sustainability

MDPI

Article

C4I System Security Architecture: A Perspective on Big Data Lifecycle in a Military Environment

Seungjin Baek and Young-Gab Kim *

Department of Computer and Information Security and Convergence Engineering for Intelligent Drone, Sejong University, Seoul 05006, Korea; specialbaek@sju.ac.kr
* Correspondence: alwaysgabi@sejong.ac.kr

Abstract: Although the defense field is also one of the key areas that use big data for security reasons, there is a lack of study that designs system frameworks and presents security requirements to implement big data in defense. However, we overcome the security matters by examining the battlefield environment and the system through the flow of data in the battlefield. As such, this research was conducted to apply big data in the defense domain, which is a unique field. In particular, a three-layered system framework was designed to apply big data in the C4I system, which collects, manages, and analyzes data generated from the battlefield, and the security measures required for each layer were developed. First, to enhance the general understanding of big data and the military environment, an overview of the C4I system, the characteristics of the 6V's, and the five-phase big data lifecycle were described. While presenting a framework that divides the C4I system into three layers, the roles and components of each layer are described in detail, considering the big data lifecycle and system framework. A security architecture is finally proposed by specifying security requirements for each field in the three-layered C4I system. The proposed system framework and security architecture more accurately explain the unique nature of the military domain than those studied in healthcare, smart grids, and smart cities; development directions requiring further research are described.

Keywords: big data security; security architecture; C4I system framework; Internet of Battlefield Things (IoBT)

check for updates

Citation: Baek, S.; Kim, Y.-G. C4I System Security Architecture: A Perspective on Big Data Lifecycle in a Military Environment. *Sustainability* **2021**, *13*, 13827. https://doi.org/10.3390/su132413827

Academic Editor: Marc A. Rosen

Received: 18 November 2021
Accepted: 12 December 2021
Published: 14 December 2021

1. Introduction

In the hyper-connected era of the fourth industrial revolution, big data technology is being used to create better lives for people and generate increased profits in the fields of commerce, healthcare, public services, science, military, etc. Big data systems provide users valuable information by collecting, storing, and analyzing the large amount of data generated via smartphones [1,2] and the Internet of Things (IoT) [3,4] in various fields. As the use of big data systems increases, new technologies are emerging based on big data, and existing technologies are becoming more advanced, increasing the complexity of system architectures. Such advanced technologies such as distributed file systems [5], non-relational SQL [6], cloud systems, machine learning [7], security, etc., are considered part of the big data system. For example, Ge et al. [8] suggested that IoT operated in the big data lifecycle and provided an efficient operation plan in the big data system. In addition, as the importance of security for safely operating big data systems is further emphasized, research on measures to reduce risks and analyze security threats to big data systems is continuing. One of the most frequently mentioned keywords in big data security issues is personal data protection. For example, Abouelmehdi et al. [9] suggest ways to strengthen privacy protection in big data systems in the medical field by reviewing privacy laws established in various countries around the world.

Big data systems have different characteristics depending on the operating environment and purpose of use. Since the source, format, volume, and analysis method of the data are different, it is necessary to study the structure and security requirements of the system depending on the field of use. The defense sector is also one of the critical areas that use big data, and intensive study is required to understand the unique operating environment and apply strengthened security requirements. In particular, the command control communication computer and intelligence (C4I) system to which big data are applied requires more careful study. This is because the C4I system is a core command and control system and requires a stable system structure and the highest level of security. However, unfortunately, there is a lack of study that designs system frameworks and presents security requirements to implement big data in defense. Since it is very limited to access and disclose the systems in the military for security reasons, many researchers have not actively conducted academic research related to the C4I big data system, and it is difficult to find well-organized research results related to the C4I big data system framework and security architecture. However, in this paper, we overcome limitations without using any classified information just because we understood the battlefield and the system according to the flow of data handled in the battlefield. Therefore, by examining the C4I system in detail from the data point of view, it was possible to design a stable system framework and derive the security requirements to each framework layer. Finally, the security architecture for the C4I big data system was designed based on the system framework and security requirement to which big data were applied.

The study's main contributions are as follows:

1. Design a three-layered C4I system framework that can implement a big data system: The C4I system has all 6V's characteristics of big data, and the layered framework confirmed the correlation between the big data lifecycle and the big data framework.
2. Derivation of detailed four-layered security architecture for the military C4I big data system: We derived security requirements for the entire C4I big data system.
3. Introduction of potential directions focusing on C4I big data systems' security to guide security managers in the military: Security managers can estimate the security level of the entire system through the security architecture.

This research's methodology is that a logical development was made to propose a new C4I big data system framework and security architecture while presenting a reasonable and analytical basis. First, the correlation between the big data and the C4I system is analyzed, and then the C4I big data system framework is presented. In addition, by specifically presenting the components of each part of the C4I big data system framework, it is possible to present security threats and requirements to be considered at each part. As a result, it is possible to propose a newly proposed security architecture by applying security requirements based on the system framework.

The remainder of this paper is organized as follows: Section 2 provides background knowledge about big data in the military environment. To understand the military environment, this section presents an overview of C4I systems from a data perspective and how the defense information system is applied to the 6V's characteristics of big data by examining the details of 6V's. In addition, a five-step big data lifecycle is presented. Section 3 proposed a layered framework for a big data system and a C4I system. In addition, the roles and components of each layer are specifically presented. Section 4 provides an analysis of the security requirements for each layer of the C4I system and the four-layered big data system architecture. Section 5 proposes a comparative analysis of the system framework and security architecture studied in the military field. Finally, Section 6 concludes the study.

2. Understanding Big Data in Military Environment

It is crucially necessary to understand the military environment from a data perspective to use big data technology in defense fields. Section 2.1 provides an overview of C4I systems from a data perspective and its security concerns. In addition, Section 2.2 examines that the unique characteristics of big data (6V's) can also be applied to big military data. Section 2.3 provides background knowledge to understand the C4I big data system by explaining the five-phase big data lifecycle.

2.1. Overview of C4I Systems from a Data Perspective and Its Security Concerns

The traditional military operational environment, from a system perspective, focused on the capabilities of communication to establish command and control (C2) among deployed combat elements and headquarters. Communication has become an important factor in establishing the concept of command, control, and communication (C3) as another alphabet C that is essential in addition to C2 in the battlefield. The system to support C3 is used as a means of sharing information and visualizing battlefields, but the role of information analysis and determination depends on the knowledge and experience of the commander and staff. However, the modern concept of C4I demands high efficiency in decision making and control because of the transformative developments of information technology that is deeply embedded in advances in the military environment [10]. Therefore, the analysis of battlefield data, which relies on human knowledge and experience, can be systematically analyzed by appropriate algorithms within the C4I system.

As the importance of battlefield data and the value of analysis output increases, it is necessary to understand the C4I system from a data point of view (i.e., the process of data creation, management, analysis, and utilization). First, data are generated from various devices such as sensors, radars, and UAVs in all operational environments, such as combat areas, combat support facilities, and civilian areas. The raw data are transmitted to the military information system that can manage it efficiently, and an appropriate analysis algorithm is applied to provide valuable information. Based on the analyzed results, the command center establishes the best operational plan and supports effective operations by sharing it with the battlefield. Wang et al. [11] reviewed several techniques to build a C4I system architecture. However, those techniques focused on mission-oriented approaches such as situational awareness and collaborative decision-making abilities. In this paper, we propose a framework based on the data flow in C4I systems. A C4I system framework from a data point of view is described in detail in Section 3.

The importance of the security of defense information systems has been continuously emphasized. Ahmad et al. [12] analyzed the C3I systems' cyber threats and security issues by proposing vulnerabilities, attack vectors, and countermeasures. Although they suggested countermeasures against vulnerabilities and attack vectors in a C3I system, it did not present the security requirements for each component of the system architecture. In addition, existing studies [13–15] suggested security requirements applicable to the Internet of Battlefield Things (IoBT), cloud systems, and wireless sensor networks, respectively, but studies about the security architecture applicable to the holistic C4I system framework have been lacking. In this study, we propose a security architecture for C4I big data systems based on the data perspective. Security architecture for C4I systems is explored further in Section 4.

2.2. 6V's in the Big Military Data

De Mauro et al. [16] proposed the definition of big data as "the information asset characterized by high volume, velocity, and variety to require specific technology and analytical methods for its transformation into value". Basically, this definition refers to big data as information assets, and the four fundamental features of big data, information, technology, method, and impact, are considered to comprise the definition. In general, however, the nature of big data is usually described in several words starting with V. Since Doug Laney in 2001 [17] defined big data using 3V's, namely volume, velocity, and variety, many V's have emerged up to 11V's, and complexity was added, which can be abbreviated as $V_3^{11} + C$ [18].

Although $V_3^{11} + C$ is valuable for characterizing big data, many other studies [19–21] referred to the characteristics of big data using the 6V's: volume, velocity, variety, variability, veracity, and value. Big data systems that are implemented in C4I systems also have the characteristics of 6V's, which are described in detail as follows and briefly addressed in Table 1.

Table 1. Big data characteristics of 6V's can be applicable to the military C4I system.

Brief Description		In the Military C4I System
Volume	Large amount of data	– The origin of the Network Centric Warfare concept and the emergence of many devices such as IoBT in the battlefield – Surveillance assets generate a large volume of data such as video/image, radio signal, etc.
Velocity	Speed of data growth Rate of data flow	– The increases in the IoBT and the development of information assets on the battlefield accelerate the rise in battlefield data. – High performance to support real (or near-real) time decision making in a military operation
Variety	Different types of data from different kind of data source.	– Intelligent elements such as HUMINT, SIGINT, MASINT, GEOINT, etc., create different types, the volume of battlefield data
Variability	Change of data flow rate, format, structure, volume, etc.	– The defense data center secures scalability by introducing a cloud system. – Conversion to data in a form suitable for C4I system terminal devices
Veracity	Data accuracy	– In the battlefield, it is important to ensure that the data can be trusted for military operation and military decision-making.
Value	Valuable input/output data	– Input data are provided by assets in the battlefield. – Output data are used for the combat assets by making a decision.

- Volume: The huge amount of data (or datasets) stored and processed in a big data system [17]. Since the US military highlighted the Network Centric Warfare [22] concept in 1998, information sharing between elements of the battlefield has become more active, particularly the IoBT and surveillance assets that shoot Video/Image have exponentially increased the amount of data to be processed in a C4I system. For example, terabytes (10^{12}) of data volume have become common to handle at the end-user level, and zettabytes (10^{21}) of data volume is no longer an unfamiliar word for military data managers. Due to the increase in the volume of data on the battlefield, Defense Data Center requires sufficient storage space to store it and technology to manage it efficiently.
- Velocity: Velocity, in big data, concerns mainly two aspects: the speed of data growth and the rate of data flow [18]. The speed of data growth is closely related to the characteristics of data volume introduced above. The increases in the IoBT and the development of information assets on the battlefield accelerate the rise in battlefield data. The data flow rate means that real-time (or near-real-time) data transmission/reception and processing capabilities are required to implement Observe-Orient-Decide-Act (OODA) in a real-time military operation. Improved wire/wireless communication network infrastructure provides data transmission with minimal latency, and advanced computing technology enables data processing (or analysis) in real time.
- Variety: Military intelligence types include human intelligence (HUMINT), signal intelligence (SIGINT), measurement and signature intelligence (MASINT), geospatial intelligence (GEOINT), open-source intelligence (OSINT), etc. This information is collected from the human, the IoBT, sensors, unmanned aerial vehicles (UAVs), satellites, etc., operated on the battlefield, and the collected data are stored in the database as structured, unstructured, or semi-structured formats. As a result, C4I systems must be able to process different forms of data from different kinds of sources on the battlefield.
- Variability: This feature refers to change in a dataset, whether in the data flow rate, format, structure, and/or volume. For example, data volume implies the need to scale up or scale down virtualized resources to efficiently handle the additional processing

load [19]. As a system architecture to handle the variability of data, the cloud can dynamically scale systems in virtual environments. It also applies relational databases and NoSQL DBs as data types change.

- Veracity: Just as input (data) must be accurate and reliable in I/O systems to achieve the desired results, data accuracy is one of the most important requirements in C4I big data systems. For example, as big data are used for decision-making, it is important to ensure that they can be trusted [23]. Inaccurate data such as repetition of meaningless data, noise, or typos may result in poor-quality output data. The system requires technology to eliminate inaccurate data. For instance, statistical methods, integrating/aggregating data with high precision technology, even machine learning algorithms, etc., are the technologies used to ensure data quality.
- Value: From a data point of view, both the input data collected in a C4I system and the output data generated after big data analysis should be of sufficient value. Therefore, by using a C4I system with big data, the commander must be able to make a decision and effectively support the activities of the combatants.

2.3. Five-Phase Big Data Lifecycle

The data, the input resource of big data, can be divided according to the procedure operated within the system. In research by the National Institute of Standards and Technology (NIST) [24], the data were processed by the big data application provider in five steps: collection, preparation/curation, analytics, visualization, and access. Patgiri [25] suggested as another five-step big data management cycle: acquisition, preprocessing, storage, analytics, and visualization. Both studies present the data management process from data generation/collection to visualization/utilization. However, with the recent increase in the security and privacy requirements for data used in the big data system, data destruction must be considered if the data are no longer necessary for the intended purpose or if the data provider withdraws consent. Koo et al. [26] proposed a five-phase data lifecycle that considers the disposal of data: collection, storage, analytics, used, and destruction. Therefore, we will apply those five-phase lifecycles to a C4I big data system.

- Collection: In practice, big data systems collect large volumes and diverse formats of data from several unique areas such as healthcare, economics/industry, smart cities, science, military, etc. Structured data are exemplified by consistently structured data and can be described efficiently in a relational model [19]. Structured data conform to a database model, which is largely characterized by the various fields to which data belong, such as name, address, age, etc., and by the data type for each field such as numeric, currency, alphabetic, name, date, and address [27]. Unstructured data refer to information that either does not have a predefined data model or are not organized in a predefined way. Photos, graphic images, video, text, voice records, streaming sensor data, and so forth can be categorized as unstructured data. Semi-structured data are a data type in which both the characteristics of structured and unstructured data are reflected. Word-processing documents, including metadata such as author name and created date, and photos uploaded to social network service (SNS) with tags are representative examples of semi-structured data.
- Storage: Big data systems are not simply satisfied with storing the raw data collected in the previous phase. In the storage step, the data preparation process (i.e., data aggregation and integration, data cleaning/cleansing, data partition, data indexing, etc.) must be included to store the large volumes and diverse formats of data appropriately. This phase briefly introduces two main technologies (i.e., distributed file system and MapReduce) to implement data storage. Distributed file systems are the most popular infrastructure that can store massive data sets in multiple distributed storage repositories [21,28]. MapReduce is a data-intensive programming model for processing large data sets in a cluster of distributed storage nodes [21,29]. Practically, Hadoop provides a framework for both Distributed File System and MapReduce.

- Analytics: This phase, big data analytics, generates useful knowledge by analyzing a large amount of previously collected and stored data. For example, Husamaldin et al. [20] suggested that big data analytics can be categorized into four aspects: descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics. To obtain meaningful information or knowledge, various techniques are used, such as statistical analysis, data aggregation, data clustering, machine learning, etc.
- Utilization: In each field to which big data technology is applied, the primary purpose of the utilization phase is to produce valuable information and knowledge through data analysis. For example, it is easier for the company to produce products desired by consumers in the commercial field by analyzing purchasing trends during Internet shopping. In addition, data collected for academic research can be quickly and accurately analyzed to provide valuable results. In the military domain, it provides evidence for commanders to make decisions and suggests the direction in which each combat element acts. Therefore, to effectively use big data, a visualization tool or a decision-making tool that accurately understands and expresses users' needs is essential.
- Destruction: As requirements regarding the security and privacy of big data become stricter, it has become crucial to manage data according to regulations. Basically, privacy data should be destroyed without delay after exceeding the data retention period unless otherwise specified in other laws and regulations. In addition, data must be destroyed if they are no longer necessary for the intended purpose or if the data provider withdraws consent [26]. In the military field, the destruction of data based on security regulations is an essential element to ensure the safety of military operations and protect the C4I system.

3. C4I System Framework with Big Data in the Battlefield Environment

In this section, the structural characteristics of the big data system and the C4I system are clearly understood, and the C4I big data system framework is newly presented by considering the relationship between the two systems. First, Section 3.1. defines a system framework classified into four layers by interpreting big data from a system perspective in order to understand the structural characteristics of a general big data system. The general big data system framework provides a basic guide for applying big data technology to the C4I system operated on the battlefield. Section 3.2. classifies the C4I system into three layers according to the flow of data created and managed in the battlefield. By classifying the C4I system, which includes complex and diverse components, into three layers from the data point of view, it can be said that it is a clear standard for applying big data technology was presented. Consequently, Section 3.3. newly presents the C4I big data system framework through structural fusion based on the correlation between the previously defined big data system framework and the C4I system framework. In addition, we examine the possibility of practical system implementation by presenting the necessary components for each part of the framework presented in detail.

3.1. Layered Big Data System Framework from a System Perspective

The layered big data system framework refers to a system structure adequately implemented to manage the data covered by the big data lifecycle. According to the system configuration perspective of big data, this paper proposes four layers of a framework: infrastructure, data, platform, and application layers. This classification extended from three frameworks (i.e., infrastructure, platform, and processing) of big data systems mentioned in NIST [24] and the three layers (i.e., infrastructure, big data fabrics, and big data platform as a service and application) presented by Kune et al. [21].

- Infrastructure Layer: This layer refers to physical or virtualized resources required in the entire process of the big data lifecycle for data collection, storage, analysis, utilization, and destruction. The infrastructure consists of a server, storage, network devices, the IoT, sensors, Internet web services, and so on. Additionally, in recent years, cloud computing, which is featured by excellent scalability and use of information system resources, on-demand resource provisioning, and ease of parallel computing, has been in the spotlight as a big data infrastructure that guarantees the 6V's.

- Data Layer: This layer stores and manages data in a big data system. To deal with the vast amounts and various forms of data in the data layer, it is necessary to have an appropriate system. For example, partitioning, indexing, and MapReduce technologies have been applied to distributed file systems to store vast amounts of data in multiple repositories and efficiently manage data. In addition, databases such as relational databases and NoSQL are required for various types of data.
- Big Data Platform Layer: The big data platform may be defined as middleware for implementing the big data function of the system [21]. The Hadoop ecosystem is a representative platform for implementing big data. For example, HDFS, MapReduce, YARN, etc., provide functions as unique modules for storing, processing, and managing data in big data systems; Cloudera, which is a commercial product that enhances security and management, is a representative of a Platform as a Service (PaaS).
- Application Layer: Depending on the field in which the big data system and processed data are used, appropriate analysis methods and visualization tools should be applied, which are covered by the application layer. In other words, algorithms for obtaining analysis results required by users or equipment and software that users can directly use are necessary in the application layer.

3.2. Three-Layered C4I System Framework in Terms of Battlefield Data Flow

On the battlefield, a C4I system is a tool that exchanges information between numerous networked battlefield elements under the concept of "Sensor to Shooter" and supports the commander's determination and shooter's action effectively. In particular, it is divided into three layers according to the processes of generating, processing, and using battlefield data used in the C4I system, which are expressed in the data generation, data processing, and data usage layers. The details of each layer are presented below and are depicted in Figure 1.

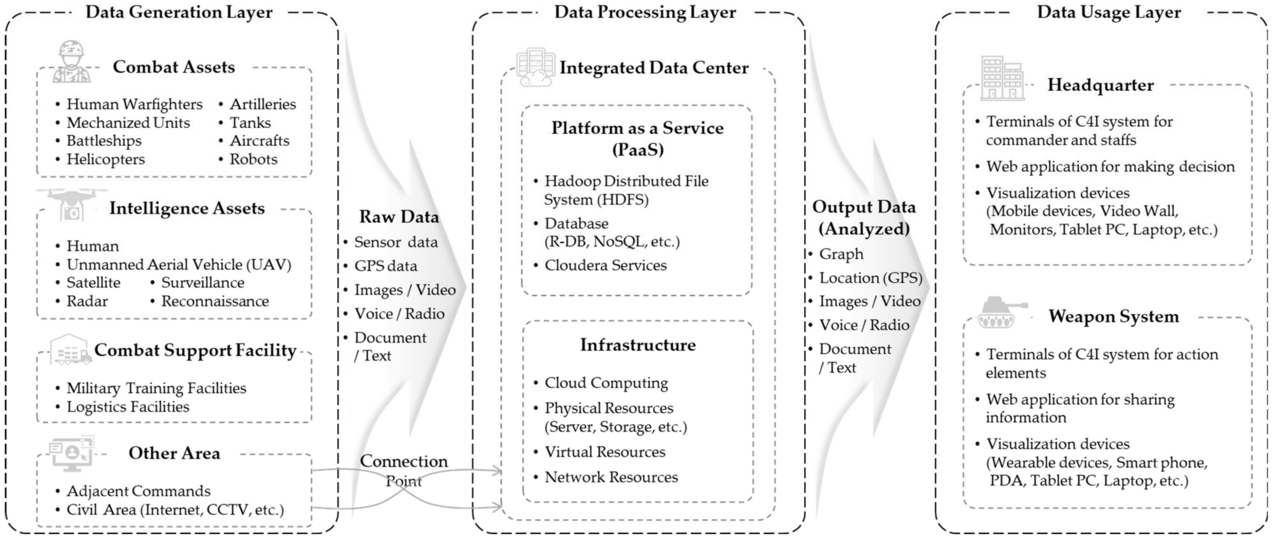

Figure 1. Three-layered C4I system framework classified according to the flow of battlefield data.

- Data Generation Layer: This layer generates large volumes and diverse types of battlefield data and transmits those data to the next step, the data processing layer. The elements that generate combat data are operated in a wide variety of places, such as combat areas, military camp areas, and civil areas. Examples of combat data generators are intelligence assets, combat assets, combat support facilities, other commands, the Internet, etc. Combat data also consist of sensor data from various

IoBTs, GPS signal, image/video, voice/radio signal, text/documents, etc. It contains almost all forms of data covered by a typical big data system.

- Data Processing Layer: This layer runs in the integrated data center, where major big data lifecycle phases are implemented to store and analyze the collected data. The data processing layer can be divided into infrastructure and platform (or Platform as a Service (PaaS)). Infrastructure means physical/virtual resources such as cloud computing, network devices, etc., and the big data system platform may be, e.g., Hadoop, Cloudera, etc.
- Data Usage Layer: This layer helps commanders make decisions by visualizing the data analyzed in the data processing layer as valuable information. Real-time combat information is directly available to the commander and staff who make situational judgments and decisions at headquarters and the combat/intelligence assets on the battlefield.

3.3. *Proposed C4I Big Data System Framework*

As mentioned previously, the three-layered C4I system framework was classified based on the data flow for the generation, processing, and use of battlefield data and had a close relationship with each stage of the big data lifecycle and layered system architecture, as discussed in Sections 2.3 and 3.1, respectively.

The correlation between the C4I system framework and the big data lifecycle is presented in Figure 2. First, the data generation layer mainly corresponds to data collection. The data-processing layer plays a role in storing and analyzing collected data, and the data usage layer is a step involving using valuable information provided by big data. Finally, destruction, in the big data lifecycle, is an action that occurs in all layers of the framework. As mentioned in a previous study [26], data security is an essential factor in the lifecycle; each layer of a C4I system should be able to delete unnecessary or illegal data immediately. Lifecycle steps are overlapped finely on each layer, which means that the matters necessary to link the framework steps exist in mutual stages.

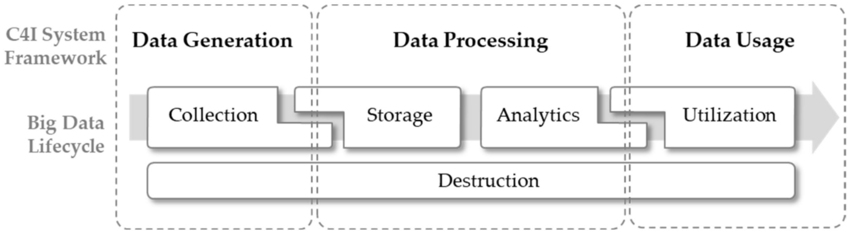

Figure 2. Description of the correlation between the three-layered C4I system framework and the big data lifecycle.

The four-layered big data framework is also related to the framework of the C4I system classified according to the flow of battlefield data, and Figure 3 expresses this association. For example, in the data generation layer of a C4I system, infrastructure can be referred to as IoBT devices and the network that connects those devices. As a result, Figure 3 suggests the components needed in each layer of a C4I system for each big data framework applied to the three-layered C4I system. However, the big data platform layer is a platform provided as a service to implement a big data system, and thus, the data generation layer and data usage layer are not applicable.

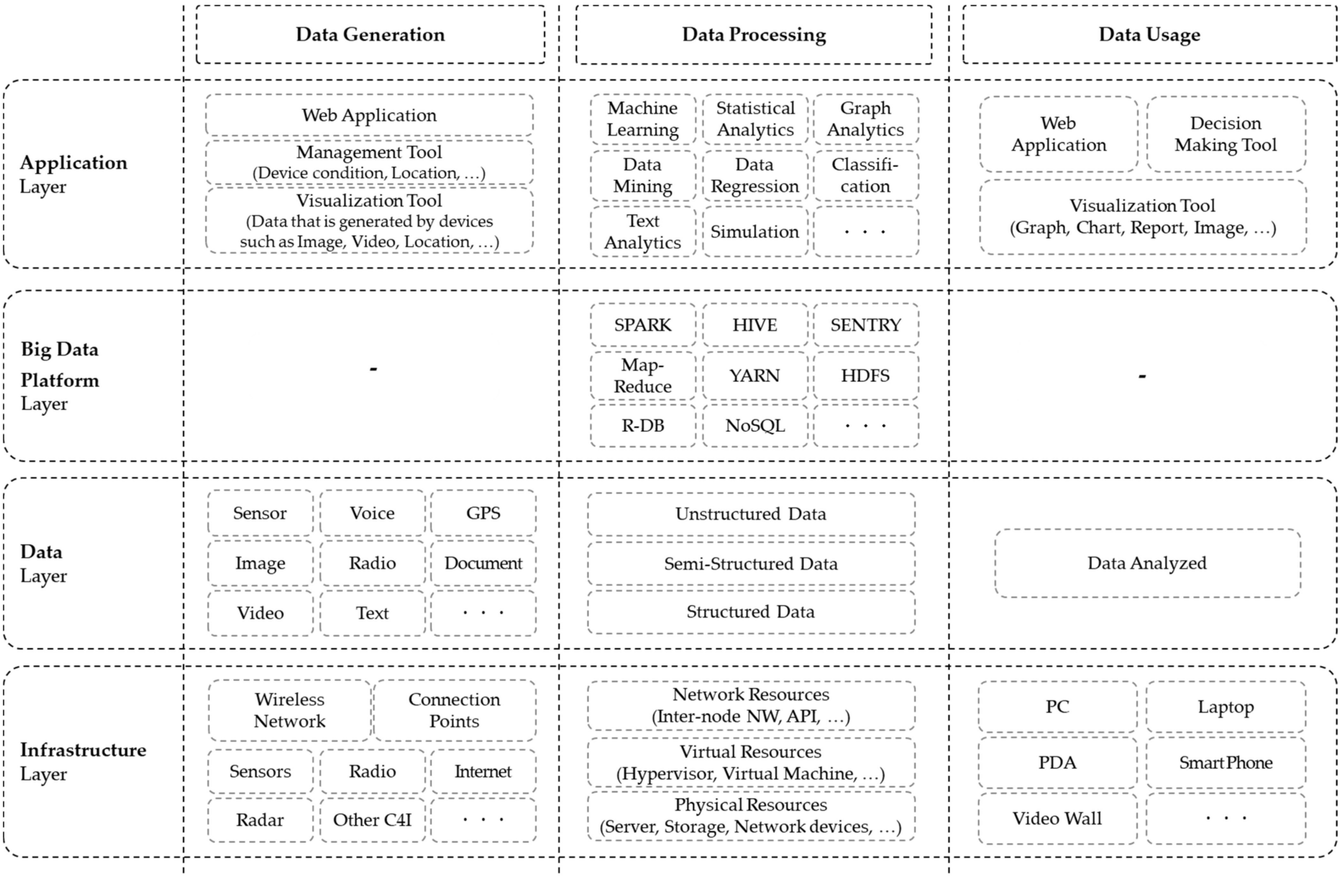

Figure 3. Description of the correlation between three-layered C4I system framework and big data system architecture.

4. Security Architecture for the Military C4I Big Data System

To safely operate big data systems within the boundaries of regulations related to information protection, Damiani et al. [30] and Cloud Security Alliance [31] have analyzed security threats, and several international organizations have proposed security standards and guidelines [32–34]. In addition, some studies [35,36] have presented security considerations applicable to general big data systems, whereas Abouelmehdi et al. [9] and Hossain et al. [37] mention security requirements that should be uniquely applied to specific fields such as healthcare and smart grids. This section more closely examines the security of a military C4I system with big data. By analyzing the security threats to be considered in the layers of the C4I system framework classified in Section 3 and presenting requirements, we propose a security architecture suitable for a C4I system with big data.

4.1. Three-Layered C4I System Framework in Terms of Battlefield Data Flow

C4I is a defense information system used in military operations that has strict security characteristics. First, all information covered in the C4I system is important and sensitive information that can affect national security. Therefore, a C4I system must safely manage data by maintaining a high level of security under any circumstances. Second, the enemy's grave threat to the friendly forces' C4I system is continual. Managers responsible for information security in a C4I system should analyze enemy threats preferentially and supplement them proactively. Third, like the distinct hierarchical structure of military organizations, the differences in C4I system user authority are also apparent. Because of these differences in users' authority, the system requires clear authentication and differentiates data use and access scope. Finally, as needed, a C4I system is an isolated network and has a connection point with an external network. Therefore, security measures for contact points with external networks are essential, and the verification of data introduced from the outside is necessary.

4.2. Security Threat and Requirements Related to C4I Big Data System Framework

In this section, we analyze security threats and present the security requirements for each of the three layers classified in the C4I big data system framework. Figure 4 comprehensively represents the security requirements applied to each layer of the C4I system framework, and details are provided in the subsection.

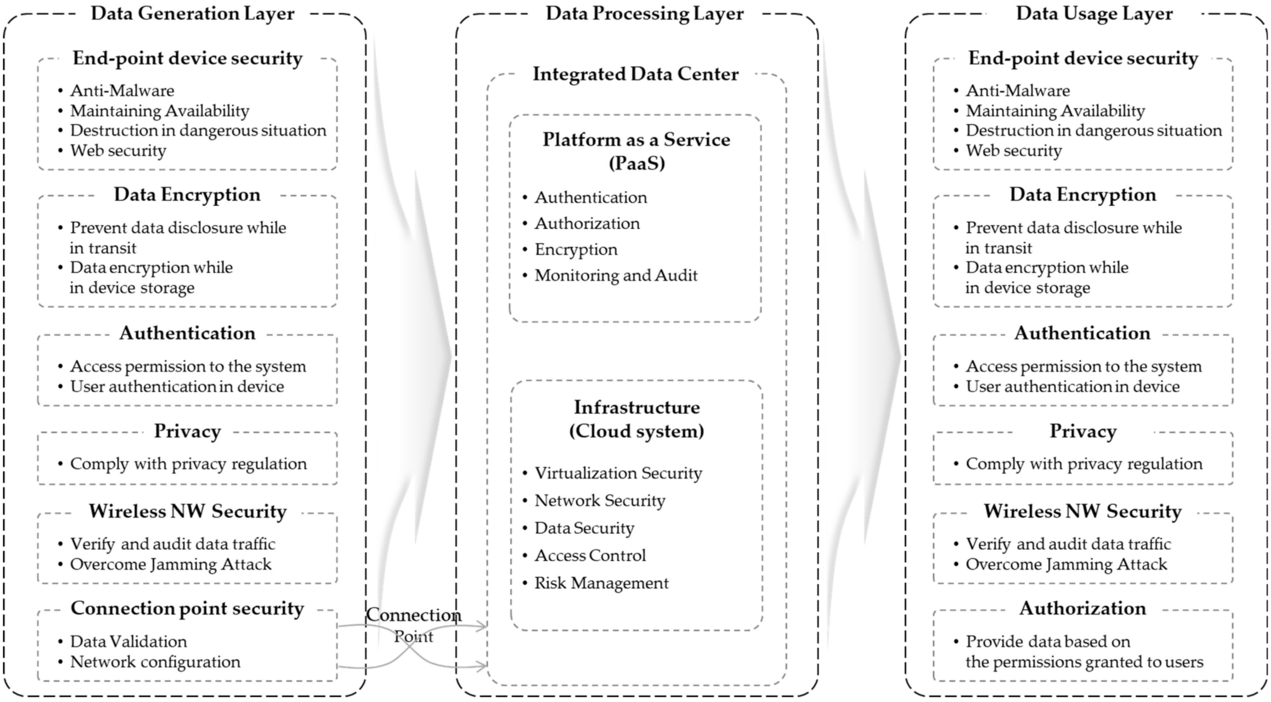
Data Generation Layer
End-point device security
• Anti-Malware
• Maintaining Availability
• Destruction in dangerous situation
• Web security
Data Encryption
• Prevent data disclosure while in transit
• Data encryption while in device storage
Authentication
• Access permission to the system
• User authentication in device
Privacy
• Comply with privacy regulation
Wireless NW Security
• Verify and audit data traffic
• Overcome Jamming Attack
Connection point security
• Data Validation
• Network configuration
Connection Point
Data Processing Layer
Integrated Data Center
Platform as a Service (PaaS)
• Authentication
• Authorization
• Encryption
• Monitoring and Audit
Infrastructure (Cloud system)
• Virtualization Security
• Network Security
• Data Security
• Access Control
• Risk Management
Data Usage Layer
End-point device security
• Anti-Malware
• Maintaining Availability
• Destruction in dangerous situation
• Web security
Data Encryption
• Prevent data disclosure while in transit
• Data encryption while in device storage
Authentication
• Access permission to the system
• User authentication in device
Privacy
• Comply with privacy regulation
Wireless NW Security
• Verify and audit data traffic
• Overcome Jamming Attack
Authorization
• Provide data based on the permissions granted to users

Figure 4. Proposed security requirements applied into the 3-layered C4I framework.

4.2.1. Data Generation Layer

In the data generation layer, many networked military assets generate a variety of and a large amount of battlefield data. Human fighters, mechanized units, UAVs, radars, etc., are the military assets, regarded as IoBT, that provide battlefield data. Based on the characteristics of the military-operated IoT, various studies have been conducted on the security requirements. In particular, among the studies that presented systematical security requirements, Mavroeidakos et al. [38] addressed the security threats, such as malware, botnet, ransomware, etc., and cyberattacks such as a physical layer attack, network attack, and man-in-the-middle attack, etc. Puthal et al. [39] also suggested possible security threats of IoT applications based on confidentiality, integrity, and availability, which is referred to as the CIA triad. Therefore, we intend to present the security requirements by analyzing the security threats of IoBT operating within the scope of the C4I system as follows.

As with the general IoT, the IoBT is vulnerable to malware infections; therefore, measures must be implemented to prevent infection with ransomware or abuse as a botnet. In addition, the physical space of this layer is a combat area, and adversaries act in the same space. If the enemy acquires the friendly forces' IoBT equipment, they would have unauthorized access to our networks, steal built-in information, and provide false information. Additionally, the enemy can hijack wired and wireless network signals to attack unauthorized network access through sniffing and spoofing. A C4I system is an isolated network. However, sometimes the data from another area, such as other levels of C4I systems or the Internet in the civil area, are also required. That is, connection points with external networks exist; therefore, the security management of those connections, which can be external data inflow channels, is crucial. Therefore, we propose six security requirements for the data generation layer.

- End-Point Security: The end-point devices that generate data may be referred to as the IoBT, which are connected through a network. The IoBT basically needs anti-malware functionality to compensate for the vulnerability to ransomware infection. In addition, considering the operating environment of the battlefield, the availability of the device should be increased, and data could be erased urgently in preparation for dangerous situations.
- Authentication: There are two considerations for the authentication in the data generation layer: The first is user authentication in the device. The device can verify that the user accesses the device by applying multi-factor authentication such as username/password and biometric authentication. The second is access permission to the system. By checking whether the device is registered as part of the system, network access control (NAC) may be applied to verify unique identifiers such as IP and MAC addresses of the device.
- Data Encryption: Encryption is necessary to ensure data confidentiality, and two points of view must also be considered: when data are collected and stored on a device and when data are transmitted and received over a network.
- Privacy Issue: Privacy-preserving issues for collected big data are a vital part of security. If individual combatants' personal information covered in a C4I system is collected, privacy issues must thoroughly be considered as well. For example, not only general personal information (e.g., the combatant's name, class, and gender) but also location information and bio-signals must be sensitively managed. Therefore, the system must comply with the country's privacy regulations such as the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA).
- Wireless Network Security: Confidence, availability, and integrity must be guaranteed for wireless traffic communication on the battlefield. Confidentiality can be strengthened through the data encryption (in-transit or at-rest) presented above, but further supplementary measures are needed to prevent interruptions to availability caused by jamming attacks and any harm to integrity caused by traffic hijacking or injection attacks from the enemy. Therefore, to protect the wireless network in the data generation layer, it is necessary to have the ability to overcome traffic jamming attacks and to prove that traffic has not been affected during transmission/reception by verifying data traffic.
- Connection Point Security: In an isolated network, connection points with external networks are vulnerable factors that should be dealt with carefully. In particular, to transmit the data from the external network to the inside, the network for the connection point must be configured safely, and the format and safety of the data must be verified and transmitted.

4.2.2. Data Processing Layer

1. Infrastructure Represented by Cloud Computing

Cloud computing has excellent scalability. The cloud can efficiently operate limited physical resources by dividing server performance and storage space; therefore, cloud computing is a representative infrastructure for implementing a big data system characterized by the 6V's. Our research group previously studied the security requirements for a cloud-based defense C4I system [40,41]. To derive the security requirements in our prior research, we referenced the security requirements identified by the Security Controls Baseline (FedRAMP) [42], Cloud Computing Security Requirements Guide (DISA) [43], Security Certification Guide for Cloud Service (Korea Internet and Security Agency) [44], Security Requirements for Server Virtualization System (Telecommunications Technology Association) [45], and the Security Guidelines of National and Public Institution for Cloud Computing (National Intelligence Service) [46]. The requirements consist of five categories with 28 items; virtualization security (9 items), data security (9 items), network security

(1 item), access control (4 items), and risk management (5 items) [40]. In this paper, two additional items for the connection point are proposed. Physical contacts with different levels of C4I systems or the Internet in the data generation layer exist in the data center's infrastructure. Therefore, two items were added: verification of data flowing through connection points for data security and security management of connection points existing in the isolated network in network security. The added contents are displayed in bold font in Figure 5.

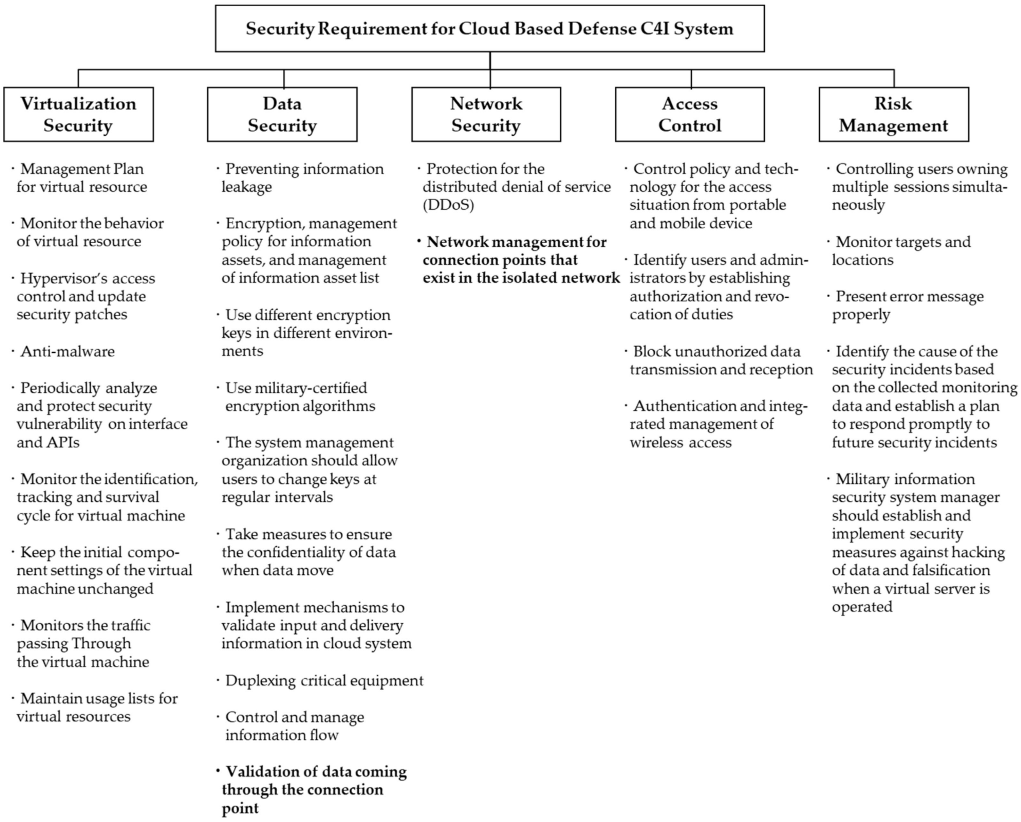
Security Requirement for Cloud Based Defense C4I System
Virtualization Security
Data Security
Network Security
Access Control
Risk Management
· Management Plan for virtual resource
· Monitor the behavior of virtual resource
· Hypervisor's access control and update security patches
· Anti-malware
· Periodically analyze and protect security vulnerability on interface and APIs
· Monitor the identification, tracking and survival cycle for virtual machine
· Keep the initial component settings of the virtual machine unchanged
· Monitors the traffic passing Through the virtual machine
· Maintain usage lists for virtual resources
· Preventing information leakage
· Encryption, management policy for information assets, and management of information asset list
· Use different encryption keys in different environments
· Use military-certified encryption algorithms
· The system management organization should allow users to change keys at regular intervals
· Take measures to ensure the confidentiality of data when data move
· Implement mechanisms to validate input and delivery information in cloud system
· Duplexing critical equipment
· Control and manage information flow
· **Validation of data coming through the connection point**
· Protection for the distributed denial of service (DDoS)
· **Network management for connection points that exist in the isolated network**
· Control policy and technology for the access situation from portable and mobile device
· Identify users and administrators by establishing authorization and revocation of duties
· Block unauthorized data transmission and reception
· Authentication and integrated management of wireless access
· Controlling users owning multiple sessions simultaneously
· Monitor targets and locations
· Present error message properly
· Identify the cause of the security incidents based on the collected monitoring data and establish a plan to respond promptly to future security incidents
· Military information security system manager should establish and implement security measures against hacking of data and falsification when a virtual server is operated

Figure 5. Security requirements for cloud-based defense C4I system.

2. Platform as a Service (PaaS)—Hadoop, Cloudera

Earlier, in Section 4.2.1. we looked at infrastructure security that implements big data systems as physical or virtual resources. We focus on the security of Hadoop and, in this section, a platform as a service (PaaS) of processing layers that store, manage, and analyze big data. Security overview [47], published by Cloudera, which provides services based on Hadoop, describes three security issues (i.e., authentication, encryption, and authorization). In addition, some studies [48–50] convey the meanings such as authentication, authorization, encryption, monitoring the system by expressing them such as access control, permission, data security at respite, data security internode communication, inspection, and classification. Therefore, the security requirements for a data processing platform of a C4I big data system could be organized into four categories: authentication, authorization, encryption, and security monitoring and audit.

- Authentication: This is a fundamental security requirement for any information system, users must prove their identity, and the system necessarily verifies whether the C4I big data system could be organized into four categories: authentication, authorization, encryption, and security monitoring and audit.
- Authentication: This is a fundamental security requirement for any information system, users must prove their identity, and the system necessarily verifies whether the user has accessibility. Clarifying the identity of users accessing the system amid the threat of continuous network penetration by the enemy is a necessary security factor

to ensure C4I big data system managers can operate the entire system stably. Authentication on the Hadoop platform uses many different access control technologies (i.e., Access Control Lists (ACLs), HDFS extended ACLs, and role-based access control (RBAC)) and applies the Kerberos mechanism. Kerberos is widely used as an authentication mechanism applicable to most Hadoop clusters such as HDFS, MapReduce, and YARN [47].

- Authorization: In a military C4I system, the range of access to data assigned to users varies widely depending on the hierarchical differences in the organization and the role of the staff. All activities within the Hadoop cluster, such as data access, use, view, and administrative modification, must be properly executed within the authority assigned to the user or administrator, and the authorization mechanism must be applied for this purpose. In Hadoop clusters such as HDFS, MapReduce, and YARN, access control is applied through POSIX-style permissions that grant permission to each file and directory. Subdivided ACLs are also applied, or Apache RANGER is used to manage authorization for each cluster [47].
- Encryption: Encryption is the last line of defense when a hacker obtains complete access to our data [51]. Sensitive data that need to be protected, whether stored in storage or in transit, must be encrypted so that its contents are not disclosed and must not be decrypted by unauthorized users. In military C4I systems, data encryption is a crucial security requirement that protects against confidentiality breaches due to enemy threats. Hadoop cluster guarantees encryption not only for data-at-rest but also for data-in-transit by applying transport layer security (TLS) and secure socket layer (SSL) [47].
- Security Monitoring and Auditing Log: The system security administrator should monitor the behavior of the Hadoop cluster and quickly recognize events that deviate from the set security criteria. Sometimes by leaving all actions generated in the Hadoop cluster, the stored log is analyzed to find the cause and effect of the problem. Ganglia and Nagios are open-source-based monitoring tools [48] that can also be applied to Hadoop, and the Cloudera manager also provides high-performance security monitoring capabilities for the big data platform [47].

4.2.3. Data Usage Layer

The results of big data analysis in a C4I system are visualized for the use of combat/intelligence assets at the command control headquarters. Therefore, both the decision-making tool and visualization device operating in the military C4I system must meet the security requirements of the end-point device. That is, all terminals operated in this layer are part of the IoBT connected to the network and have security characteristics similar to those of devices used in the data generation layer. However, due to the characteristics of C4I systems, the connection point does not exist in the data usage layer; therefore, the connection point security is excluded from the six security requirements suggested in Section 4.2.1. Then, since the range of information provided is limited according to the user's authority, authorization is added as a security requirement. Therefore, six security requirements (e.g., end-point device security, authentication, data encryption, privacy issue, wireless network security, and authorization) must be satisfied to ensure safe system operation.

4.3. Proposed Security Architecture for Big Data Implemented in a Military C4I System

Security architecture is a unifying framework and reusable services that implement policy, standards, and risk management decisions [51]. Designing security architecture requires an understanding of the overall system structure and specific security requirements. In other words, security architecture describes how the security countermeasures are positioned and how they relate to the overall systems architecture. Since we proposed a three-layered C4I system framework, four-layered big data architecture, and security requirements, a security architecture for big data implemented in a military C4I system can be designed. To understand C4I big data systems, we examined how the components of each layer of a C4I system are configured in the big data system architecture. The layers of security architecture are classified based on the layers presented in the big data architecture; infrastructure security, data security, big data platform security, and application security layers. For each layer, specific security requirements are positioned. Figure 6 shows the security architecture of the big-data-implemented C4I system. The proposed security architecture based on a C4I system structure can contribute to the technical implementation of the security requirements, the presentation of security policies and guidelines, and risk management when the C4I system applies big data.

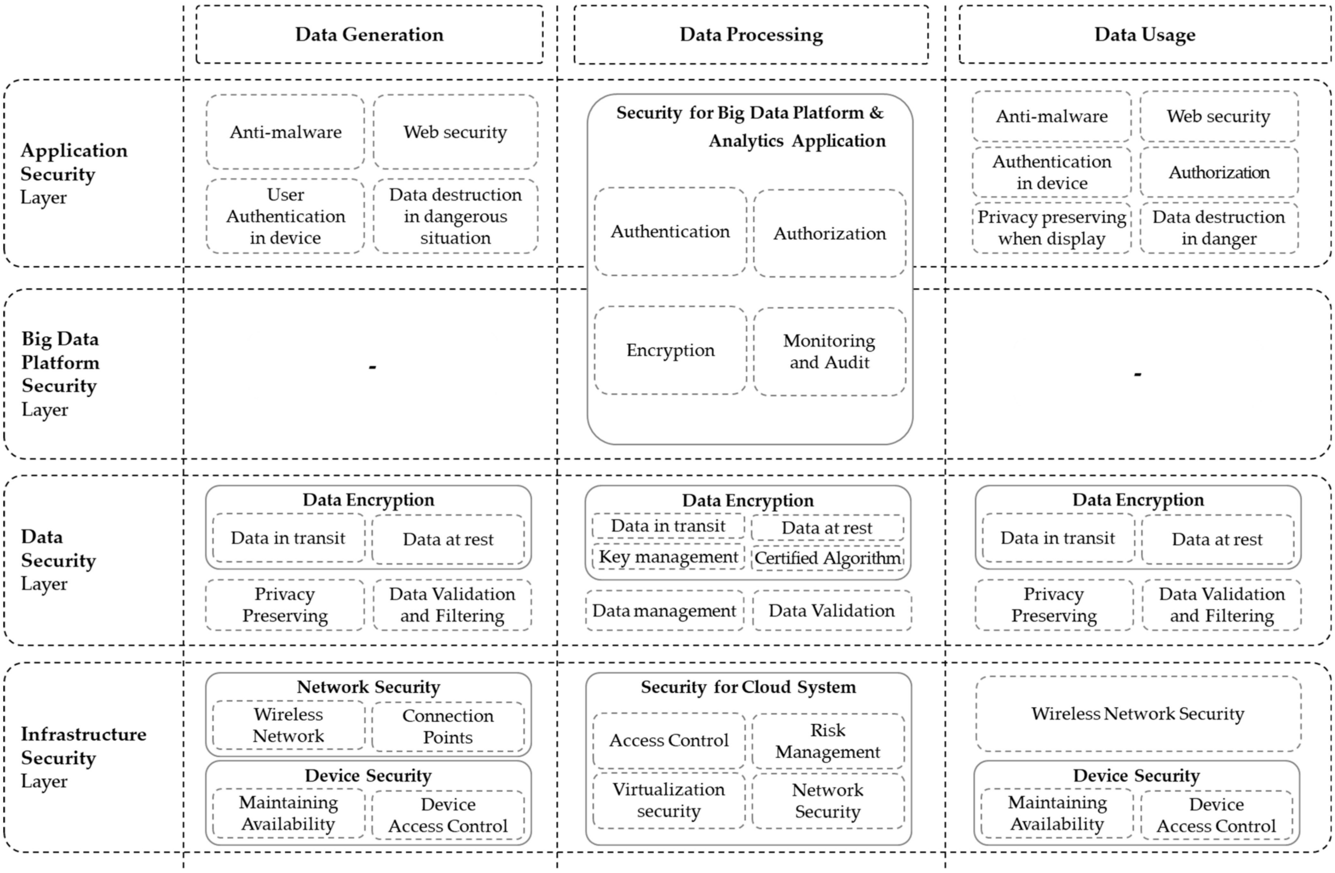

Figure 6. Proposed security architecture for big data implemented in a C4I system.

First, the infrastructure security layer requires security measures for physical or virtual resources and networks that constitute the layer of each C4I system framework. The infrastructure of the data generation and data usage layers are mainly composed of end-point physical devices and communicate over a wireless network; therefore, device security and network security are required. Since the infrastructure of the data processing layer applies the cloud computing system, the security criteria for the cloud system introduced into the military domain should be applied.

The data security layer enhances the level of security for the overall data covered by the system. Proper encryption and data verification are essential because the confidentiality and integrity of the data in the C4I system must be thoroughly guaranteed. All three layers of the C4I system must be encrypted when data are stored, transmitted, and received over the network, which is necessary not only to ensure data confidentiality but also to comply with privacy preservation requirements. In addition, in the data processing layer, data security requires more effort to safely manage data than other layers because big data for analysis are stored and handled in the data processing layer.

The big data platform security layer applies to the data processing with big data platforms such as Hadoop and Cloudera. Four security requirements are presented for this security layer: authentication, authorization, encryption, and monitoring and auditing. However, as mentioned before, data generation and usage layers do not use Hadoop or Cloudera as a big data platform; therefore, there is no need to consider security factors.

In an application security layer, security for user-centered software and hardware is defined for each layer. For example, the authentication of software directly accessed by users is strengthened, and the security of the web environment is required for data generation and usage layers. In the data processing layer, which is equivalent to big data analytics, security factors of the big data platform layer are also considered to require the same level of security because it provides Hadoop-based services.

As a result, the security architecture of the proposed big data C4I system can be said to be a structure of a security system essential for each hierarchical structure. When designing a C4I system with big data technology, the security architecture presents security points to consider.

5. Evaluation and Discussion

This paper proposes an improved new system framework and security architecture for the military big data applied to the C4I system. The system framework and security architecture provide a practical and systematic framework to establish a more safe C4I big data system efficiently. This section analyzes the proposed system framework and security architecture by comparing them with the previously studied big data in the military domain.

As presented in Table 2, evaluation fields related to the C4I big data system and its security, such as lifecycle and system architecture of big data, C4I big data system framework, security requirements, and security architecture, are analyzed. Previous studies, which are presented as comparative papers in Table 2, could not fully explain the structural characteristics and security issues of the C4I system to which big data are applied. Ahmad et al. [12] reviewed vulnerability, attack vector, and security countermeasures of the C3I system, but it could not carefully analyze the C3I system from the big data perspective. Alghamdi et al. [52] presented a model-based architecture framework for the security of the C4I system, but this study has limitations that are difficult to apply to the practical environment due to the lack of information on the structural characteristics of the C4I system. Moreover, other studies [53–55] were interested in the big data system framework, but these studies did not carefully deal with the security issues.

The C4I system has a very extensive and complex structure, from weapon systems operated on the battlefield to defense information systems that analyze and visualize data. The proposed C4I big data system framework contributes to designing a practical framework by analyzing the structure of the entire C4I system from the perspective of battlefield data. In addition, the elements, such as IoBT, cloud computing, Hadoop ecosystem, DBMS, machine learning, etc., presented for each part of the layered C4I big data system framework can provide the fundamental concept to construct a practical system.

The proposed security architecture analyzes security threats according to the roles and characteristics of each field classified in the system framework and systematically presents the necessary security elements. In particular, designing a specialized security architecture for the C4I big data system by aggregating security requirements for various technical elements based on the big data system is the beginning of academic research on big data security in the military domain. It contributes to improving the level of overall system security by presenting what security measures are required in which part of the practical system as an architecture. In addition, it provides security managers with a potential direction to safely manage the entire system.

Table 2. Comparison of the proposed method to those used in the military domain of big data application.

	System			Security	
	Big Data				
	Lifecycle	System Architecture	C4I System Framework	Requirements	Architecture
Ahmad et al. (2021) [12]	–	–	Four categories centered on the communication network	62 countermeasures based on 27 vulnerabilities and 22 attack vectors	–
Alghamdi et al. (2011) [52]	–	–	Department of defense architectural framework (DODAF) models	Briefly mentioned INFOSEC (confidentiality, integrity, and availability)	An example of a security architecture model based on cyber threats and operational situation
Shukla et al. (2018) [53]	Raw data, collection, filtering/classification, analysis, storage, and visualization	Briefly mentioned Hadoop core architecture	–	Authorization, auditing, and authentication	–
Zhang et al. (2019) [54]	–	–	Resource, capability, platform, management layer	Briefly mentioned about data security	–
Song et al. (2015) [55]	Data generation, management, and analysis	–	Five-layered architecture (repository, platform, service, application, and portal)	–	–
Proposed method	Collection, storage, analytics, utilization, and destruction	Four-layered architecture (infrastructure, data, platform, and application)	Three-layered framework (data generation, data processing, and data usage)	34 security requirements	Four-layered security architecture

Defense information systems, especially C4I systems, have very limited access and information disclosure according to security policies in the military. In particular, it is very difficult to obtain results by configuring an operational environment to verify the effectiveness of the system-wide framework and security architectures of the C4I big data system. In this study, the layered system framework and security architecture were proposed by analyzing the system's characteristics, but in the future, it must be studied in detail how to apply each requirement to a practical operating system. First, for military-used big data analytics, the security threats and security requirements that exist when using analytical technologies such as static methods, data classification, data mining, machine learning, etc., should be further developed in the military big data domain. Second, in addition to the security requirements presented in this paper, it is necessary to specify technical algorithms suitable for the defense sector for each requirement. For example, authentication was suggested as a necessary security measure for all presented layers, but the technical algorithms used for authentication in each layer are different. Presenting the most suitable algorithm for each environment is a challenge that requires further research. Therefore, to implement the required authentication, encryption, data validation, and private preservation for each part of the entire system, a more secure C4I big data system could be implemented if a practical algorithm considering the military environment was studied. Finally, it should be possible to analyze security considerations to strengthen the interoperability between IoBT and big data systems operated as one part of a defense C4I system. The IoT is already a source that provides input data to the system, and many studies have been conducted in connection with big data systems. However, research in the security field to safely and efficiently operate the IoBT and big military data requires further development.

6. Conclusions

This paper contributes to the safe and efficient use of big data in the defense domain by specifically presenting the security requirements necessary to apply big data technology to a defense C4I system. Defense C4I systems contain highly sensitive information, which can affect national security and the life of each combatant; therefore, they require a high level of protection because the presence of an enemy poses a severe and persistent threat. Therefore, to research how to operate a defense C4I system with big data safely, the system was first classified into three layers (data generation, data processing, and data usage layers), and this three-layered system framework was presented. This classification examined the C4I system in line with the functions of collecting, processing, and using data generated on the battlefield, and the C4I system was analyzed as a big data system by presenting the connection with both the big data lifecycle and a four-layered big data system architecture. By specifying the interrelationship between the three-layered C4I system framework and the four-layered big data system architecture, the components constituting each layer were clearly recognized, and the necessary security requirements were presented. The security architecture was layered into infrastructure security, data security, big data platform security, and application security. As a result, big data technology can be applied to C4I systems more safely and effectively by presenting specific security elements for C4I systems segmented according to functions and roles from sensors to shooters.

The US Department of Defense is trying to efficiently and safely establish a defense big data system by defining a platform for operating big data systems in the defense sector [56]. In addition, the importance of big data in the defense sector is increasing globally because each combatant and combat equipment produces and uses valuable information on battlefields. For example, as smart warriors, drones, unmanned robots, etc., are used in combat, efforts are being devoted to effectively analyzing the data generated here and using them in combat. In conclusion, if additional considerations are developed based on the security architecture presented in this paper, a security-enhanced big data system could be established in the defense sector.

Author Contributions: The authors contributed to this paper as follows: S.B. wrote this article and designed the system framework and security architecture; Y.-G.K. supervised and coordinated the investigation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C2012635).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Laurila, J.K.; Gatica-Perez, D.; Aad, I.; Blom, J.; Bornet, O.; Do, T.M.T.; Dousse, O.; Eberle, J.; Miettinen, M. From big smartphone data to worldwide research: The Mobile Data Challenge. *Pervasive Mob. Comput.* **2013**, *9*, 752–771. [CrossRef]
2. Anshari, M.; Lim, S.A. E-government with big data enabled through smartphone for public services: Possibilities and challenges. *Int. J. Public Adm.* **2017**, *40*, 1143–1158. [CrossRef]
3. Hajjaji, Y.; Boulila, W.; IFarah, R.; Romdhani, I.; Hussain, A. Big data and IoT-based applications in smart environments: A systematic review. *Comput. Sci. Rev.* **2021**, *39*, 100318. [CrossRef]
4. Marjani, M.; Nasaruddin, F.; Gani, A.; Karim, A.; Hashem, I.A.T.; Siddiqa, A.; Yaqoob, I. Big IoT data analytics: Architecture, opportunities, and open research challenges. *IEEE Access* **2017**, *5*, 5247–5261.
5. Ergüzen, A.; Ünver, M. Developing a File System Structure to Solve Healthy Big Data Storage and Archiving Problems Using a Distributed File System. *Appl. Sci.* **2018**, *8*, 913. [CrossRef]
6. Ali, W.; Shafique, M.U.; Majeed, M.A.; Raza, A. Comparison between SQL and NoSQL Databases and Their Relationship with Big Data Analytics. *Asian J. Res. Comput. Sci.* **2019**, *4*, 1–10. [CrossRef]
7. Roh, Y.; Heo, G.; Whang, S.E. A Survey on Data Collection for Machine Learning: A Big Data—AI Integration Perspective. *IEEE Trans. Knowl. Data Eng.* **2021**, *33*, 1328–1347. [CrossRef]
8. Ge, M.; Bangui, H.; Buhnova, B. Big data for internet of things: A survey. *Future Gener. Comput. Syst.* **2018**, *87*, 601–614. [CrossRef]

9. Abouelmehdi, K.; Beni-Hssane, A.; Khaloufi, H.; Saadi, M. Big data security and privacy in healthcare: A Review. *Procedia Comput. Sci.* **2017**, *113*, 73–80. [CrossRef]
10. Russell, S.; Abdelzaher, T. The internet of battlefield things: The next generation of command, control, communications and intelligence (C3I) decision-making. In Proceedings of the MILCOM 2018–2018 IEEE Military Communications Conference (MILCOM), Los Angeles, CA, USA, 29–31 October 2018.
11. Wang, M.; Cao, S. A survey on C4ISR system architecture technique. *Glob. J. Eng. Technol. Adv.* **2020**, *2*, 54–66. [CrossRef]
12. Ahmad, H.; Dharmadasa, I.; Ullah, F.; Babar, A. A Review on C3I Systems' Security: Vulnerabilities, Attacks, and Countermeasures. *arXiv* **2021**, arXiv:2104.11906.
13. Agadakos, I.; Ciocarlie, G.F.; Copos, B.; George, J.; Leslie, N.; Michaelis, J. Security for Resilient IoBT Systems: Emerging Research Directions. In Proceedings of the IEEE INFOCOM 2019—IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Paris, France, 29 April–2 May 2019.
14. Cho, S.; Hwang, S.; Shin, W.; Kim, N.; In, H.P. Design of Military Service Framework for Enabling Migration to Military SaaS Cloud Environment. *Electronics* **2021**, *10*, 572. [CrossRef]
15. Gupta, R.; Sultania, K.; Singh, P.; Gupta, A. Security for wireless sensor networks in military operations. In Proceedings of the 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, India, 4–6 July 2013; pp. 1–6. [CrossRef]
16. De Mauro, A.; Greco, M.; Grimaldi, M. A formal definition of Big Data based on its essential features. *Libr. Rev.* **2016**, *65*, 122–135. [CrossRef]
17. Laney, D. 3D data management: Controlling data volume, velocity and variety. *META Group Res. Note* **2001**, *6*, 1.
18. Patgiri, R.; Ahmed, A. Big Data: The V's of the Game Changer Paradigm. In Proceedings of the 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Sydney, NSW, Australia, 12–14 December 2016; pp. 17–24. [CrossRef]
19. Chang, W.L.; Grady, N. Big Data Public Working Group Definitions and Taxonomies Subgroup NIST Big Data Interoperability Framework: Volume 1, Definitions. *NIST Spec. Publ.* **2015**. [CrossRef]
20. Husamaldin, L.; Saeed, N. Big Data Analytics Correlation Taxonomy. *Information* **2010**, *11*, 17. [CrossRef]
21. Kune, R.; Konugurthi, P.K.; Agarwal, A.; Chillarige, R.R.; Buyya, R. The anatomy of big data computing. *Softw. Pract. Exp.* **2016**, *46*, 79–105. [CrossRef]
22. Pushkar, A. Network-centric warfare: Its origin and future. *Accel. World's Res.* **1998**, *124*, 28–35.
23. Hammad, K.A.I.; Fakharaldien, M.A.I.; Zain, J.M.; Majid, M. Big data analysis and storage. In Proceedings of the 2015 International Conference on Operations Excellence and Service Engineering, Orlando, FL, USA, 10–11 September 2015.
24. Chang, W.L.; Grady, N. NIST big data interoperability framework: Volume 2 Big Data Taxonomy. *NIST Spec. Publ.* **2017**, *1500*, 10. [CrossRef]
25. Patgiri, R. A Taxonomy on Big Data: Survey. *arXiv* **2018**, arXiv:1808.08474.
26. Koo, J.; Kang, G.; Kim, Y.-G. Security and Privacy in Big Data Life Cycle: A Survey and Open Challenges. *Sustainability* **2020**, *12*, 10571. [CrossRef]
27. Murthy, P.; Bharadwaj, A.; Subrahmanyam, P.A.; Roy, A.; Rajan, S. Cloud Security Alliance Report on Big Data Taxonomy. September 2014. Available online: https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Taxonomy.pdf (accessed on 25 August 2021).
28. Shvachko, K.; Kuang, H.; Radia, S.; Chansler, R. The Hadoop Distributed File System. In Proceedings of the IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), Incline Village, NV, USA, 3–7 May 2010; pp. 1–10.
29. Dean, J.; Ghemawat, S. MapReduce: Simplified data processing on large cluster. *Commun. ACM* **2008**, *51*, 107–113. [CrossRef]
30. Damiani, E.; Ardagna, C.A.; Zavatarelli, F.; Rekleitis, E.; Marinos, L. *Big Data Threat Landscape and Good Practice Guide*; European Union Agency for Network and Information Security (ENISA): Athens, Greece, 2016.
31. Big Data Working Group. Expanded Top Ten Big Data Security and Privacy Challenges. Cloud Security Alliance. 2013. Available online: https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenges.pdf (accessed on 25 August 2021).
32. Greene, T.; Shmueli, G.; Ray, S.; Fell, J. Adjusting to the GDPR: The Impact on Data Scientists and Behavioral Researchers. *Big Data* **2019**, *7*, 140–162. [CrossRef]
33. Stallings, W. Handling of Personal Information and Deidentified, Aggregated, and Pseudonymized Information Under the California Consumer Privacy Act. *IEEE Secur. Priv. Mag.* **2020**, *18*, 61–64. [CrossRef]
34. Khan, R.A.; Alka, K. An Improved Security Threat Model for Big Data Life Cycle. *Asian J. Comput. Sci. Technol.* **2018**, *7*, 33–39. [CrossRef]
35. Rajan, S. (Ed.) *Top 10 Big Data Security and Privacy Challenges*; Cloud Security Alliance: Seattle, WA, USA, 2012; Available online: https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Top_Ten_v1.pdf (accessed on 25 August 2021).
36. Yang, K.; Lee, D.; Kim, K.; Yoon, H. Analysis of Security Threat and Security Requirements of the Bigdata System. *J. Secur. Eng.* **2016**, *13*, 501–514. [CrossRef]
37. Hossain, E.; Khan, I.; Un-Noor, F.; Sikander, S.S.; Sunny, S.H. Application of Big Data and Machine Learning in Smart Grid, and Associated Security Concerns: A Review. *IEEE Access* **2019**, *7*, 13960–13988. [CrossRef]

38. Mavroeidakos, T.; Chaldeakis, V. Threat landscape of next generation IoT-enabled smart grids. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*; Springer: Cham, Switzerland, 2020.
39. Puthal, D.; Ranjan, R.; Chen, J. Big Data Stream Security Classification for IoT Applications. In *Encyclopedia of Big Data Technologies*; Springer: Cham, Switzerland, 2018. [CrossRef]
40. Koo, J.; Kim, Y.-G.; Lee, S.H. Design of Security Architecture for the Cloud-Based Korea Military Command and Control System. *J. Korean Inst. Commun. Inf. Sci.* **2020**, *45*, 400–408. [CrossRef]
41. Koo, J.; Oh, S.-R.; Lee, S.H.; Kim, Y.-G. Security Architecture for Cloud-Based Command and Control System in IoT Environment. *Appl. Sci.* **2020**, *10*, 1035. [CrossRef]
42. FedRAMP Security Controls Baseline. Available online: https://www.fedramp.gov/documents/ (accessed on 11 August 2021).
43. Cloud Computing Security Requirements Guide. Available online: https://public.cyber.mil/dccs/dccsdocuments/ (accessed on 11 August 2021).
44. Security Certification Guide for Cloud Service. Available online: https://www.kisa.or.kr/public/laws/laws3_View.jsp?cPage=1&mode=view&p_No=259&b_No=259&d_No=91&ST=&SV= (accessed on 11 August 2021).
45. Security Requirements for Server Virtualization System. Available online: http://www.tta.or.kr/data/ttas_view.jsp?rn=1&pk_num=TTAK.KO-10.0708 (accessed on 11 August 2021).
46. *Security Guidelines of National and Public Institution for Cloud Computing*; National Intelligence Service: Seoul, Korea, 2016; pp. 101–103.
47. Cloudera Inc. Cloudera Security Overview. August 2020. Available online: https://docs.cloudera.com/cdp-private-cloud-base/7.1.3/security-overview/cm-security-overview.pdf (accessed on 15 August 2021).
48. Perwej, Y. The Hadoop Security in Big Data A Technological Viewpoint and Analysis. *Int. J. Sci. Res. Comput. Sci. Eng. IJSRCSE* **2019**, *7*, 1–14.
49. Shrihari, M.R.; Manjunath, T.N.; Archana, R.A.; Hegadi, R.S. Research Challenges in Big Data Security with Hadoop Platform. In *International Conference on Recent Trends in Image Processing and Pattern Recognition*; Springer: Singapore, 2019; Volume 1037. [CrossRef]
50. Martis, M.; Pai, N.V.; Pragathi, R.S.; Rakshatha, S.; Dixit, S. Comprehensive Survey on Hadoop Security. *Emerg. Res. Comput. Inf. Commun. Appl.* **2019**, 227–236. [CrossRef]
51. Peterson, G. *Security Architecture Blueprint*; Arctec Group, LLC.: Graz, Austria, 2007.
52. Alghamdi, A.S. Proposed methodology to enhance C4I systems security on architectural level. *Sci. Res. Essays* **2011**, *6*, 6095–6103. [CrossRef]
53. Shukla, V.; Singh, B.; Kumar, M.; Negi, K. Big Data Analytics in C4I Systems. In Proceedings of the 2018 International Conference on Automation and Computational Engineering (ICACE), Greater Noida, India, 3–4 October 2018.
54. Zhang, J.; Wang, G.; Wang, S. Command and Control System Construction in Big Data Era. *J. Phys. Conf. Ser.* **2019**, *1168*, 032022. [CrossRef]
55. Song, X.; Wu, Y.; Ma, Y.; Cui, Y.; Gong, G. Military simulation big data: Background, state of the art, and challenges. *Math. Probl. Eng.* **2015**, *2015*, 298356. [CrossRef]
56. *DoD Digital Modernization Strategy*; Department of Defense: Arlington, VA, USA, 2019.