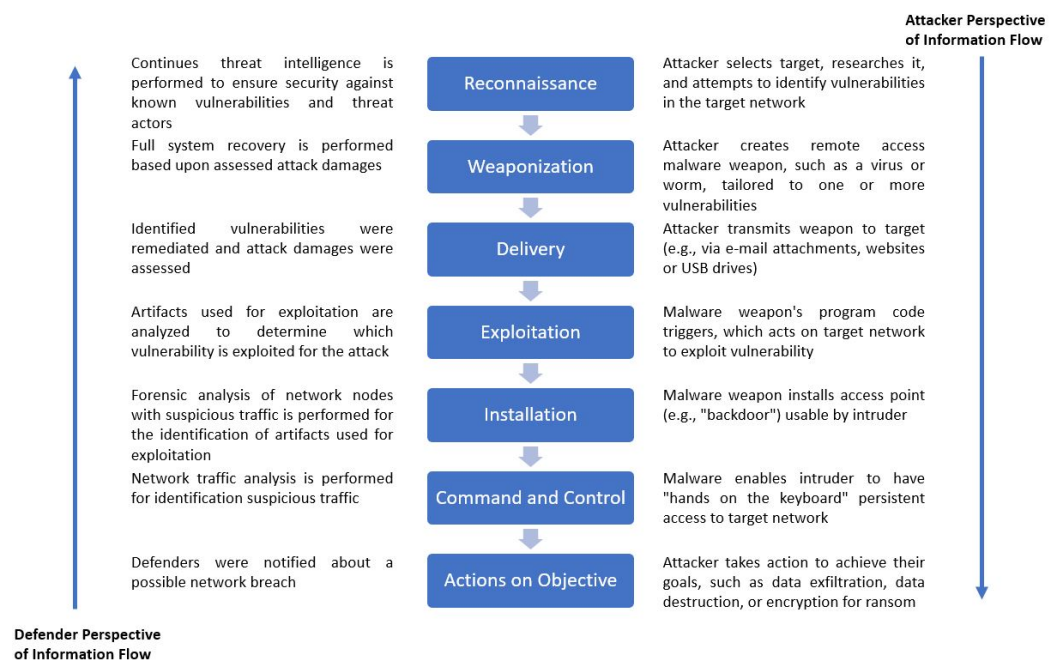


Figure 7. General attack and defence scenario using Cyber Kill Chain.

## 6. Countermeasures

### 6.1. OSINT Fingerprint Cleanup

Employees of organizations unintentionally or by lack of cybersecurity awareness use their organizational email addresses. Sometimes they also use the same passwords to make accounts for different digital services. Most digital services are vulnerable to cyber attack and data breaches. The cyber attack and data breaches reveals those organizational credentials, which can be retrieved by cyber attackers for launching attacks on the organization. This problem can be dealt with non-technical and technical means. Non-technical means include raising cybersecurity awareness using training and awareness session for organizational employees.

Technical measures include proactively identifying the uses of organizational credentials on different digital services and removing those credentials or any related information.

- **Digital fingerprint cleanup**  
Cleaning up digital fingerprints over the internet can be a difficult task. The user accounts of the employees of an organization can be deleted from various online web services and applications. However, removing the information from various search engines and web caching services may require sending some legal notices. Due to the complexity of this task, multiple applications and services were developed in order to facilitate the process of digital fingerprint cleanup. The details of such services and applications are presented in Table 19.

**Table 19.** Tools used digital finger print cleanup.

ID	Name	Description	Reference
1	Justdeletemr	Open-source web browser extension that list the details of multiple websites account deletion process.	[84]
2	Accountkiller	Online web application that assist in deleting accounts and canceling different subscriptions.	[85]

## 6.2. Avoiding Cyber Tracking

When the employees of an organization are surfing the web or using different web applications, their activity is tracked by different methods. This activity tracking is usually conducted for performance improvements and personalized services. However, cyber attackers can use this benign functionality for their advantage in different cyber attack scenarios. Therefore, employees of an organization should know the consequences of activity-tracking applications to avoid their usage. Some technical solutions that can help to avoid tracking by such web and mobile application are given below:

- **Anonymous browsing**  
Multiple services provide anonymous browsing capabilities. They mostly use TOR, VPN, and Proxy services in order to avoid tracking from internet services provider and different cyber adversaries. How effective these services are is questionable, due to the cyber vulnerabilities present within those services. However, its better safe then sorry, therefore, their usage is recommended while communicating over unsecured computer networks. It should also be noted that various APT and sophisticated hacker groups also setup these services in order to tap the information flowing in those services channels. The details of the services that are designed to provide anonymous browsing are presented in Table 20.

**Table 20.** Tools used for secure browsing.

ID	Name	Description	Reference
1	TOR	The onion routing protocol browser for anonymous browsing.	[86]
2	I2P	Invisible Internet project, similar to TOR, but involve multiple encryption schemes on all layers of communication.	[87]
3	locabrowser	Web-browser-based proxy to surf the internet from different location and by emulating different devices.	[88]
4	Epic privacy browser	Free browser developed by keeping user privacy in mind, it has built-in proxies for surfing the internet from multiple locations.	[89]

- **Secure operating systems**  
As stated in the above section, some of the services that are designed to provide anonymous browsing are actually being used by APT and hacker groups for tapping user information. Therefore, to tackle this issue, the concept of secure operating systems is gaining momentum. The secure operating systems are being designed from ground up by considering security and privacy principles from ground up. These operating systems are developed using freely available tools but are configured by security professionals in order to preserve user privacy and provide security against cyber adversaries. The details of such operating systems that are designed to preserve user privacy are presented in Table 21.

**Table 21.** Secure operating systems.

ID	Name	Description	Reference
1	Whonix	Open-source debian based Linux distro designed by keeping user security and privacy in mind.	[90]
2	Tails	Developed to preserve user privacy and anonymity.	[91]
3	Subgraph OS	It is developed to reduce the risk on endpoints and is resistant to network based attacks.	[92]

### 6.3. Cyber Imposter

One way of avoiding information gathering by cyber attackers using OSINT tools for an organizational employees is by posing as a imposter on digital media and hiding the true identity. This could be done for sourcing or hiring employees for sensitive projects and jobs, whose public advertisement may attract cyber attackers. Variety of tools are developed for hiding the identify of individuals on digital media, which allows organizational employees to operate safely within digital domains. The details of such tools are given below:

- **Email services**  
Email services are required for authentication and verification of many social media applications. In some case the data breach occurs through social media applications releasing the emails of the users. Many services provide temporary email addresses, which can be used for creating and verifying social media application accounts to preserve anonymity of organizational employees. Some of tools that assist in creating fake social media profiles are presented in Table 22.

**Table 22.** Tools used for temporary email services.

ID	Name	Description	Reference
1	10minutemail	Online web application provides a disposable or temporary email address and inbox for up to 10 min.	[93]
2	20minutemail	Online web application provides a disposable or temporary email address and inbox for up to 20 min.	[94]
3	minuteinbox	Online web application provides a disposable or temporary email address and inbox for up to 1 month.	[95]

- **Fake persona generation**  
To operate in cyberspace securely, sometime it is required to have a fake persona to avoid any inherit threat poised by OSINT tools and techniques. These fake persona contains personally identifiable information, which makes individuals unique. Some

tools are being developed to generate fake persona who want to operate in cyber space without disclosing their own identities. The details of such tools are presented in Table 23.

**Table 23.** Tools used for fake persona generation.

ID	Name	Description	Reference
1	fakenamegenerator	Free PHP based web application that can generate random user.	[96]
2	randomuser	Free open source NODE.JS based API that can generate users based upon a given input.	[97]
3	faker.js	Free open source NODE.JS script to generate random users using a browser plugin.	[98]

## 7. Artificial Intelligence in the Field of OSINT

Cybersecurity threats are evolving with time. Hackers are using advanced mechanisms for breaching the security measures that can penetrate the traditional detection and prevention technologies. Therefore, more active and advanced methods for analysing and classifying large amounts of data from various sources, including network traffic, domain name system logs, proxy logs, and devices logs, are needed. The AI methods achieved significant success to address challenging problems in different fields, including but not limited to natural language processing, computer vision, and robotics.

Open-source companies consider AI methods to scan different news from around the world to analyse accelerating trends. AI methods are also used to analyse sentiment for marketing, and political campaigns [99] as well as to fact check fake news and detect deep fakes across social media platforms [100]. Therefore, processing large amounts of data from different public sources through AI methods would represent a significant advantage in terms of efficiency and accuracy. The incorporation of AI methods in OSINT is not yet well-considered.

Therefore, responding to an increasing need to process large amounts of data regarding security threats, the field of malicious cyber activities detection via OSINT tools and artificial intelligence (AI) methods are receiving attention [101]. The ultimate target of using AI methods is to analyse public information and detect cyberthreats automatically.

Different social media sources have substantially increased the magnitude of data that intelligence analysts have to sort out. The current trend in improving different technological aspects allows for more intelligence data to be gathered; it is challenging to extract that intelligence into helpful information. These enormous archives of data are known as Big Data. The data handling capabilities of AI methods mean that Big Data and AI methods are strongly connected. Intelligence analysts realized that the OSINT tools find it challenging to cope with the unstructured nature of the data affiliated with open-source intelligence, presenting it as the key challenge they face today.

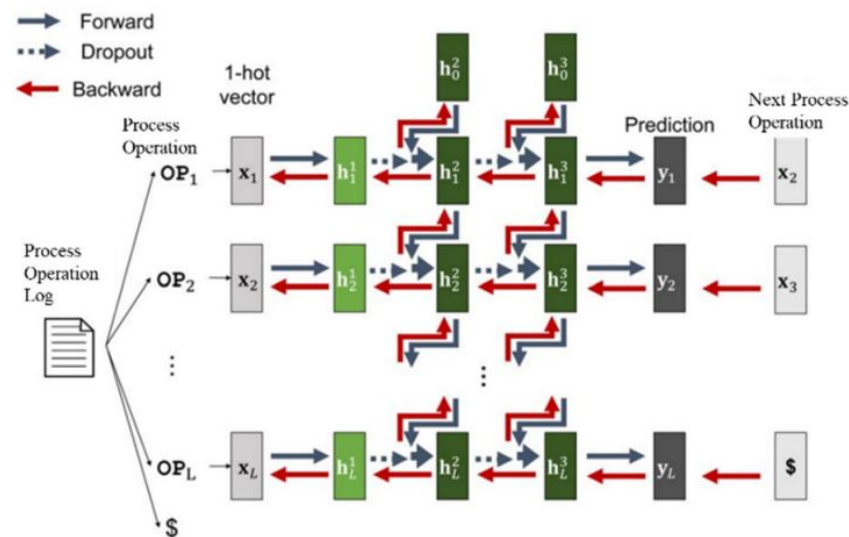
These huge archives effectively make detecting significant features akin to “finding a specific needle in a stack of needles” [102]. However, this challenge can be addressed by AI methods that effectively scan these archives. AI methods allow for abnormal and normal patterns to be detected within data with a greater magnitude of accuracy. AI methods can parse large amounts of information to recognize patterns and anomalies, analyse and classify affiliations, and summarize data insights. AI methods extract informative features from these otherwise insurmountable archives.

The consideration of AI methods in OSINT includes disseminating through otherwise insurmountable archives, autonomous imagery, web crawlers, language processing, event detection, and pattern identification. When combined with AI methods, Big Data techniques, for example, web crawling is effective to digest comprehensive social media or other websites in search of specific details [103].

The significant advantages of using AI methods in OSINT boils down to streamlining how information is sifted, sorted, and disseminated to be most helpful. We highlight the use of AI methods in the process of OSINT [103]. The intelligence analyst obtains extracted information from an AI method. This method may identify keywords or sentences tagged in Twitter or chatroom posts. The analyst discovers the presence of a symbol relating to a specific malicious entity. In fact, web Crawlers gather data from hundreds of sources, including social media, English or foreign language media, news clippings, etc. The gathered data is processed through AI methods developed to detect patterns or create a geographic/temporal profile.

An intelligence profile is then produced according to the user’s preferences. Though none of these examples may seem innovative for the OSINT process, AI methods efficiently streamline the process. It allows the investigator to invest more time in the analysis. Therefore, significant advantage of AI methods in the OSINT process exists. The intersection of all these analyses of AI methods in the OSINT process and the further improvement of the methods could lead to an AI technique capable of rapidly detecting malicious information across different sources. Last but not least, AI techniques suggest what we see as promising paths for future work and improvements in the field of open-source intelligence.

Vinayakumar et al. [104] assessed the usage of different deep-learning methods including recurrent neural network (RNN), long short-term memory (LSTM), and other traditional machine learning classifiers to analyse DNS logs in a local area network (LAN) to discover the domain name as either benign or malicious. The generic structure of RNN is depicted in Figure 8. Deep-learning-based methods worked better than other classical machine learning classifiers. In fact, deep-learning-based methods are characterized by consolidated capabilities to extract the informative features implicitly. The LSTM model presented the highest accuracy to detect malicious activities in all experiments compared to the other deep-learning methods.



**Figure 8.** An RNN architecture for understanding malicious behaviour considering the process operations is presented [105].

The reason is that deep-learning-based methods usually manually avoid complex data preprocessing and feature extraction. Traditional methods also depend on expert knowledge and some potential features, which are challenging to discover. Chen et al. [106] also considered deep-learning-based LSTM model and a CNN model. For this purpose, they used the fully qualified domain names (FQDNs) of DNS packets as the input and exploited an end-to-end detection method. They filtered the detection outcomes of the LSTM model with the grouped filtering technique to reduce the false-positive rate further. They used the requested FQDNs of DNS packets as payloads. They trained LSTM and CNN models

to analyse the payloads to find whether there contains DNS covert channel traffic. In their experiments, the LSTM model performed better than the CNN model.

## 8. Discussion and Conclusions

The OSINT tools provide a cyber attacker with the capability of gathering victim information without active engagement. From a defender perspective, a defender can use open-source threat intelligence information to identify attackers. However, as Sun Tzu [1] said, “know your enemy and know yourself; you need not fear the result of a hundred battles”. The OSINT tools can also be employed as a defensive mechanism to map the information exposure for an attacker to protect himself. We presented different OSINT tools and mapped them with Cyber Kill Chain to identify their usage in an active cyber attack and defence engagement.

We presented OSINT tool usage scenarios in real cybersecurity incidents ranging from simple printer exploitation for propaganda to tracking military movements. The dangers of OSINT for governments, military, private and public enterprises, and ordinary people just using the Internet have never been more significant. We presented some known technical countermeasures against OSINT. However, no matter how many technical walls we place in front of an attacker, there is always somebody behind the walls who does not understand the risk of publicly sharing seemingly unrelated information.

Cybersecurity awareness can play a vital role in defence against OSINT-based attacks. One way to increase awareness against such attacks is by cybersecurity exercises. Most cybersecurity exercises focus on traditional attack and defence scenarios with good technical security operations. However, we need to develop exercise scenarios that incorporate human cybersecurity elements. These scenarios can be as simple as harvesting company employee details from LinkedIn to guessing their user credentials based upon publicly available information.

**Author Contributions:** Conceptualization, M.M.Y.; writing—original draft preparation, M.M.Y.; writing—review and editing, M.U., H.U., M.H. and K.M.; supervision, B.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received funding from the Norwegian University of Science and Technology. This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1A2B5B01070067).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Tzu, S. The art of war. In *Strategic Studies*; Routledge: Abingdon, UK, 2014; pp. 86–110.
2. Hwang, Y.W.; Lee, I.Y.; Kim, H.; Lee, H.; Kim, D. Current Status and Security Trend of OSINT. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 14. [CrossRef]
3. Lee, Y.J.; Park, S.J.; Park, W.H. Military Information Leak Response Technology through OSINT Information Analysis Using SNSes. *Secur. Commun. Netw.* **2022**, *2022*, 10. [CrossRef]
4. Martin, L. Cyber Kill Chain®. 2014. Available online: [https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/Gaining\\_the\\_Advantage\\_Cyber\\_Kill\\_Chain.pdf](https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/Gaining_the_Advantage_Cyber_Kill_Chain.pdf) (accessed on 3 April 2022).
5. Tabatabaei, F.; Wells, D. OSINT in the Context of Cyber-Security. In *Open Source Intelligence Investigation*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 213–231.
6. Revell, Q.; Smith, T.; Stacey, R. Tools for OSINT-Based Investigations. In *Open Source Intelligence Investigation*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 153–165.
7. Tagarev, T.; Sharkov, G.; Stoianov, N. Cybersecurity and resilience of modern societies: A research management architecture. *Inf. Secur.* **2017**, *38*, 93–108.
8. Khanna, P.; Zavarsky, P.; Lindskog, D. Experimental analysis of tools used for doxing and proposed new transforms to help organizations protect against doxing attacks. *Procedia Comput. Sci.* **2016**, *94*, 459–464. [CrossRef]

9. Doxing: What It Is and How to Protect Yourself | NortonLifeLock. Available online: <https://us.norton.com/internetsecurity-privacy-what-is-doxing.html> (accessed on 3 April 2022).
10. Homepage—Maltego. Available online: <https://www.maltego.com/> (accessed on 3 April 2022).
11. What is Defense-in-Depth?—Definition. Available online: <https://www.forcepoint.com/cyber-edu/defense-depth> (accessed on 3 April 2022).
12. Tagarev, T.; Stoianov, N. Scoping the Scenario Space for Multi-sector Cybersecurity Analysis. In *Digital Transformation, Cyber Security and Resilience of Modern Societies*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 203–217.
13. He, Q.; Meng, X.; Qu, R.; Xi, R. Machine learning-based detection for cybersecurity attacks on connected and autonomous vehicles. *Mathematics* **2020**, *8*, 1311. [CrossRef]
14. Jang, S.; Li, S.; Sung, Y. Fasttext-based local feature visualization algorithm for merged image-based malware classification framework for cybersecurity and cyber defence. *Mathematics* **2020**, *8*, 460. [CrossRef]
15. Hayes, D.R.; Cappa, F. Open-source intelligence for risk assessment. *Bus. Horizons* **2018**, *61*, 689–697. [CrossRef]
16. Xu, L.; Li, Y.; Fu, J. Cybersecurity investment allocation for a multi-branch firm: Modeling and optimization. *Mathematics* **2019**, *7*, 587. [CrossRef]
17. OSINT Framework. Available online: <https://osintframework.com/> (accessed on 26 September 2019).
18. Qusef, A.; Alkilani, H. The effect of ISO/IEC 27001 standard over open-source intelligence. *PeerJ Comput. Sci.* **2022**, *8*, e810. [CrossRef] [PubMed]
19. Pieterse, H.; Va not Wout, C.; Kahn, Z.; Serfontein, C. Specialised Media Monitoring Tool to Observe Situational Awareness. In Proceedings of the International Conference on Cyber Warfare and Security, Albany, NY, USA, 17–18 March 2022; Volume 17, pp. 244–252.
20. Kassim, S.R.B.M.; Li, S.; Arief, B. How national CSIRTs leverage public data, OSINT and free tools in operational practices: An empirical study. *Cyber Secur. A Peer-Rev. J.* **2022**, *5*, 251–276.
21. Kanta, A.; Coisel, I.; Scanlon, M. A survey exploring open source Intelligence for smarter password cracking. *Forensic Sci. Int. Digit. Investig.* **2020**, *35*, 301075. [CrossRef]
22. Yeboah-Ofori, A.; Brimicombe, A. Cyber intelligence and OSINT: Developing mitigation techniques against cybercrime threats on social media. *Int. J. Cyber-Secur. Digit. Forensics (IJCSDF)* **2018**, *7*, 87–98. [CrossRef]
23. Glassman, M.; Kang, M.J. Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT). *Comput. Hum. Behav.* **2012**, *28*, 673–682. [CrossRef]
24. Mercado, S.C. Sailing the Sea of OSINT in the Information Age. In *Secret Intell Reader*; Routledge: Abingdon, UK, 2009; Volume 78.
25. Jesson, J.; Matheson, L.; Lacey, F.M. *Doing Your Literature Review: Traditional and Systematic Techniques*; Sage: Thousand Oaks, CA, USA, 2011.
26. Dns Recon and Research, Find and Lookup Dns Records. Available online: [DNSdumpster.com](https://DNSdumpster.com) (accessed on 26 September 2019).
27. Registered Domain Names Search—Find Registered Domain Names. Available online: <https://instantdomainsearch.com/> (accessed on 26 September 2019).
28. Free WHOIS—Domain Name Lookup | IP2WHOIS.com. Available online: <https://www.ip2whois.com/> (accessed on 26 September 2019).
29. GitHub—lanmaster53/recon-ng: Open Source Intelligence Gathering Tool Aimed at Reducing the Time Spent Harvesting Information from Open Sources. Available online: <https://github.com/lanmaster53/recon-ng> (accessed on 26 September 2019).
30. Internet Archive: Wayback Machine. Available online: <https://archive.org/web/> (accessed on 26 September 2019).
31. We Leak Info—Data Breach Search Engine. Available online: <https://weleakinfo.to/> (accessed on 26 September 2019).
32. Cryptome. Available online: <https://cryptome.org/> (accessed on 26 September 2019).
33. GitHub—0x09AL/raven: Raven Is a Linkedin Information Gathering Tool That Can Be Used By Pentesters to Gather Information about an Organization Employees Using Linkedin. Available online: <https://github.com/0x09AL/raven> (accessed on 26 September 2019).
34. Social Searcher—Free Social Media Search Engine. Available online: <https://www.social-searcher.com/> (accessed on 26 September 2019).
35. Social Media Monitoring Wiki—A Wiki of Social Media Monitoring Solutions. Available online: [https://en.ryte.com/wiki/Social\\_Media\\_Monitoring](https://en.ryte.com/wiki/Social_Media_Monitoring) (accessed on 26 September 2019).
36. Yamin, M.M.; Katt, B.; Kianpour, M. Cyber Weapons Storage Mechanisms. In *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*; Springer: Cham, Switzerland, 2019; pp. 354–367.
37. Shodan. Available online: <https://www.shodan.io/> (accessed on 25 September 2019).
38. CENSYS. Available online: <https://censys.io/> (accessed on 25 September 2019).
39. About GreyNoise Intelligence. Available online: <https://www.greynoise.io/> (accessed on 25 September 2019).
40. Vulnerability Database. Available online: <https://vuldb.com/> (accessed on 26 September 2019).
41. NVD—Home. Available online: <https://nvd.nist.gov/> (accessed on 25 September 2019).
42. CNNVD. Available online: <http://www.cnnvd.org.cn/> (accessed on 25 September 2019).
43. Home—FSTEC of Russia. Available online: <https://fstec.ru/en/> (accessed on 25 September 2019).
44. Exploit Database—Exploits for Penetration Testers, Researchers, and Ethical Hackers. Available online: <https://www.exploit-db.com/> (accessed on 25 September 2019).

45. Vulnerability and Exploit Database. Available online: <https://www.rapid7.com/db/> (accessed on 25 September 2019).
46. Oday.Today Agreement-Oday.today Exploit Database: Vulnerability: Oday: New Exploits: Buy and Sell Private Exploit: Shellcode by Oday Today Team. Available online: <http://w.oday.today/> (accessed on 25 September 2019).
47. Find Email Addresses in Seconds • Hunter (Email Hunter). Available online: <https://hunter.io/> (accessed on 25 September 2019).
48. Have I Been Pwned: Check If Your Email Has Been Compromised in a Data Breach. Available online: <https://haveibeenpwned.com/> (accessed on 25 September 2019).
49. Namechk | Username, Domain, and Trademark Search | Username Registration. Available online: <https://namechk.com/> (accessed on 25 September 2019).
50. Find People for Free | Get Their Contact Info | Thatsthem. Available online: <https://thatsthem.com/> (accessed on 25 September 2019).
51. Username Search-Search for Any Username or Email Address to Find the Identity Amongst Billions. Available online: <https://usersearch.org/index.php> (accessed on 25 September 2019).
52. Default Passwords | CIRT.net. Available online: <https://cirt.net/passwords> (accessed on 25 September 2019).
53. Default Passwords List—Select Manufacturer. Available online: <https://default-password.info/> (accessed on 25 September 2019).
54. Default Password Lookup Utility. Available online: [https://www.fortypoundhead.com/tools\\_dpw.asp](https://www.fortypoundhead.com/tools_dpw.asp) (accessed on 25 September 2019).
55. Router Passwords Community Database—The Wireless Router Experts. Available online: <https://www.routerpasswords.com/> (accessed on 25 September 2019).
56. Reconstructor. Available online: [www.reconstructor.org](http://www.reconstructor.org) (accessed on 25 September 2019).
57. Microsoft—Office File Analysis Tool. Available online: <https://book.hacktricks.xyz/generic-methodologies-and-resources/basic-forensic-methodology/specific-software-file-type-tricks/office-file-analysis> (accessed on 25 September 2019).
58. Download VMware Workstation Player | VMware. Available online: <https://www.vmware.com/no/products/workstation-player/workstation-player-evaluation.html> (accessed on 25 September 2019).
59. Oracle VM VirtualBox. Available online: <https://www.virtualbox.org/> (accessed on 25 September 2019).
60. Free Virtual Machines from IE8 to MS Edge—Microsoft Edge Development. Available online: <https://developer.microsoft.com/en-us/microsoft-edge/tools/vms/> (accessed on 25 September 2019).
61. Genymotion Android Emulator | Cloud-Based Android Virtual Devices | Develop—Automate Your Tests-Validate with Confidence. Available online: <https://www.genymotion.com/> (accessed on 25 September 2019).
62. BlueStacks-Not Another Android Emulator—6x Faster Than Any Phone. Available online: <https://www.bluestacks.com/blog/bluestacks-exclusives/performance-new-bluestacks4-en.html> (accessed on 25 September 2019).
63. The Best Android Emulator For PC & Mac | Andy Android Emulator. Available online: <https://www.andyroid.net/> (accessed on 25 September 2019).
64. Free Android Emulator on PC and Mac-Download NoxPlayer. Available online: <https://www.bignox.com/> (accessed on 25 September 2019).
65. Zone-H.org-Unrestricted Information | Defacements Archive. Available online: <http://www.zone-h.org/archive> (accessed on 25 September 2019).
66. Ransomware Abuse Tracker. Available online: [https://ransomwaretracker.abuse.ch/downloads/RW\\_DOMBL.txt](https://ransomwaretracker.abuse.ch/downloads/RW_DOMBL.txt) (accessed on 25 September 2019).
67. Malware Domains. Available online: [mirror1.malwaredomains.com/files/domains.txt](http://mirror1.malwaredomains.com/files/domains.txt) (accessed on 25 September 2019).
68. Malware Patrol. Available online: <https://www.malwarepatrol.net/> (accessed on 25 September 2019).
69. Yamin, M.M.; Katt, B. *A Survey of Automated Information Exchange Mechanisms Among CERTs*; CERC: Cambridge, UK, 2019; pp. 311–322.
70. GitHub—armbues/ioc\_parser: Tool to Extract Indicators of Compromise from Security Reports in PDF Format. Available online: [https://github.com/armbues/ioc\\_parser](https://github.com/armbues/ioc_parser) (accessed on 26 September 2019).
71. GitHub—InQuest/Python-iocextract: Advanced Indicator of Compromise (IOC) Extractor. Available online: <https://github.com/InQuest/python-iocextract> (accessed on 26 September 2019).
72. Threat Research | FireEye Inc. Available online: <https://www.fireeye.com/blog/threat-research.html> (accessed on 26 September 2019).
73. Techniques—Enterprise | MITRE ATT&CK™. Available online: <https://attack.mitre.org/> (accessed on 26 September 2019).
74. IBM X-Force Exchange. Available online: <https://exchange.xforce.ibmcloud.com/> (accessed on 25 September 2019).
75. MISP—Malware Information Sharing Platform and Threat Sharing—The Open Source Threat Intelligence Platform. Available online: <https://www.misp-project.org/> (accessed on 26 September 2019).
76. Threat Exchange Network—Blueliv Community. Available online: <https://community.blueliv.com/#!/discover> (accessed on 26 September 2019).
77. A Hacker Forced 50,000 Printers to Spread PewDiePie Propaganda—And the Problem Is Much Bigger Than You Know. Available online: <https://www.forbes.com/sites/thomasbrewster/2018/12/03/a-hacker-forced-50000-printers-to-spread-pewdiepie-propagandaand-the-problem-is-much-bigger-than-you-know/> (accessed on 26 September 2019).
78. Shodan Search Engine. Available online: <https://www.shodan.io/> (accessed on 26 September 2019).



79. Port 9100 Printing—Hacking Printers. Available online: [http://hacking-printers.net/wiki/index.php/Port\\_9100\\_printing](http://hacking-printers.net/wiki/index.php/Port_9100_printing) (accessed on 26 September 2019).
80. GitHub-RUB-NDS/PRET: Printer Exploitation Toolkit—The Tool That Made Dumpster Diving Obsolete. Available online: <https://github.com/RUB-NDS/PRET> (accessed on 26 September 2019).
81. Hern, A. Fitness tracking app Strava gives away location of secret US army bases. *Guardian* **2018**, *28*, 2018.
82. Corcoran, K. Not Having Snapchat Is the Worst Thing about Living on a Giant Aircraft Carrier, According to HMS Queen Elizabeth’s Youngest Sailor. 2017. Available online: <https://www.sciencedirect.com/topics/computer-science/social-network-site> (accessed on 26 September 2019).
83. Lovell, K.N.; Heering, D. exercise nePtune: MaritiMe cybersecurity training using the navigational siMulators. In Proceedings of the Fifth Interdisciplinary Cyber Research Conference 2019, Tallinn, Estonia, 29 June 2019; p. 34.
84. GitHub-justdelete/justdelete.me: A Directory of Direct Links to Delete Your Account from Web Services. Available online: <https://github.com/justdelete/justdelete.me> (accessed on 26 September 2019).
85. Available online: [ACCOUNTKILLER.COM](https://ACCOUNTKILLER.COM) (accessed on 26 September 2019).
86. Tor Project | Anonymity Online. Available online: <https://www.torproject.org/> (accessed on 26 September 2019).
87. I2P Anonymous Network. Available online: <https://geti2p.net/en/> (accessed on 26 September 2019).
88. Easy Web Browsing From Multiple Locations | LocaBrowser. Available online: <https://www.locabrowser.com/> (accessed on 26 September 2019).
89. Epic Privacy Browser, a Secure Chromium-Based Web Browser That Protects Your Privacy and Browsing History | a Free VPN Privacy Browser. Available online: <https://www.epicbrowser.com/> (accessed on 26 September 2019).
90. Whonix™—Anonymous Operating System. Available online: <https://www.whonix.org/> (accessed on 26 September 2019).
91. Tails—Privacy for Anyone Anywhere. Available online: <https://tails.boum.org/> (accessed on 26 September 2019).
92. Subgraph OS. Available online: <https://subgraph.com/> (accessed on 26 September 2019).
93. 10 Minute Mail. Available online: <https://10minutemail.com/> (accessed on 26 September 2019).
94. 20 Minute Mail—Temporary E-Mail 10 Minute and More-Temp Mail, Fake Email. Available online: <https://www.20minutemail.com/> (accessed on 26 September 2019).
95. MinuteInbox | 10 Minute Mail Address. Available online: <https://www.minuteinbox.com/> (accessed on 26 September 2019).
96. Generate a Random Name-Fake Name Generator. Available online: <https://www.fakenamegenerator.com/> (accessed on 26 September 2019).
97. GitHub—RandomAPI/Randomuser.me-Node: Source Code that Powers Randomuser.me. Available online: <https://github.com/berteltoorp/Randomuser.me-Source> (accessed on 26 September 2019).
98. GitHub—Marak/faker.js: Generate Massive Amounts of Realistic Fake Data in Node.js and the Browser. Available online: <https://github.com/faker-js/faker> (accessed on 26 September 2019).
99. Pastor-Galindo, J.; Nespola, P.; Mármol, F.G.; Pérez, G.M. The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends. *IEEE Access* **2020**, *8*, 10282–10304. [[CrossRef](#)]
100. Ozbay, F.A.; Alatas, B. Fake news detection within online social media using supervised artificial intelligence algorithms. *Phys. A Stat. Mech. Its Appl.* **2020**, *540*, 123174. [[CrossRef](#)]
101. Branco, E.P. Cyberthreat Discovery in Open Source Intelligence Using Deep Learning Techniques. Ph.D. Thesis, Universidade de Lisboa, Lisboa, Portugal, 2017.
102. Van Schaack, B. Leveraging Big Data for LOAC Enforcement: Finding the Needle in a Stack of Needles. In *Big Data and International Humanitarian Law*; Lieber Institute: Baltimore, MD, USA, 2021.
103. Future, R. How Artificial Intelligence Is Shaping the Future of Open-Source Intelligence. 2020. Available online: <https://www.recordedfuture.com/open-source-intelligence-future> (accessed on 26 September 2019).
104. Vinayakumar, R.; Soman, K.; Poornachandran, P. Detecting malicious domain names using deep learning approaches at scale. *J. Intell. Fuzzy Syst.* **2018**, *34*, 1355–1367. [[CrossRef](#)]
105. Lu, S.; Li, Q.; Zhu, X. Stealthy Malware Detection Based on Deep Neural Network. *J. Phys. Conf. Ser.* **2020**, *1437*, 012123. [[CrossRef](#)]
106. Chen, S.; Lang, B.; Liu, H.; Li, D.; Gao, C. DNS covert channel detection method using the LSTM model. *Comput. Secur.* **2021**, *104*, 102095. [[CrossRef](#)]