Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 1290129, 14 pages
https://doi.org/10.1155/2022/1290129

Review Article

Current Status and Security Trend of OSINT

Yong-Woon Hwang[1], **Im-Yeong Lee**[1], **Hwankuk Kim**[2], **Hyejung Lee**[3], **and Donghyun Kim**[4]

[1]*Department of Software Convergence, Soonchunhyang University, Asan 31538, Republic of Korea*
[2]*Department of Information Security Engineering, Sangmyung University, Cheonan 31066, Republic of Korea*
[3]*Department of Innovation and Convergence, Hoseo University, Cheonan 31066, Republic of Korea*
[4]*Department of Computer Science, Georgia State University, Atlanta 30303, GA, USA*

Correspondence should be addressed to Im-Yeong Lee; imylee@sch.ac.kr

Received 17 November 2021; Revised 4 January 2022; Accepted 26 January 2022; Published 18 February 2022

Academic Editor: Yan Huo

Recently, users have used open-source intelligence (OSINT) to gather and obtain information regarding the data of interest. The advantage of using data gathered by OSINT is that security threats arising in cyberspace can be addressed. However, if a user uses data collected by OSINT for malicious purposes, information regarding the target of an attack can be gathered, which may lead to various cybercrimes, such as hacking, malware, and a denial-of-service attack. Therefore, from a cybersecurity point of view, it is important to positively use the data gathered by OSINT in a positive manner. If exploited in a negative manner, it is important to prepare countermeasures that can minimize the damage caused by cybercrimes. In this paper, the current status and security trends of OSINT will be explained. Specifically, we present security threats and cybercrimes that may occur if data gathered by OSINT are exploited by malicious users. Furthermore, to solve this problem, we propose security requirements that can be applied to the OSINT environment. The proposed security requirements are necessary for securely gathering and storing data in the OSINT environment and for securely accessing and using the data collected by OSINT. The goal of the proposed security requirements is to minimize the damage when cybercrimes occur in the OSINT environment.

1. Introduction

Recent developments regarding the Internet of Things (IoT) and big data have caused the amount of data to increase boundlessly and accelerate the advancement of open-source intelligence (OSINT). The paths of information collection, including IoT, are becoming increasingly diverse, and the data are analyzed based on big data. Thus, deriving intelligence is becoming more important. Here, intelligence is translated as mental intelligence, confidentiality, and information. In the military and spy worlds, it is referred to as espionage. Every country worldwide gathers information regarding other countries. The amount of information collected by countries is called Intelligent Surveillance and Reconnaissance (ISR). Information collection methods like ISR are of three types, as Table 1 shows: OSINT, human intelligence (HUMINT), and technical intelligence (TECHINT) [1].

OSINT is the most basic method of collecting information, which is a form of collecting data through open sources (internet, broadcasting, papers, etc.) and processing them [2]. As open information is used, there are advantages, e.g., the information is collected in real-time, and the data are accessed easily and collected at a low cost. However, the importance of the information is lower than that of other information collection methods.

HUMINT indicates that humans extract or steal information. Simply, it refers to spies or secret agents. It has the advantage of obtaining high-quality information like first-class confidential information, however, there always exists a risk of betrayal and double espionage because people are involved [2].

Table 1: Information collection method.

OSINT	Gather intelligence using open information, data, and software. (i) Open information: information available in everyday life, such as internet, broadcasting, papers, and journals.
HUMINT	Intelligence gathered by humans through activities (e.g., spying, undercover operation). (i) White agent: can collect open information but espionage is not permitted. (ii) Black agent: steals confidential information (first-class confidential information, high-quality information) in secret.
TECHINT	Technology and information assets are used to gather enemy intelligence. (i) IMINT: UAV, reconnaissance planes, satellites, etc., are used to gather information. (ii) SIGINT: signals such as radio waves and radar signals are analyzed to gather information. (iii) MASINT: devices other than IMINT and SIGINT are used to gather information.

Recently, TECHINT emerged as an information collection method that uses technology and information assets to gather enemy intelligence. Here, the technology and information assets refer to devices that have the latest technologies for collecting information, such as imagery intelligence (IMINT) and signals intelligence (SIGINT). Its disadvantages are that the costs are high, and the reliability of the acquired information is low when a problem occurs in signals and radio waves.

Each of the three information collection methods has advantages and disadvantages, depending on the environment. In this paper, we examine and explain (See Section 2) OSINT, which is the basis for information-gathering methods. Currently, all users use OSINT technology when searching for data online. On this basis, users obtain information about the data they are looking for. However, from the perspective of cybersecurity, the use of data gathered by OSINT is a double-edged sword.

(i) On the positive side, data gathered by OSINT can be used as a means of resolving cybersecurity threats, which can track down cybercriminals or prevent cyberattacks before occurring.

(ii) On the negative side, data gathered by OSINT becomes the basis for attackers to create cybersecurity threats. In other words, an attacker can set a target to attack based on data, and after gathering related information, they can engage in various cybercrimes, such as hacking, malware, and denial-of-service (DoS) attacks [3].

To solve this problem, it is important to establish basic security requirements in the stages when data are collected and stored in the OSINT environment, as well as when users access the data. Currently, because anyone can access OSINT, the problem is that security-related requirements are not taken into consideration.

Therefore, this paper will explain the current status and security trends of OSINT. In particular, we focus on security awareness by mentioning the importance of OSINT from the perspective of cybersecurity and providing additional security requirements to resolve security threats occurring in the OSINT environment. It is expected to address the future problem of cybercrimes occurring when attackers misuse data collected by OSINT, and the goal is to reduce the cybercrime occurrence rate through security technology and minimize the damage in the event of an occurrence.

The security requirements proposed in this paper are basics, which can be applied to the OSINT environment, where data importance is high, rather than all OSINT environments. Here, the data importance refers to the data that are worth providing confidentiality and integrity for, such as security elements, because the value of the data processed by OSINT is high. Data with low importance are data that anyone can easily access and check, and it does not significantly affect cybersecurity threats. Therefore, it is necessary to apply basic security requirements and security technologies to data with high importance.

This paper consists of the following sections: Section 2 describes the background of OSINT. In detail, OSINT definition, structure, advantages, disadvantages, and examples of using OSINT are described. Section 3 mentions the basic requirements in OSINT and explains the security threats and cybercrimes arising when collected data are maliciously used. Section 4 describes the importance of OSINT from the cybersecurity perspective and presents common security requirements that are needed to solve the security threats mentioned in Section 3. Section 5 mentions the future challenges or necessary research in the OSINT field, and Section 6 concludes the paper.

2. Background of OSINT

This section describes the definition and structure of OSINT. It also describes the advantages and disadvantages of OSINT, and the examples of using OSINT.

2.1. Definition and Structure of OSINT. OSINT is a compound word for open source and intelligence. It refers to the overall process in which anyone can collect and analyze information based on open-source information and create useful information. Before discussing OSINT, we define each term as follows:

(i) *Intelligence*: refers to information and espionage. Specifically, it refers to information collected, processed, and reduced to satisfy explicit or understood needs [4].

(ii) *Open-source data (OSD)*: refers to unprocessed general data. Examples include images, photographs, survey data, audio data, metadata, and datasets, which can be obtained from public information.

(iii) *Open-source information (OSINF)*: refers to general data that have been partially filtered based on requirements or certain criteria. Examples include books, articles, and papers written on certain topics,

and they are characterized by some filtering before being processed. The OSINT data are the result of collecting and processing data according to the purpose of the OSINT tool. Therefore, OSINF is an essential prerequisite for OSINT, and investigators/information producers collect information for OSINT.

(iv) *OSINT*: it refers to data processed through open sources. In detail, it refers to data that has been processed through a search and filtering process to satisfy a specific request or standard purpose. The information is directly used in all intelligence contexts, and a large amount of data are summarized, sorted, and output for the OSINT tool.

(v) *Validated OSINT (OSINT-V)*: it is OSINT with a high degree of certainty/veracity. Data must be checked (verified) using a reputable OSINT source or a source that is not of OSINT. Validation is essential because some malicious users (attackers) tamper with OSINT analysis, produce inaccurate OSINT information, and spread it.

Media companies, colleges, journalists, and scholars have been analyzing OSINT data in the private sector, hundreds of years ago before the advent of the internet. In 2001, Wikipedia was established in the U.S., a nonprofit organization and website that collects, analyzes, and discloses OSINT. It is the world's largest private nonprofit OSINT collection, analysis, and open site on the internet. Currently, the importance of OSINT has rapidly emerged as much information overflows because of the development of computers and the internet in the twenty-first century. Since guidelines for OSINT tools are publicly available, companies and users can set up OSINT tools according to their purpose and collect data [5].

One of the most basic and important steps of OSINT is the search and collection of public information, and the collection of public information is primarily related to the search and acquisition of relevant specialized data. It is necessary to find ways to search and use a vast amount of information that has explosively increased nowadays. To find the desired information, it is necessary to design a structured OSINT search process by developing a series of search processes consistently and systematically [2].

To effectively collect data online, it is important to identify relevant websites corresponding to the invisible Web or deep Web and securing and organizing the list of these websites plays a key role. As they are not searched for in regular methods, their web addresses are obtained through offline sources, such as library searches, references in relevant books, expert reports, and interviews. OSINT data are collected by creating and managing separate lists of collected useful websites or online information [2, 6, 7].

Figure 1 illustrates the basic structure of OSINT. The OSINT process consists of collecting, processing, analyzing, and reporting data after identifying the specified data. Each organization has a modified structure of OSINT according to its purpose and requirements because OSINT requirements vary from one organization to another. However, the OSINT process consists of five steps as shown in Figure 1.

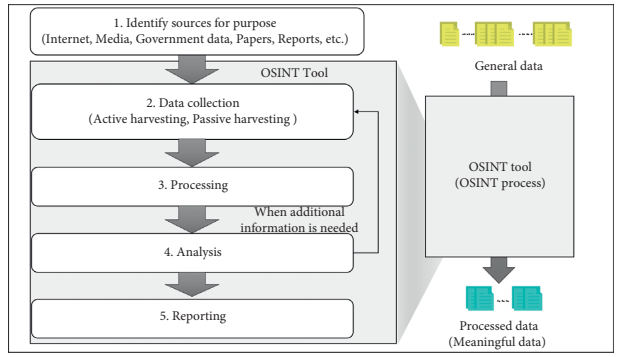

Figure 1: Structure of OSINT.

(i) *Step 1. Identifying the source*: set the information that the investigator (user) wants to obtain among numerous data. One needs to know where and how to get this information.

(ii) *Step 2. Data collection*: it is the stage to bring related data from identifying the source. When bringing in data, the harvesting step is classified into two types, namely active and passive, depending on the data collection method. In the active harvesting step, information is directly collected using a program or script on the target. The active type has a characteristic, which is that logs are left behind because it directly accesses the target. In the passive harvesting step, information is collected using Google, Netcraft, Whois, Recon-NG, Shodan, etc. The passive type has a characteristic, which is that no separate log is left behind because the information is collected using third-party applications.

(iii) *Step 3. Processing*: it is a step to obtain meaningful information by processing and refining the information obtained in Step 2. Since there is a lot of information in Step 2, the task of filtering much information is important in the processing step. Furthermore, it is important to consider the association between information, and the processing step requires a high-difficulty task that needs a lot of experience and perspectives.

(iv) *Step 4. Analysis*: in which the data refined in the processing step are processed according to the investigation purpose. For example, suppose there are evidence data (information) A, B, and C obtained by collecting and processing a variety of information to support an argument. Finally, a conclusion is reached if it is proved by A, B, and C that the argument is correct. If additional information is needed in the analysis step, the data collection and processing steps are continuously repeated to find an association between information to derive meaningful information.

(v) *Step 5. Reporting*: it is a method to summarize the contents up to Step 4 and write it in the form of a report. The reports are distributed and evaluated in various forms, such as evidence and analysis reports, depending on the organization (institution) that uses them. They include all source data that indicate the accuracy of data to give credibility to the data for the argument and evidence. As a result, much of the general data is processed into data that meets the criteria set by the investigator, resulting in meaningful data [2, 8, 9].

2.2. Advantages and Disadvantages of OSINT. As Table 2 shows, OSINT has advantages and disadvantages, depending on how the collected data are used [10, 11]. The advantages of using OSINT are as follows:

(i) *Fast/real-time information collection*: information collected by OSINT is quickly obtained through open sources, and the data are tracked in real-time. To obtain desired data, the user searches data by relying on a variety of OSINF—such as searching the internet data, watching YouTube/TV, and reading books—rather than collecting information from one place. It has the advantage of ensuring prompt data access.

(ii) *Secure acquisition of much data*: the data collected by OSINT secures much data that supplements the gathering of secret intelligence. In HUMINT, the obtained data are few because only a few agents are used, however, in OSINT, there is the advantage that much data can be obtained through open sources. Excellent data (meaningful data) are obtained if a considerable amount of data is processed using OSINT. Furthermore, as OSINF is accessible by anyone, it is legal, and there is the advantage of low risk in terms of a security issue, which means that data are obtained securely.

(iii) *Clarity of sources*: in HUMINT, the credibility of data is questionable because the source of the information that the agent obtains is unclear. In contrast, data collected by OSINT ensures credibility because the clarity of the open sources is guaranteed by a validation process.

(iv) *Convenience and ease of access*: not everyone can access data easily because data access rights are set in such a way that only authorized users can access confidential and high-quality data. In contrast, anyone can easily access information collected by OSINT and use data conveniently according to the user's requirement.

(v) *Low cost*: OSINT has the advantage of obtaining data at a low cost, compared to the cost of training agents in HUMINT and the cost of collecting data using the latest equipment, such as satellites and unmanned aerial vehicles (UAV) in TECHINT.

Disadvantages of using OSINT are as follows:

(i) *The amount of information is too large*: the more information the user has, the harder it is for the user to output reliable data using OSINT. If incorrect information is mixed among the evidence data of several factors supporting an argument, it may reduce the credibility of the data, which may result in false information in the argument data. Currently, because much data is searched for in open sources, it takes time and effort to detect false information and select reliable data.

(ii) *Organizational perception and prejudice of intelligence agencies*: in the organizational culture of intelligence agencies, the value of data collected by OSINT is underestimated, and the importance of data is not considered because anyone can access and use the data.

(iii) *Security issues and technical constraints*: intelligence agencies use internal computer networks because of security issues, which limit the use of OSD using the internet. As a result, analysts at the intelligence agencies exhibit a passive attitude toward the use of OSINT data. Computer security experts are endeavoring to prepare methods of freely using OSD while solving security problems.

(iv) *Cornerstone of cybercrimes when misused*: anyone can access the data collected by OSINT. However, there is the disadvantage that the data collected by OSINT can be the basis of committing cybercrimes because of users with malicious goals. Therefore, research is required on security requirements (measures) and technology that can minimize the damage of cybercrimes, even if users use OSINT's data for malicious purposes.

2.3. Examples of Using OSINT. In recent years, all internet users have been using various OSINT technologies when searching for data online. They collect OSINT data in some form, regardless of whether they are companies, schools, universities, or individuals. The intelligence and investigation agencies of major countries, including the UK and US, have recognized the importance of OSINT early on and are investing systematically and actively building these systems. Related companies are also developing several types of OSINT solutions. In South Korea, a variety of research and education are underway using OSINT [12]. The subsequent paragraphs show the typical examples of OSINT uses [13].

In terms of law enforcement agencies, police use OSINT sources to protect citizens from abuse, sexual violence, identity theft, and other crimes. It is done by monitoring social media channels for keywords and photographs that help prevent crimes before they increase. Law enforcement agencies use OSINT to monitor and track criminal networks in many countries. For example, using OSINT tactics, they collect information about criminals (persons of interest) and

Table 2: Advantages and disadvantages of OSINT.

Advantages	Disadvantages
(i) Fast/real-time information collection (ii) Secure acquisition of large data (iii) Clarity of sources (iv) Convenience and ease of access (v) Low cost	(i) The amount of information is too large (ii) Organizational perception and prejudice of intelligence agencies (iii) Security issues and technical constraints (iv) Cornerstone of cybercrimes when misused

create a complete profile of each criminal. Furthermore, they use OSINT sources to track online counterfeiting and copyright violations and use them as tools for dealing with various cybersecurity threats.

In terms of business corporations, information is power. Businesses use OSINT sources to investigate new markets, monitor competitors' activities, plan marketing activities, and predict anything that may affect the current operation and future growth. In the past, the use of OSINT sources was limited to large corporations with sufficient intelligence budgets, however, nowadays, because of the broad use of the internet, small companies with limited budgets can use OSINT sources and incorporate the obtained information in their business plans.

As opposed to the above advantageous uses, OSINT sources can be used in malicious ways, and terrorist organizations can use OSINT sources to plan attacks. They can gather information about the target (when investigating the target location) before attacking, analyze social media sites to secure more fighters, obtain military information disclosed accidentally by the government (e.g., a method of making explosives), and use various media channels to spread propaganda around the world [13, 14]. Furthermore, data collected by OSINT can become a cornerstone for committing various cybercrimes (See Section 3).

In summary, using OSINT data is important because the results may be a double-edged sword, depending on the aspect of using the data collected by OSINT. For information on overseas, OSINT projects, and open sources, refer to the following papers [5, 15].

3. OSINT General Requirements and Security Threats

3.1. OSINT General Requirements. The requirements of OSINT vary depending on the purpose, organization, and data to be derived from OSINT. Accordingly, the process of OSINT varies. Figure 1 illustrates the basic process of OSINT. The data processed using the OSINT process is stored in the built OSINT database, and it is used and analyzed. In this paper, it is indicated that the data processed by the OSINT tool is stored in the OSINT tool storage. Figure 2 is a schematic diagram of the process of storing and utilizing data collected in OSINT. If the user misuses the stored data, the contents of the security threat are also briefly included. The requirements are composed of the aspect of collecting and storing data and the aspect in which the user accesses and uses the collected data. OSINT data collectors, who collect and store data with OSINT tools, should basically provide data curation, data integrity, and reliability.

(i) *Data curation*: in a museum or art gallery, the term "curator" refers to a person who decides which works to exhibit. In other words, curation is a term that refers to an act of selecting and providing data in the Big Data era, where much data exists. Curation is an essential element for finding the most valuable information by efficiently using a limited "time resource" in the Big Data era. Data curation is provided for the goals of data search, data quality assurance, value addition, reuse, and preservation over time, which includes the creators/recorders and the selection and evaluation of record repositories [16]. In the past, the curation process was carried out using simple information collection. Recently, data are processed more sophisticatedly using data-based deep analysis and machine learning using artificial intelligence (AI) to increase the use-value of data. It is a requirement needed in the OSINT data collectors who collect general data and process it as valuable data.

(ii) *Guarantee of data integrity*: when storing the collected data in the OSINT tool storage, data integrity must be ensured. Here, integrity refers to maintaining the consistency and accuracy of data, and the stored data must not be modified (forgery and alteration) by someone without permission [17, 18]. If anyone accesses and modifies data in an open space, the reliability of data may degrade. In addition, users may accept and spread incorrect information because of the tempered data, laying a cornerstone for cybercrimes like fake news [16]. Therefore, a guarantee of data integrity is an essential requirement in the OSINT process.

(iii) *Guarantee of data reliability*: data reliability was an essential element when users used the collected data [16]. To guarantee data reliability, it was required to validate the data integrity and data sources. Usually, the validation process of sources was performed through the OSINT process, and guaranteeing the sources of the basis for claiming the legitimacy of data and providing integrity for the sources were essential requirements to increase data reliability.

Most OSINT tools did not consider the requirements needed when accessing and using stored data. It means that the collected data can be used by everyone, including general users. Regarding security, one of the biggest problems was that the users of OSINT might accidentally disclose sensitive assets and information on the internet [19]. It was a severe problem because OSINT was used for security purposes. As

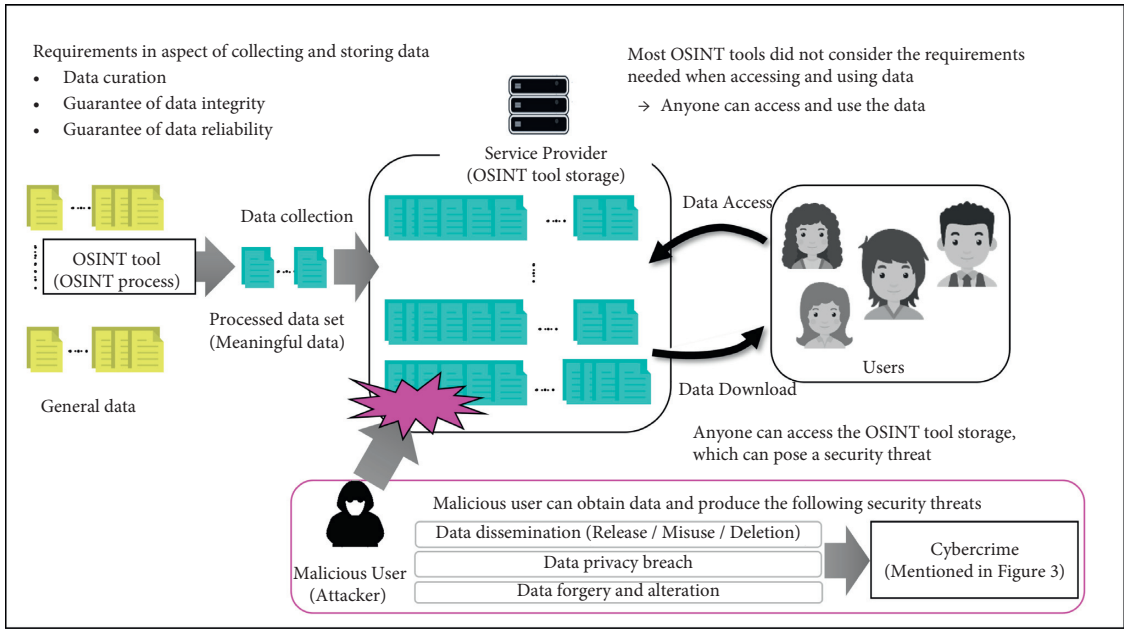

Figure 2: OSINT process and possible security threats.

the source information was disclosed as the word OSINT suggested, anyone could access and use data. Therefore, it was important for users to use data with ethical awareness [16]. Here, ethical use referred to rules that define the allowed actions or proper behaviors. In short, it meant that users should use data for legitimate purposes and not malicious purposes. If users did not have ethical consciousness, it would be the basis for causing various cybercrimes (See Section 3.2).

3.2. Security Threats of OSINT. Figure 2 shows that various security threats, such as data dissemination, data privacy breach, and data forgery and alteration exist in the OSINT environment, which can lead to cybercrime. In Figure 2, the cybercrime mentioned refers to various cybercrimes, such as hacking, data loss, denial of service attack, spreading viruses, and fake news, as well as illegal use in various fields, such as games and financial shopping [19–21]. Figure 3 is a schematic diagram of the contents of cybercrimes that can occur through security threats. Since anyone has access to the data, a malicious user can access the stored data. The acquired data can be disseminated, which can be the basis for hacking financial crimes and virus spread. It can also falsify data, giving users inaccurate information or spreading fake news to create confusion.

Moreover, establishing the reliability of data extracted by OSINT tools is a very difficult problem. It is required to decide who would determine the data reliability and whether the reliability of sources was credible. In the OSINT domain, general trust was often based on perspectives, ideologies, prejudices, beliefs, or product marketing, regardless of value or truth [4]. Public information obtained in secret might be reported early or analyzed immediately and used for a purpose of trusting or disbelieving [2]. Therefore, when general data are processed using OSINT, the evidence for ensuring the data reliability and the guarantee of the data sources must be validated. Thereafter, if the processed data are altered by malicious users, the data reliability would drop, which implied that this needed to be dealt with in advance.

Another issue related to open sources is the concern about privacy breach in the internet age. People use the internet for various purposes, including data collection, analysis, and communication. Ceaselessly expanding social media platforms, such as Facebook and Instagram, are information channels that offer little protection from hackers and malicious attackers [22]. As a result, public trust in privacy was waning, and the problem was that companies (data mining companies) knew information about users, however, the audience knew little about what the data mining companies knew about them [4]. Therefore, security is required for sensitive information, such as personal information. Despite applying general requirements of OSINT mentioned in Section 3.1, various security threats arise, which further contribute to cybercrimes (See Figure 3). The security threats that arise are as follows:

(i) *Data dissemination (release/misuse/deletion)*: if the confidentiality of the collected data is not provided, attackers can obtain data and create various security threats based on them, which might cause the damage of data loss. The collected information of OSINT might vary, depending on the organization and the specialty field, however, if the attacker deleted data, there would be a delay when the users of OSINT data search and check the data they want. Assuming the financial sector, an attacker might use the collected data to obtain users' personal information, financial information, and information about various elements. It would be used as basic data for hacking and might lead to cybercrimes for obtaining sensitive or financial information and monetary gains, such as identity theft and financial

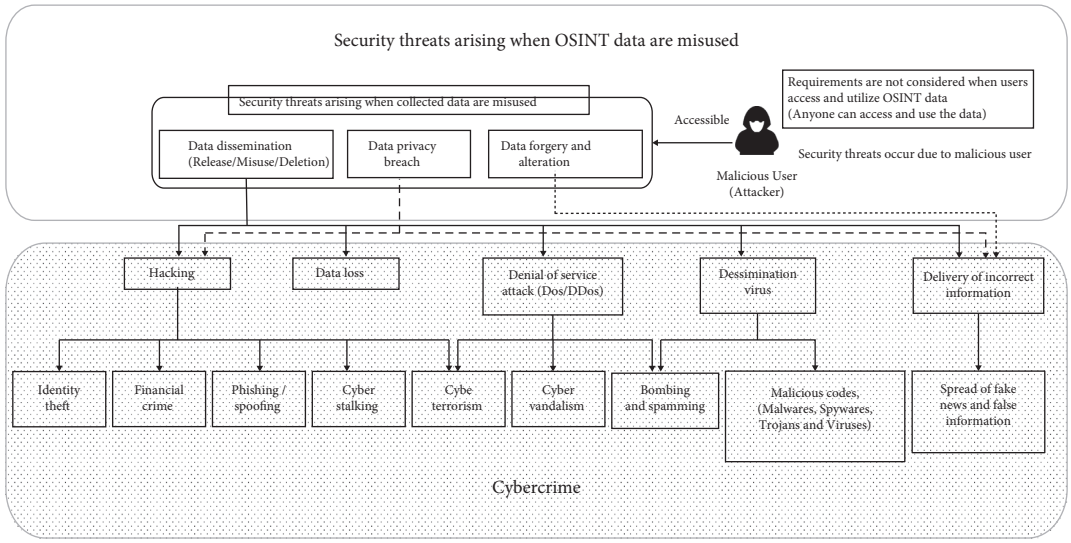

Figure 3: OSINT security threats and cybercrimes.

crimes [21]. Also, because attackers hacked systems based on the obtained OSINT data, it was critical in the national security aspect because not only financial crimes but also, on a larger scale, cyber terrorism activities, such as network overload and DoS attack, could arise [19]. Therefore, if the data collected by OSINT are sensitive and important, it is important to provide confidentiality and the integrity of data, and various requirements, such as user authentication and access control for accessing data, were additionally needed. Furthermore, requirements, such as data backup and recovery, were additionally needed to respond to a loss of data collected by OSINT.

(ii) *Data privacy breach*: if an attacker identified the contents of the collected data, the data for user information (personal information) included in the data could be collected, based on which various security threats could arise [23]. As a typical example, there was a possibility of cybercrimes through the leaking of personal information, which was used as criminal data by the attacker who wanted monetary gains, and cybercrimes, such as identity theft and voice phishing, might occur [24]. The leak of personal information itself caused damages like personal information infringement, however, there was a possibility of additional secondary and tertiary damage. In particular, the attacker set the attack targets based on the personal information of the users and attacked them by spreading viruses or malware. Therefore, additional security measurements were needed, such as anonymity and de-identification, which protect sensitive information, such as the confidential information of users contained in the data collected by OSINT.

(iii) *Data forgery and alteration*: attackers forged or altered the collected data, causing various security threats. If numerically expressed data were altered, the different results of statistical data or figures would be output, and users might misunderstand the information, including the false values in the information. A typical example is fake news, which is rapidly spread to users through social media by manipulating the content and source (evidence data) of data [21]. As a result, the users would be confused about the authenticity of the information, making the collected data less dependable. Therefore, additional security requirements must be in place to ensure the data integrity and guarantee the data source to respond to forgery and alteration of data collected by OSINT.

As such, additional security requirements are needed in OSINT to deal with security threats and cybercrimes (see Section 4). If the importance of the data collected by OSINT was high, security technology, such as user authentication and access control for accessing the collected data, was additionally required. It minimized the security threats that users produced using the data collected by OSINT by making sure in advance that the users would not be able to perform malicious activities, such as data leakage, forgery, and alteration, by accessing data.

4. Role of OSINT from Cybersecurity Perspective

In this section, we explain the importance of OSINT from a cybersecurity perspective. Furthermore, we propose commonly needed security requirements to solve with security threats arising from the misuse of OSINT data.

4.1. Importance of OSINT for Cybersecurity. Regarding cybersecurity, the aspect of using data collected by OSINT could be viewed as a two-edged sword [25]. If the data collected by OSINT were used in the positive aspect, a

considerable amount of data could be obtained compared to secret intelligence data, and on this basis, trends and situations of enemy countries or countries where there were no spies could be examined [2, 3]. Furthermore, if data collected by OSINT were used properly in the security aspect, cybercrimes, such as cyber security threats and cyber terrorism activities, that might occur in cyberspace could be prevented in advance. At present, studies have been continuously conducted to respond to cyberattacks using OSINT [14]. According to a report by the U.S. Office of Homeland Security, the use of data collected by OSINT included general intelligence, advanced warnings, domestic counterterrorism, protection of important infrastructures (including cyberspace), protection against critical terrorism, and emergencies in the domain of important missions [16]. Therefore, the management of data collected by OSINT was crucial because the use of OSINT data was important in terms of cybersecurity. In other words, it would be important for intelligence, security, and public safety agencies in terms of cybersecurity to collect a considerable amount of data from various sources, including criminal records of terrorism incidents and cybersecurity threats, process them into valuable (meaningful) data using OSINT, and securely manage the processed data [20].

If the data collected by OSINT were used in the negative aspect, the attacker could set the target based on profiling. Then, after gathering target information, the attacker commits various cybercrimes, such as SPAM, malware, hacking, DoS attack, phishing, the violation of digital property rights, confidential information infringement, and dissemination of false or confidential information [26, 27]. Cybercriminals could not be easily tracked or caught because they use anonymity and camouflage opportunities through web-based communication to perform malicious activities. Most of these cybercrimes were aimed at user identity theft, stealing sensitive information, and monetary gains. However, the following cases were considered serious cybercrimes in terms of national security: crimes of disrupting legitimate network operations, overloading networks, or denying network services by exploiting loopholes, bugs, improper configuration of software services, or raising false political issues by disseminating incorrect information [28–30].

As OSINT had both positive and negative aspects depending on the data utilization from the cybersecurity perspective, users must make effective use of data collected by OSINT. In other words, the use of OSINT must be limited to legal activities and nonmalicious purposes, and basic security requirements (measures) were additionally needed to minimize the damage, even if the attackers misused the collected information [24].

4.2. Essential Security Requirements When Using OSINT. Various security threats existed in OSINT. As Section 3 mentioned, the basic security threats included data dissemination (release/misuse/deletion), data privacy breach, data forgery, and alteration. To solve these problems, OSINT tools commonly needed additional basic security requirements applied to cloud or IoT environments [18, 31, 32]. Figure 4 is a schematic diagram of the security requirements required when using OSINT. The security requirements in Figure 4 consisted of the security requirements needed in data collection, the storage stage of OSINT, and the security requirements needed when users accessed and used the OSINT data. The details of the requirements are as follows:

4.2.1. Security Requirements Needed When Collecting and Storing OSINT Data

(i) *Data encoding/data encryption*: in general, integrity existed in the data processed by OSINT, however, because they were publicly available information, data confidentiality must be provided, depending on the importance of the data. Confidentiality refers to the prevention of unauthorized access to secure information, and only legitimate users can check the data [18]. Data encoding and encryption technology could be applied for this. The term "data encoding" referred to changing the shape of the information stored in a file to something else according to the purpose and format used when storing or transmitting data. It aimed to increase the data usability in other systems and reduce the space required for storage. The term "data encryption" referred to a method of using an encryption algorithm to conceal the contents of data so that only authorized users could read them, and the purpose was to provide data confidentiality. Encryption was used by mixing symmetric and public-key methods, and lightweight algorithms were often used to reduce the computational amount of encryption [23–34]. However, proxy re-encryption or attribute-based encryption was used if the scope was an environment, where anyone could access and obtain encrypted data in an open space. Proxy re-encryption is an encryption technology that converts the ciphertext without decrypting it so that the proxy can decrypt the ciphertext encrypted with user A's public key with user B's private key [35–37]. Attribute-based encryption is a technology that encrypts and decrypts users' attributes (e.g., affiliation, position, etc.) and an access structure created based on them [38–40]. Both proxy re-encryption and attribute-based encryption are cryptographic techniques, in which users (many unspecified users) accessed and obtained data (ciphertext) when satisfying the policies/conditions of the ciphertext. In other words, the confidentiality of the data is provided, The data can be securely protected because only the users who were authorized to access the data could see the contents of data. Also, if the data were encrypted and stored in the OSINT tool, the ciphertext requested by the user can be easily searched for and transformed without the process of decrypting the ciphertext using searchable encryption [41, 42] and homomorphic encryption [43, 44]. The

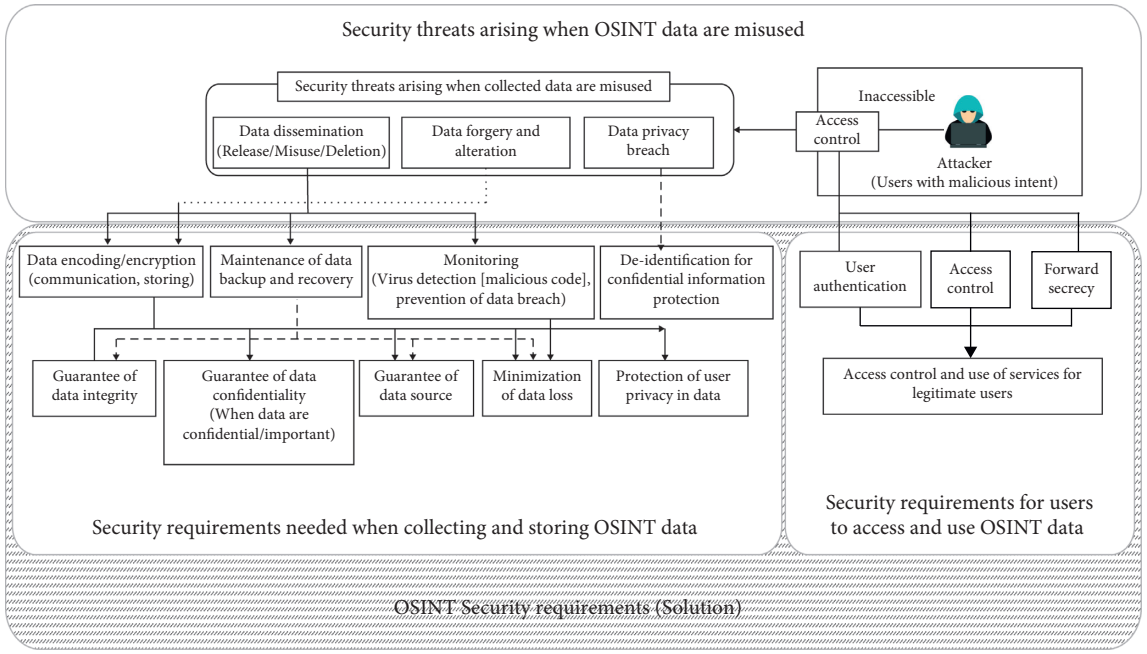

Figure 4: Security requirements needed when using OSINT.

aforementioned encryption techniques were usually used in cloud environments, where the service providers were not trusted. Therefore, it is a basic security requirement to securely store and manage the data collected in the OSINT environment, which is an open space. As such, security elements (data integrity, confidentiality, and guarantee of data source) were provided for the data processed and stored through data encoding/encryption processes. It prevented the loss of data (prevention of data leakage, user privacy breach, etc.) in advance.

(ii) *Maintenance of data backup and recovery*: damage caused by the loss of data because of system crashes, data alteration, and data deletion could not be ignored. Therefore, settings for the backup and recovery of data sets collected by OSINT were important. Data backup and recovery were the processes of backing up data in the case when data loss occurred and configuring the security systems, and in the end, it facilitated the recovery of the lost data [45]. However, considering the backup cost, data recovery cost, and loss cost, it was inefficient to configure the backup and recovery for all data because data collected by the OSINT tool were big data—although they might vary depending on the environment. Therefore, research was required on the operation methods of data backup and recovery to minimize the damage caused by data loss, and this was an additional security requirement.

(iii) *Monitoring*: if OSINT service providers provided additional security elements, such as security audits and security management, the security of the data collected by OSINT would be strengthened. In particular, monitoring should be performed in real-time to detect viruses and malware and prevent users from leaking data [46, 47]. Security audits were activities performed to check whether security activities were conducted appropriately in companies [48]. Security audits were performed according to the policies defined in the OSINT tool and the security activities configured. These were important because security was managed for users and stored data.

(iv) *De-identification of Personal Information*: as data processed by OSINT were open, the users who obtained data could know the contents of personal and sensitive information of users contained in the data [6, 25]. Therefore, if legitimate users obtained data, security technology was required to protect privacy, such as personal and confidential information expressed inside the data. The technology that protected privacy included de-identification, which anonymized personal information. De-identification was a privacy protection technique for reducing the risk of private life infringement by providing statistical anonymity to big data containing sensitive information [49, 50]. Based on this, pseudonymous information and anonymous information can be obtained. Here, pseudonymous information referred to the processing of personal information based on methods, such as partially deleting personal information or replacing some or all of it so that certain users could not recognize it without additional information. Anonymous information referred to the processing of personal credit information so that specific individuals, i.e.,

the subjects of credit information, could not be recognized. The information could not be identified, even if it was combined with additional information, and it was not subject to the Personal Information Protection Act. Therefore, research and security requirements were needed for anonymization of sensitive data, such as de-identification technique, to protect sensitive information contained inside OSINT data. It would minimize cybercrimes, such as privacy breaches and identity theft, caused by user privacy breaches [51, 52].

(v) *Other requirements for guaranteeing data integrity and sources*: the provision of data integrity and guarantee of sources were essential elements because they were related to data reliability for the users of OSINT data. Data integrity and source guarantee are important factors in the stage of collecting and storing data in the OSINT environment. To provide them, signature technology and blockchain technology were typically used. Signature technology referred to a technology that confirmed the signee and showed that the signee signed the data. A variety of signature technologies have been studied and used to increase the reliability of the data sent by the data owner (data sender) [53–55]. Blockchain technology referred to unchanging shared ledgers used for the efficient process of recording transactions and tracking assets in a business network. Among various features, data integrity was provided because transactions for data were recorded, and the recorded transactions were difficult to modify or delete, even for the system administrator [56, 57]. To date, studies have been continuously conducted to ensure data integrity using signature and blockchain technologies in IoT or cloud environments. It can also be used as a way to provide integrity and guarantee sources for data collected and processed by OSINT tools and stored. The use of signature technology or the introduction of blockchain is one of the ways to guarantee integrity and sources. However, the adoption of blockchains should be carefully considered according to the environment since blockchain technology requires high infrastructure.

4.2.2. Security Requirements for Users to Access and Use OSINT Data

(i) *User authentication*: usually, the authentication process was performed to determine who the user was and whether the user had the right when the user tried to access and obtain data on IoT, cloud, or Web. Simply, authentication referred to a process of inputting a user ID and password for accurate verification of the user. In particular, the authentication process was required to access servers, such as a cloud, and it was also required between the users for sending and receiving data [58–61]. OSINT data used in the open space environment was publicly available, and anyone could access and use them. It was considered an advantage and a disadvantage, and as mentioned, authentication technology for users trying to access the OSINT tool was needed, depending on the importance of the data. Authentication technology was typically classified into knowledge-based authentication, ownership-based authentication, biometric authentication, and behavior-based authentication. Usually, it was basic to authenticate a user with a token and certificate issued from the OSINT tool provider after registering the user. Also, one method is to use an identity authentication service using a decentralized identity (DID), which has become an issue recently [62, 63]. DID is a technology in which the user has the authority to control his/her own information, and authentication can be performed with a minimum amount of information compared to the identification method controlled by the existing central system. As a zero-knowledge-proof method, authentication can be performed without disclosing user information [64]. The two methods are the authentication methods that could protect the privacy and sensitive information of users. The method of applying authentication technology might be different, depending on the environment, however, currently, many security threats might exist since any user could access the OSINT tool. They might induce various cybercrimes, as mentioned in Figure 3. Therefore, the malicious use of OSINT data by users should be prevented in advance by at least adopting an authentication process.

(ii) *Access control*: the term access control referred to a function that permitted or denied someone from using something (service). In general, it referred to the user's rights to a service. Regarding information protection, the procedure for access control was conducted to identify the user by the user ID and perform the authentication using the password, token, or signature. Afterward, by granting a security level and a service privilege level according to the Access Control List, authorization/permission would be granted, based on which, the user could use the service [65]. Assuming that the OSINT tool service provider issued a token after user registration, the issued token would contain the rights to access the OSINT tool and the rights and services to use data. In other words, anyone could access the open space and use the data, and the various cybercrimes mentioned in Figure 3 could occur. To deal with this problem, access control technology was required in addition to authentication technology.

(iii) *Forward secrecy*: forward secrecy was important when the OSINT tool was examined from a cloud perspective. Forward secrecy referred to the encrypted communications and sessions recorded

in the past that could not be retrieved [17]. In other words, it was to assume that the user had a token containing the rights for accessing the OSINT tool. If the user's registration period expires or the registration for the OSINT tool is withdrawn, the user's right to access the OSINT tool must expire. Withdrawn user should not be able to access the OSINT tool to obtain the content of the collected data. Moreover, backward secrecy was not considered because when registering in the OSINT tool, the user received the service rights for reading the information in the OSINT tool.

5. Challenges of OSINT

This section describes the research challenges required for future OSINT development. Research should be conducted on the OSINT process for the efficient extraction of the data that the user wants from countless big data in OSINT. Furthermore, additional research is required according to the situation to improve the security of the collected data. As Section 3 mentioned, research for providing general requirements should be provided, and additional research is required to proactively prevent security threats and cybercrimes that might occur if the users use the data of OSINT maliciously (See Section 4). The elements needed for future OSINT development might vary depending on the situation, however, in common, the following challenges should be continuously studied [13, 15]. It is similar to the requirements mentioned earlier but will have a major impact on the evolution of OSINT.

(i) *Efficient and reliable data filtering*: to extract the data that the user wants from OSINT, much data should be collected and effectively filtered [13]. It consumes a huge amount of time and human resources, depending on the amount of data. Organizations or users will utilize automation tools (organizations have their own AI filtering tools) and skill sets to filter data according to purpose. However, the accuracy and reliability of the data extracted when there are software defects in the set of automation tools and techniques are questionable. Therefore, it is important to continuously check the automation tool that is the standard for data filtering, and research on the verification of the extracted data is necessary. It remains a challenge for collectors who collect and filter data from OSINT.

(ii) *Provides data transparency*: the reliability of the collected data was a critical issue in the aspect of the users using OSINT's data. In particular, the verification of sources for claiming the legitimacy of data during the OSINT process increased data reliability significantly. However, in the case of obtaining OSINT data by illegal means, the user might intentionally discard or hide important sources, however, no countermeasure existed in OSINT. It is important to keep a record of the sources on which the data extracted from OSINT in the future is based on credibility. Through this, users need to be provided with data transparency, and research on this still remains a challenge. It also requires the integration and collaboration of many OSINT tools to provide data reliability and transparency [15].

(iii) *Lack of validation of privacy management procedures*: many companies, such as Facebook and Google, collect much data from online users for commercial intelligence. Data that was collected online included not only general data produced by users but also sensitive information, such as names, birthdays, addresses, and passport numbers. Many companies have revealed that they collect and manage data anonymizing to justify data collection, however, it is unknown whether this is being done properly [13]. It is similar to the privacy problem that may arise when collecting data depending on the purpose of the OSINT tool. If anonymous information was mixed in a large amount of data, it would be questionable whether the processed data were reliable and whether the provided anonymization method was properly executed for the data. Therefore, research into the validation of privacy management procedures in OSINT data still remains a challenge. From a legal perspective, OSINT should use data while respecting the data protection policy according to the law [13].

6. Conclusions

In this paper, we explained the current status and security trends of OSINT and proposed security requirements needed in OSINT from a cybersecurity perspective. Specifically, to deal with security threats arising when data collected in the OSINT environment are misused, we proposed basic security requirements according to the steps of collecting and storing data in OSINT and the steps involved when the users accessed the data collected by OSINT. They were similar to the security requirements required in cloud environments or IoT environments. It was crucial to provide data confidentiality and integrity for important data through data encoding/encryption. Anonymization techniques, such as de-identification, were required to protect user privacy in data. Data backup and recovery processes were also needed to minimize data loss, and users accessing the data collected in OSINT should be managed through authentication and access control. Furthermore, based on forward secrecy, users whose registration was canceled should not have access to the data collected by OSINT. In the OSINT environment where these security requirements were satisfied, attackers would not be able to commit cybercrimes easily. The best method for dealing with cybercrimes was to use the data collected in OSINT only in a positive aspect, and it was important for users to have ethical awareness.

In future research, we will propose a secure OSINT model based on the security requirements presented in this

paper. Moreover, we need to conduct a study to improve the security performance of the secure OSINT model. It can be accomplished by testing whether the security threats mentioned in this paper can arise in the secure OSINT model.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea Government (MSIT) (No. 2021-0-00358, AI·Big data based Cyber Security Orchestration and Automated Response Technology Development)

References

[1] P. Casanovas, "Cyber warfare and organised crime. a regulatory model and meta-model for open source intelligence (OSINT)," *Ethics and Policies for Cyber Operations*, pp. 139–167, 2017.

[2] W. H. Lee, M. W. Yun, and J. S. Park, "Intelligence in the internet Era: understanding OSINT and case analysis," *Korean Security Journal*, vol. 34, pp. 259–278, 2013.

[3] K. Shin, fnm au, J. Yoo et al., "A study on building a cyber at tack database using open source intelligence (OSINT)," *Jouranl of Information and Security*, vol. 19, no. 2, pp. 113–121, 2019.

[4] B. H. Miller, "Open source intelligence (OSINT): an oxymoron?" *International Journal of Intelligence & Counter Intelligence*, vol. 31, no. 4, pp. 702–719, 2018.

[5] M. E. Hayden, *Guide to Open Source Intelligence (OSINT)*, Tow Center for Digital Journalism, Columbia University, New York, NY, USA, pp. 1–61, 2019.

[6] S. Chauhan and N. K. Panda, "Open source intelligence and advanced social media search," *Hacking Web Intelligence Open Source Intelligence and Web Reconnaissance Concepts and Techniques*, Elsevier, Amsterdam, Netherlands, pp. 15–32, 2015.

[7] S. Chauhan and N. K. Panda, "Understanding browsers and beyond," *Hacking Web Intelligence Open Source Intelligence and Web Reconnaissance Concepts and Techniques*, Elsevier, Amsterdam, Netherlands, pp. 33–52, 2015.

[8] M. Danda, "Open source intelligence and cybersecurity," Unpublished Master's Thesis, Webster University, Webster Groves, MO, USA, 2019.

[9] A. Kanta, I. Coisel, and M. Scanlon, "A survey exploring open source Intelligence for smarter password cracking," *Forensic Science International: Digital Investigation*, vol. 35, Article ID 301075, 2020.

[10] W. Chun, "Open source intelligence in the information age," *Journal of National Intelligence Studies*, vol. 1, no. 1, p. 151, 2008.

[11] T. Dokman and T. Ivanjko, "Open source intelligence (OSINT) issues and trends," *The Future of Information Sciences*, pp. 191–196, 2020.

[12] L. Benes, "OSINT, new technologies, education: expanding opportunities and threats, a new paradigm," *Journal of Strategic Security*, vol. 6, no. 3, pp. 22–37, 2013.

[13] N. A. Hassan and R. Hijazi, "The evolution of open source intelligence," *Open Source Intelligence Methods and Tools*, pp. 1–20, Apress, Berkeley, CA, USA, 2018.

[14] D. Wells, "Taking stock of subjective narratives surrounding modern OSINT," *Open Source Intelligence Investigation*, pp. 57–65, 2016.

[15] J. Pastor-Galindo, P. Nespoli, F. Martinez Perez, and G. M. Perez, "The not yet exploited goldmine of OSINT: opportunities, open challenges and future trends," *IEEE Access*, vol. 8, pp. 10282–10304, 2020.

[16] F. Tabatabaei and D. Wells, "OSINT in the context of cybersecurity," in *Open Source Intelligence Investigation: From Strategy to Implementation, B. Akhgar*, P. S. Bayerl and F. Sampson, Eds., Springer, Cham, Switzerland, pp. 213–231, 2016.

[17] F. Alkhudhayr, S. Alfarraj, B. Aljameeli, and S. Elkhdiri, "Information security: a review of information security issues and techniques." a review of information security issues and techniques," in *Proceedings of the 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, pp. 1–6, Riyadh, Saudi Arabia, May 2019.

[18] R. Barona and E. M. Anita, "A survey on data breach challenges in cloud computing security: issues and threats," in *Proceedings of the 2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, pp. 1–8, Kollam, India, April 2017.

[19] A. Yeboah-Ofori and A. Brimicombe, "Cyber intelligence and OSINT: developing mitigation techniques against cybercrime threats on social media," *International Journal of Cyber-Security and Digital Forensics*, vol. 7, no. 1, pp. 87–98, 2018.

[20] H. Chen, R. H. Chiang, and V. C. Storey, "Business intelligence and analytics: from big data to big impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165–1188, 2012.

[21] W. Kim, O.-R. Jeong, C. Kim, and J. So, "The dark side of the Internet: attacks, costs and responses," *Information Systems*, vol. 36, no. 3, pp. 675–705, 2011.

[22] G. Li, Z. Cai, G. Yin, Z. He, and M. Siddula, "Differentially private recommendation system based on community detection in social network applications," *Security and Communication Networks*, vol. 2018, Article ID 3530123, 2018.

[23] M. Siddula, Y. Li, X. Cheng, Z. Tian, and Z. Cai, "Privacy-enhancing preferential lbs query for mobile social network users," *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1–13, 2020.

[24] B. J. Koops, J. H. Hoepman, and R. Leenes, "Open-source intelligence and privacy by design," *Computer Law & Security Review*, vol. 29, no. 1, pp. 676–688, 2013.

[25] P. Chen, "Data mining applications in e-government information security," *Procedia Engineering*, vol. 29, pp. 235–240, 2012.

[26] X. Zheng, Z. Cai, and Y. Li, "Data linkage in smart internet of things systems: a consideration from a privacy perspective," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 55–61, 2018.

[27] W. Sun, Z. Cai, Y. Li, and F. Liu, "Security and privacy in the medical internet of things: a review," *Security and Communication Networks*, vol. 2018, pp. 1–10, 2018.

[28] R. Buch, D. Ganda, P. Kalola, and N. Borad, *World of Cyber Security and Cybercrime*, STM Journals 2017, vol. 4, no. 2, pp. 18–23, 2017.

[29] B. Akhgar, "Osint as an integral part of the national security apparatus," *Open Source Intelligence Investigation*, Springer, Cham, Switzerland, pp. 3–9, 2016.

[30] I. Vacas, I. Medeiros, and N. Neves, "Detecting network threats using OSINT knowledge-based IDS," in *Proceedings of the 2018 14th European Dependable Computing Conference (EDCC)*, pp. 128–135, Lasi, Romania, 2018.

[31] A. Singh and K. Chatterjee, "Cloud security issues and challenges: a survey," *Journal of Network and Computer Applications*, vol. 79, pp. 88–115, 2017.

[32] R. Kumar and R. Goyal, "On cloud security requirements, threats, vulnerabilities and countermeasures: a survey," *Computer Science Review*, vol. 33, pp. 1–48, 2019.

[33] I. Bhardwaj, A. Kumar, and M. Bansal, "A review on lightweight cryptography algorithms for data security and authentication in IoTs," in *Proceedings of the 2017 4th International Conference on Signal Processing, Computing and Control (ISPCC)*, pp. 504–509, Solan, India, September 2017.

[34] S. Sallam and B. D. Beheshti, "A survey on lightweight cryptographic algorithms," in *Proceedings of the TENCON 2018-2018 IEEE Region 10 Conference*, pp. 1784–1789, Jeju, Korea, October 2018.

[35] C. K. Chu and W. G. Tzeng, "Identity-based proxy re-encryption without random oracles," *International Conference on Information Security*, vol. 4779, pp. 189–202, 2017.

[36] Y. Yang, H. Zhu, H. Lu, J. Weng, Y. Zhang, and K.-K. R. Choo, "Cloud based data sharing with fine-grained proxy re-encryption," *Pervasive and Mobile Computing*, vol. 28, pp. 122–134, 2016.

[37] Z. Qin, H. Xiong, S. Wu, and J. Batamuliza, "A survey of proxy re-encryption for secure data sharing in cloud computing," *IEEE Transactions on Services Computing*, p. 1, 2016.

[38] S. Namasudra, "An improved attribute-based encryption technique towards the data security in cloud computing," *Concurrency and Computation: Practice and Experience*, vol. 31, no. 3, Article ID e4364, 2019.

[39] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in *Proceedings of the 13th ACM Conference on Computer and Communications Security*, pp. 89–98, New York, NY, USA, 2006.

[40] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in *Proceedings of the 2007 IEEE Symposium on Security and Privacy (SP'07)*, pp. 321–334, Berkeley, CA, USA, May 2007.

[41] L. Wu, B. Chen, K.-K. R. Choo, and D. He, "Efficient and secure searchable encryption protocol for cloud-based Internet of Things," *Journal of Parallel and Distributed Computing*, vol. 111, pp. 152–161, 2018.

[42] Y. Wang, J. Wang, and X. Chen, "Secure searchable encryption: a survey," *Journal of Communications and Information Networks*, vol. 1, no. 4, pp. 52–65, 2016.

[43] M. Van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan, "Fully homomorphic encryption over the integers," in *Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 24–43, Nice, France, 2010.

[44] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A survey on homomorphic encryption schemes: theory and implementation," *ACM Computing Surveys*, vol. 51, no. 4, pp. 1–35, 2018.

[45] J. Zhang and H. Li, "Research and implementation of a data backup and recovery system for important business areas," in *Proceedings of the 2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, vol. 2, pp. 432–437, Hangzhou, China, 2017.

[46] H. J. Syed, A. Gani, R. W. Ahmad, M. K. Khan, and A. I. A. Ahmed, "Cloud monitoring: a review, taxonomy, and open research issues," *Journal of Network and Computer Applications*, vol. 98, pp. 11–26, 2017.

[47] R. Badhwar, "Introduction to cloud monitoring security controls," in *The CISO's Next Frontier*, pp. 289–296, Springer, Cham, Switzerland, 2021.

[48] S. Majumdar, T. Madi, Y. Jarraya, and M. Pourzandi, "Cloud security auditing: major approaches and existing challenges," in *Proceedings of the International Symposium on Foundations and Practice of Security*, pp. 61–77, Montreal, QC, Canada, November 2018.

[49] M. Kayaalp, "Modes of de-identification," *AMIA Annual Symposium Proceedings*, vol. 1044, 2017.

[50] H. J. Lee, S. H. Cho, J. W. Seong, S. Lee, and W. Lee, "De-identification and privacy issues on bigdata transformation," in *Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 514–519, Busan, Korea, February 2020.

[51] H. Li, F. Guo, W. Zhang, J. Wang, and J. Xing, "(a,k)-Anonymous scheme for privacy-preserving data collection in iot-based healthcare services systems," *Journal of Medical Systems*, vol. 42, no. 3, pp. 56–59, 2018.

[52] C. Su, "Big data security and privacy protection," in *Proceedings of the 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS)*, pp. 87–89, Jishou, China, September 2019.

[53] X. Ma, J. Shao, C. Zuo, and R. Meng, "Efficient certificate-based signature and its aggregation," in *Proceedings of the International Conference on Information Security Practice and Experience*, pp. 391–408, Melbourne, VIC, Australia, December 2017.

[54] A. Buldas, D. Firsov, R. Laanoja, H. Lakk, and A. Truu, "A new approach to constructing digital signature schemes," *Advances in Information and Computer Security*, pp. 363–373, 2019.

[55] F. Rezaeibagha, Y. Mu, X. Huang, W. Yang, and K. Huang, "Fully secure lightweight certificateless signature scheme for IIoT," *IEEE Access*, vol. 7, pp. 144433–144443, 2019.

[56] W. Gao, W. G. Hatcher, and W. Yu, "A survey of blockchain: techniques, applications, and challenges," in *Proceedings of the 2018 27th International Conference on Computer Communication and Networks (ICCCN)*, pp. 1–11, Lyon, France, April 2018.

[57] W. Wang, H. Xu, M. Alazab, T. R. Gadekallu, Z. Han, and C. Su, "Blockchain-based reliable and efficient certificateless signature for IIoT devices," *IEEE Transactions on Industrial Informatics*, p. 1, 2021.

[58] A. Tewari and B. B. Gupta, "A lightweight mutual authentication protocol based on elliptic curve cryptography for IoT devices," *International Journal of Advanced Intelligence Paradigms*, vol. 9, no. 2-3, pp. 111–121, 2017.

[59] P. K. Panda and S. Chattopadhyay, "A secure mutual authentication protocol for IoT environment," *Journal of Reliable Intelligent Environments*, vol. 6, no. 2, pp. 79–94, 2020.

[60] P. Mohit, R. Amin, A. Karati, G. P. Biswas, and M. K. Khan, "A standard mutual authentication protocol for cloud computing based health care system," *Journal of Medical Systems*, vol. 41, no. 4, p. 50, 2017.

[61] M. Wazid, A. K. Das, N. Kumar, and A. V. Vasilakos, "Design of secure key management and user authentication scheme for fog computing services," *Future Generation Computer Systems*, vol. 91, pp. 475–492, 2019.

[62] A. Abraham, F. Hörandner, O. Omolola, and S. Ramacher, "Privacy-preserving EID derivation for self-sovereign identity systems," *International Conference on Information and Communications Security*, vol. 11999, pp. 307–323, 2019.

[63] D. van Bokkem, R. Hageman, G. Koning, L. Nguyen, and N. Zarin, "Self-sovereign identity solutions: the necessity of blockchain technology," 2019, https://arxiv.org/abs/1904.12816.

[64] N. V. Kulabukhova, "Zero-knowledge proof in self-sovereign identity," *CEUR Workshop Proceedings*, vol. 2507, pp. 381–385, 2019.

[65] F. Cai, N. Zhu, J. He, P. Mu, W. Li, and Y. Yu, "Survey of access control models and technologies for cloud computing," *Cluster Computing*, vol. 22, no. 3, pp. 6111–6122, 2019.