Study on Trends and Predictions of Convergence in Cybersecurity Technology Using Machine Learning

Sungwook Ryu[1], Jinsu Kim[2], Namje Park[3]*

[1] Graduate School of Future Strategy, Korea Advanced Institute of Science and Technology, Korea
[2] Graduate School of Department of Convergence Information Security, Jeju National University, Korea
[3] Department of Computer Education in Teachers College, Jeju National University, Korea
ryu76@kaist.ac.kr, kimjinsu@jejunu.ac.kr, namjepark@jejunu.ac.kr

Abstract

The indiscriminate convergence of technologies makes prediction difficult and can cause many difficulties in technology investment. This makes it difficult to choose capital investment and can induce excessive investment in inefficient technologies. Therefore, analyzing the trend of convergence technology and predicting a highly influential convergence area in the future can induce effective investment, and lead the highly influential technology to achieve great technological development. The purpose of this paper is to analyze technologies that are expected to have high influence in the future through prediction of major fusion areas and to present fusion areas that can be used as indicators of investment. The proposed mechanism selected four prominent journals in the security area and collected metadata to generate a dataset in terms of technological excellence and a dataset in terms of commercialization through patent metadata collection. Thereafter, a process of extracting a main keyword according to a topic from a metadata set by applying a Latent Dirichlet Allocation (LDA) is performed. The extracted topics and keywords are not related to topics and keywords of other years. Therefore, a dynamic topic model (DTM) is applied to analyze the trend of the extracted topics and perform prediction. DTM analyzes the topics in the fusion area classified by LDA and the trend of changing topics linked by year for each topic keyword. Finally, the association of the fusion region is analyzed to derive a fusion region with high influence. These results are believed to be used as an indicator of effective technology investment by providing a high impact area in the convergence area of cybersecurity.

Keywords: Echnology convergence, Cybersecurity technology, Machine learning, Latent Dirichlet Allocation, Dynamic topic models

1 Introduction

As commercialization of innovative technologies based on AI that are leading the 4th industrial revolution, including IoT, cloud computing, big data, 5G, robotics, and drones, are accelerating convergence among cyber technologies in various fields, the scale of cyber systems is being expanded. Accordingly, as advanced technologies converged diversely are being applied to cyber systems, the conditions and behaviors of dangerous fields and elements that induce serious security threats have become exceedingly complex as well [1].

Therefore, it is crucial to concentrate on common cyber safety challenges shared among cyber fields and systems across several industries. In this regard, the importance of cyber safety management that reduces major risk factors by achieving convergence of security technology in core fields of the 4th industrial revolution has been highlighted.

As advancements in the 4th industrial revolution usher in the "hyper-connected" era in which IT, physical, OT and IoT systems that previously operated on separate, individual networks become interconnected, core technologies that enable the 4th industrial revolution, including IoT, big data, AI, and cloud computing, are advancing our lives in a remarkable way [2-3].

In this study, major keywords of cybersecurity convergence fields and technologies are extracted and analyzed, with a focus on probabilistic topic modeling based on text analysis of papers and patents. To that end, this study major topics and core keywords are extracted from papers and patents by using the Latent Dirichlet Allocation algorithm that can be used to estimate both the distribution of keywords by topic and the distribution of topics by document. Such results offer information that can be used to identify threats.

2 Related Studies

2.1 Representative International Standards on Cybersecurity

As with information security, cybersecurity aims to achieve confidentiality, integrity, and availability. As types of attacks that threaten such goals of cybersecurity evolve along with technological advances, such attacks cause serious problems. To overcome this challenge, international standards on cybersecurity are being developed and various research on cybersecurity technology is underway [4-5].

International standards on cybersecurity are being established based on the ISO/IEC 27000 series, which is known as the existing information security management system and is being developed as specialized standards

*Corresponding Author: Namje Park; E-mail: namjepark@jejunu.ac.kr
DOI: 10.53106/160792642023052403016

targeting cybersecurity.

ISO/IEC 27100 (ISO/IEC 27100, 2019) provides an explanation of overall concepts of cybersecurity as well as cybersecurity related definitions used in series standards, where as ISO/IEC 27101 (ISO/IEC 27101, 2019) offers guidelines on developing and establishing an organizational cybersecurity framework. ISO/IEC 27102 (ISO/IEC 27102, 2019) provides guidelines on introducing cyber insurance as a way to manage the impact of a cyber incidents within the organization, while ISO/IEC 27103 (ISO/IEC 27103, 2019) demonstrates how an organization can utilize cybersecurity framework based on information security standards to achieve a systematic approach to cybersecurity management [6].

ISO/IEC 27032 (ISO/IEC 27032, 2012) provides an overview of cybersecurity, an explanation of the relationship between cybersecurity and other types of security, a definition of stakeholders and a description of their roles in cybersecurity, as well as guidance for addressing common cybersecurity issues.

2.2 Evolution of Cyber Threats

Checkpoint, the world's largest developer of security solutions, categorized cybersecurity protection into five categories [7-8]. The first generation was protected by antivirus development in a virus attack on standalone PCs. As the importance of the Internet grew in the second generation, crime was organized and malicious software emerged. Accordingly, protection was carried out through intrusion detection systems and firewalls [9]. In the third generation, the analysis of networks and software was initiated as a way to analyze vulnerabilities across the IT infrastructure, triggering intrusion prevention systems to prevent intrusion. In the fourth generation, cyberattacks are becoming larger, resulting in more sophisticated levels of attacks [10-11]. This attack has generally difficulty identifying properties, such as avoidance and polymorphism. Cyber threats in the fifth generation have evolved into largescale multivector attacks that infect a wider range of targets as attack tools are strengthened [12].

As such, the company predicts that real time response to mega attacks will be the core of nextgeneration security technology development, while implementation of such concepts as Zero Trust and SOAR for ID verification before authorizing access to all objects attempting to access a system will form the center of 5th generation cybersecurity technology development [13-14].

2.3 Analysis of Related Research Trends

This paper collects the metadata of four prominent journals and patents from 2010 to 2020 in the cybersecurity field, and categorizes major keywords into 10 topics annually based on DTM based on the collected data. LDA was used for year-by-year correlation analysis. Finally, after performing the association analysis, a fusion region between topics was derived. The purpose of this is to analyze the trend of technology by year by using machine learning based on the past metadata, and finally derive a major convergence area and use it as one of the indicators. Therefore, the analysis of related research trends was performed on the field of deriving future convergence technology areas by collecting metadata.

The study of Tahereh Saheb (2019) noting that the big data of IoT devices in the medical field is greatly influenced by the IoT Big Data Analytics (IoTBDA) Paradigm on the IoT based design, development and application of medical services. A study was conducted to confirm whether In this paper, through a review of 46 papers on IoTBDA and 86 papers on fork computing in the medical industry, IoTBDA convergence derivative factors were analyzed into three types [15].

The study of Yuan Zhou (2019) suggested a new method to analyze the convergence domain based on scientific knowledge. The study was conducted in the bioinformatics field for empirical analysis, and two stages were presented: an incubation stage and a stabilized development stage to analyze the fusion region. This study collected 8,678 papers in the field of biotechnology and 39,715 papers in the field of information technology using scientific publication citation network data for analysis. In the proposed method, Newman topology clustering for convergence cluster identification, LDA technique for cluster topic identification, and Citation-Network Data Analyzer (CDA) for intuitive representation, computer aided visualization tools for citation networks were developed [16].

A study by TaeSan Kim (2020) suggested a framework for convergence domain prediction based on machine learning. In this study, 381,062 patent metadata for signal transmission, communication and automobile fields that occurred from 2009 to 2017 were used as a dataset based on patent data from the United States Patent and Trademark Office (USPTO). The Doc2Vec technique was used to identify the semantic relevance of technologies, and Artificial Neural Networks (ANNs), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF) were applied to perform prediction [17].

3 Cybersecurity Technology Convergence

3.1 ICT Strategic Technology Trends

In order to perform a study to identify cybersecurity convergence fields and related core technologies, and to suggest the future direction of cybersecurity technology convergence, one must first understand trends in ICT where cybersecurity technologies are being applied and utilized [18]. Top Ten Strategic Technology Trends announced every year by Gartner, a global leading IT research company, are well known predictions of trends in the IT industry presenting the company's anticipation of how the industry will change (Panetta, 2019). Table 1 below shows Gartner's predictions of top ten strategic technology trends in the IT sector from 2011 to 2020.

The duration of a certain technology staying on the top ten list selected by Gartner varies depending on how long it takes for that technology to take root, and some technologies may disappear in a very short time after making the list or reappear after skipping one year [19]. Just like IT technologies that have a very close interrelation with one another, a new type of hyper security to which various forms of converged security technologies crossing over various areas are applied is being emphasized for security technologies of cyber systems to which IT technologies are applied.

Table 1. Gartner's IT top ten strategic technology trends for 11 years (2011-2020)

roto2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Cloud Computing	Cloud computing	Media tablets and beyond	Mobile device battles	Mobile device diversity and management	Computing everywhere	The device mesh	AI and advanced machine learning	AI foundation	Autonomous things	Hyperautomation
Virtualization for availability	Mobile applications and media tablets	Mobile-centric applications and interfaces	Mobile applications & HTML 5	Mobile apps and applications	The internet of things	Ambient user experience	Intelligent apps	Intelligent apps and analytics	Augmented analytics	Multiexperience
Reshaping the data center	Social communications and collaboration	Social and contextual user experience	Personal cloud	The internet of everything	3D Printing	3D printing materials	Intelligent things	Intelligent things	AI-driven development	Democratization of expertise
IT for green	Video	Application stores and marketplace	Internet of things	Hybrid cloud and IT as service broker	Advanced, pervasive and invisible analytics	Information of everything	Virtual and augmented reality	Digital twin	Digital twins	Human augmentation
Client computing	Next generation analytics	The internet of everything	Hybrid it and cloud computing	Cloud/client	Context-rich systems	Advanced machine learning	Digital twin	Cloud to the edge	Empowered edge	Transparency and traceability
Mobile applications	Social analytics	Next-generation analytics	Strategic big data	The era of personal cloud	Smart machines	Autonomous agents and things	Blockchain and distributed ledgers	Conversational platformst	Immersive experience	The empowered edge
Advanced analytics	Context-aware computing	Big data	Actionable analytics	Software-defined anything	Cloud/client computing	Adaptive security architecture	Conversational system	Immersive experience	Blockchain	Distributed cloud
Social computing	Storage class memory	In-memory computing	Mainstream in-memory computing	Web-scale IT	Software-defined applications and infrastructure	Advanced system architecture	Mesh app and service architecture	Blockchain	Smart spaces	Autonomous things
Security -- activity monitoring	Ubiquitous computing	Extreme low-energy servers	Integrated ecosystems	Smart machines	Web-scale IT	Mesh app and service architecture	Digital technology platforms	Event driven	Digital ethics and privacy	Practical blockchain
Flash memory	Fabric-based infrastructure and computers	Cloud computing	Enterprise app stores	3D printing	Risk-based security and self-protection	Internet of things platforms	Adaptive security architecture	Continuous adaptive risk and trust	Quantum computing	AI security

3.2 Trends in Cybersecurity Technology

To estimate next-generation security technology, it is useful to refer to the "Hype Cycle" developed by Gartner, as shown in Figure 1. This cycle consists of five phases - Technology Trigger, Peak of Inflated Expectations, Trough of Disillusionment, Slope of Enlightenment, and Plateau of Productivity. Of these, next generation security technology areas are mostly at the first phase, or Technology Trigger, and the second phase, or Peak of Inflated Expectations.

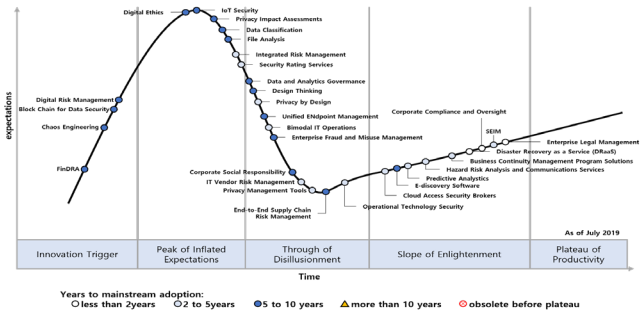


Figure 1. Risk management based prediction of next generation security technology areas by Gartner

Based on the analysis of the above data, next generation security technology is anticipated to move towards integrated cyber risk management that can respond in real time to mega attacks targeting all sectors of cybersecurity, in order to achieve safety, reliability, and resilience of hyper connected platforms based on IoT and 5G.

Therefore, it is predicted that the development of next generation security technology will move away from technology development centering on security equipment, which has been the mainstream area up to the 4th generation security, towards Predictive Intelligence that uses AI related techniques [20]. Examples include Behavioral Analytics/ Anomaly Detection, which detects abnormal behaviors through automatic learning of normal behaviors, Automated Security that protects large traffic at endpoints for processing. Moreover, vigorous research will be carried out at the national level on the Cyber Kill Chain [21-22].

3.3 Cybersecurity Fields

In this study, First, core keywords related to the top ten strategic technologies selected in the annual IT strategic technology trends from 2010 to 2020 were extracted [23-24]. Authors and experts considered several dozens of keywords through constant online meetings and several offline meetings, our selection was narrowed down to a few trends. For instance, the strategic technologies and keywords extracted from the Top Ten Strategic Technology Trends of 2019 are as shown in Table 2

Based on the compressed strategic technology trends, as shown in Table 2, main categories are first created, and then middle categories are defined by using technology names from the Hype Cycles from 2016 to 2019, and finally, small categories are determined by using specific names of relevant technologies (keywords). From 2010 to 2020, we secured keywords for the cyber security field based on the keywords of IT strategic technology trends every year [25]. Table 3 shows the final cyber security area obtained in this way.

Table 2. Strategic technologies and keywords

Strategy technology	Keyword
Artificial intelligence	Big Data, Artificial intelligence, Machine Learning, Deep Learning
5G	ALL IP, Network Slicing, Network Neutrality, Cord-cutting
Robot	Artificial intelligence, Smart Factory, Human-Robot Collaboration, Humanoid
Drone	Quadcopter, 4 Channels, Calibration, Hovering, Flight Controller
Interactive platform	Smart Secretary, Natural language processing, Amazon Skills, Corpus, Chatbot Builder
Realistic media	Realistic media, Virtual Reality, Augmented Reality, Mixed Reality
Blockchain	Distributed Ledger Technology, Blockchain, Initial Coin Offering, Byzantine Generals Problem, Consensus algorithm, Oracle Problems

4 Research Methodology

4.1 Introduction of Research Process

For keyword collection, we analyzed four prestigious academic papers (ACM CCS, USENIX Security, IEEE Security & Privacy, NDSS) in the field of information security, and collected metadata from patents and papers. In this paper, keywords were analyzed in about 4,200 papers and 3100 patents. By classifying the collected keywords over time, we conducted an analysis of the new convergence technology field and the dying convergence technology field.

Technological forecasting methodology has been developed in various forms depending on the purpose and scope of forecasting, the characteristics of the technology, and the level of data accumulation. More recently, several methodologies are combined for use rather than individually used, and each user modifies these methodologies differently in their own optimized way.

Table 3. Cybersecurity areas

Main category	Middle category	Small category	Hype cycle technology name
Network Security	Wired network security	Perimeter security	IDPS
		Secure connection	
		DDoS response	
	Wireless network security	Mobile communication network security	Mobile Threat Defense
		Wireless local area network security	
	Cloud security	Virtualization platform security	Cloud Security Assessments
		Cloud security service	
		Software defined security	
Data and application service security	Application security	Web security	Secure Web / Gateways
		Email security	
		Database security	
	Data security	Privacy protection	Data Loss / Prevention
		Data leakage prevention	
		Digital copyright infringement/ right protection	
	E-money, Fin-tech security	E-money security	The Programmable / Economy
		Blockchain security	
		Electronic transaction / abnormal behavior detection	
		Prevention of transactions and fraud	
	Digital forensics	Digital Evidence / Collection and Analysis	E-Discovery Software
		Anti-Forensic Response	

Physical security	Human/bio recognition	Biometric sensor	Biometric / Authentication / Methods
		Biometric engine	
		Human recognition and search	
		Human/bio recognition application	
	CCTV surveillance/control	Camera and storage device	Video/Image Analytics
		VMS/integrated control	
		Intelligent video surveillance	
		CCTV infrastructure protection	
	Secure search and unmanned electronic guard	Interpersonal Search Machine	
		Luggage/Luggage Finder	
		Alarm monitoring	
		Unmanned electronic security service	
System and password security	Cryptographic technique	Password design	Database Encryption
		Cryptographic side channel analysis	
		Password analysis	
	Certification/authorization technique	Universal authentication	Externalized Authorization Management
		ID management and access control	
		Bio certification	
	Security vulnerability	SW vulnerability analysis	Vulnerability Assessment
		HW vulnerability analysis	
	System security	Operating system security	Data Loss Prevention
		Virtualization security	
		System access control	
	Malware	Malware response	DLP for Mobile Devices
		Ransomware response	
	Threat analysis and control	Intelligent cyber threat analysis	SIEM
		Security information analysis and log management	
		Security control	
IoT security	Home city security	Home·City Device Security and Control	Smart City Framework
		Home City Data Privacy	
	Industrial control system security	Smart factory security	Operational Technology Security
		Infrastructure security	
		Smart energy security	
	Vehicle security	Communication security inside and outside the vehicle	Mobile Device Integration Into Automobiles
		Access control inside and outside the vehicle	
		Car intrusion detection	
		Vehicle security vulnerability diagnosis	
	Ship, ocean and air security	Prevention of hacking of autonomous ships	Autonomous Vehicles
		Shipping port communication security	
		Marine infrastructure security control	
		Unmanned vehicle security	
		Aviation infrastructure security control	
	Healthcare, Medical security	Healthcare device/sensor security	Real Time Health System Command Center
		Medical data security and sharing	
	Other ICT security	Artificial intelligence and robot security	IoT Securities

Next, the algorithms that were applied to this study's actual experiment as probabilistic topic modeling techniques, or LDA, DTM and more specifically, are introduced.

Latent Dirichlet Allocation, or LDA, is a topic modeling technique used to detect potential topics from a large number of given documents. LDA clusters documents by using topics that are latent in documents. With LDA, it is assumed that documents include more than one topic and the n-th word is chosen from each document. Then the topic with the greatest value obtained from multiplying the probability distribution of the topic in documents that contain the word by the probability distribution of the word against the topic is assigned to the n-th word. In other words, LDA is a probability model that shows what types of topics exist in each given document. Modeling methodologies for other trend analysis include dynamic topic modeling.

DTM uses a probabilistic time series model to analyze the evolution of topics over the lapse of time. The LDA can find special topics every year within a set year.

However, there may be no relationship between selected topics, which may be a problem in analyzing trends. However, the trend analysis must show a link to the annual topic change. DTM allows users to understand how the ratio of literature belonging to such topics and the topics themselves change by each period.

5 Research Experiment

5.1 Experimental Data

In this paper, we collect the articles of the prestigious Society of Information Security and the keywords for patents. About 1,600 papers from ACM CCS, 800 papers from Usenix Security, 600 papers from NDSS, and about 1,200 papers from the IEEE Security & Privacy journal were collected, and about 3,100 patents from information security related fields were collected. In addition to the papers and patents, related meta data were also collected. Table 4 below shows examples of specific information and meta data from the papers and patents collected. Table 5 is a summary of 10 topics by dividing groups with similar tendencies into topics by applying LDA to keywords collected in papers and patents.

Table 4. Examples of information and meta data from papers and patents

Type		Conference and journal scope and patent search keywords	
Paper scope	ACMCCS	Cryptography, Formal methods and theory of security, Security services, Intrusion/anomaly detection and malware mitigation, Security in hardware, etc	
	Usenix security	System security, Network security, Security analysis, Data-driven security and measurement studies, Privacy-enhancing technologies and anonymity, etc	
	NDSS	Cyber-crime defense and forensics, Security and privacy for blockchains and cryptocurrencies, Security for cloud/edge computing, etc	
	IEEE security & privacy	Application security, Attacks and defenses, Authentication, Blockchains and distributed ledger security, Cloud security, Cyber physical systems security, etc	
Patent search keywords		Application security, Authentication technique, Cloud security, Cryptographic technology, Digital forensics, Home city security, ICT security, Industrial control system security, etc	
Example of metadata of paper and patent		Title	Finding File Upload Bugs via Penetration Testing
Author		Taekjin Lee, Seongil Wi, Suyoung Lee, Sooel Son	
Abstract		An Unrestricted File Upload (UFU) vulnerability is a critical security threat that enables an adversary to upload her choice of a forged file to a target web server.	
Date		2020.11.09	
File		NDSS2020_17.pdf	

Table 5. Examples of keywords by topic from papers and patents using LDA (2020)

Paper topic	Keywords	Patent topic	Keywords
Topic 1	strategy, vulnerability, payment, crash_report, merchant, circumvention, algorithm, flash, card, attack	Topic 1	merchandise, patient, file, method, item, clinician, device, includes, base, programming
Topic 2	cache, extension, server, device, packed, defense, packing, chaperone, blacklist, compliance	Topic 2	security, data, programmable, code, switching, cloud, metadata, home, physical, vulnerability
Topic 3	access, speech, attack, service, webassembly, device, macao, keyvalue, oram, delegation	Topic 3	medical, threat, payment, data, binary, device, rolling, vehicle, healthcare, provisioning
Topic 4	unpublished_manuscript_inclusion_upcoming, prospective_author_requested_submit, event_described_paper, manager, keyless_entry, project, genetic, custos, phmon, finauth	Topic 4	token, resource, control, server, node, characteristic, packet, message, network, monitoring
Topic 5	coverage, collision, stealthy, impact, investigation, document, protection, accident, vulnerability, malware	Topic 5	provider, platform, identity, service, network, subscriber, policy, security, layer_signaling, application
Topic 6	firmware, isolation, grid, revocation, inverter, system, preview, attack, sender, email	Topic 6	droplet, related, analyte, molecule, portion, commitment, quality, warm, performance, constraint
Topic 7	skill, functional, dongle, uiscope, block, mining, ecommerce, conversation, unsolicited, privacysensitive	Topic 7	module, extracted, registered, evidence, verification, set, copy, card, sending, artifact
Topic 8	allegation, trusted, poisoning, twitter, label, path, sealer, tkperm, package, escrow	Topic 8	credential, image, pixel, administrative, standard, digital, private, encoded, region, portion
Topic 9	advice, droplet, filtering, netwarden, database, packet, besfs, seal, covert, intelligence	Topic 9	domain_name, training, score, accelerator, test, model, legitimate, train, learning, target
Topic 10	plane, censor, upload, proxy, nonce_leakage, vulnerability, fuzzing, ecdsa, enclave, trip	Topic 10	authentication, radar, principal, computing, advanced, weather, banking, login, mobile, surveillance

5.2 Data Analysis

Keywords of each topic that are extracted from papers and patents from 2010 to 2020 using the LDA algorithm are organized as shown in Table 6.

Table 6 below shows topic mapping results of the papers and patents by year in accordance with cybersecurity areas. In order to consider the topic tendency of the papers and patents as well, the topics extracted from the papers are marked in red, while the topics extracted from the patents are marked in blue.

Table 6. Mapping of cybersecurity areas with LDA

Main category	Middle category	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Network security	1. Wired network security	T3	T7, T3	T1, T10		T4		T7	T1	T1	T7	T2, T3
	2. Wireless network security	T5, T6	T1, T3	T3, T5, T3, T4, T10	T1, T8, T2, T3, T6	T4, T2, T3, T8	T3, T3, T8	T9, T4, T9	T1, T2, T8	T2, T7, T8, T10	T1, T1, T5, T8, T10	T9, T5
	3. Cloud security	T5, T8, T3, T2	T2, T6, T7, T2, T4	T2, T3, T10, T2, T4	T2, T4 T3, T6, T9	T5, T6, T7, T4, T5, T6, T7	T6, T2, T4	T7, T8 T1, T7	T2, T5 T3, T7	T3, T4, T5, T8	T6, T7	T3, T5, T2
Data and application service security	4. Application security	T1, T7, T3	T5	T1, T2, T6, T4, T8	T5, T6, T7, T8, T2, T4, T9	T5	T1, T2, T7, T4, T10	T10 T1, T7	T1	T4	T1, T4	T2, T6, T2
	5. Data security	T2, T4		T4 T3	T10	T8, T10	T8, T9 T8	T1 T7	T7	T3	T2, T3	T4, T2, T6, T7
	6. E-money, Fin-tech security	T4, T10 T1, T10		T7, T1, T2, T3, T6	T5 T2, T10	T1	T4, T10 T9	T10 T3	T3, T10 T9	T6, T7, T8, T1, T3, T4	T2, T7, T10, T1, T3, T4, T8	T2, T7
	7. Digital forensics	T3		T4		T9		T3, T5			T9	T3, T10
Physical security	8. Human/bio recognition	T2, T8	T10	T4, T7	T2	T2			T9	T9, T9, T10		T3, T9, T10
	9. CCTV surveillance/ control	T4, T7, T9				T6, T7	T7	T2				
	10. Secure search and unmanned electronic guard	T3, T5, T9	T8, T10	T1	T2, T4	T2, T9	T1, T4, T7	T10	T10	T7	T7	T4

System and password security	11. Cryptographic technique		T9	T4	T2	T3	T7 T5	T6	T2	T3, T10	T3, T5, T9, T9	T8
	12. Certification/ authorization technique	T8	T3, T5 T1, T7	T1, T7	T3, T9 T7	T4, T6 T1, T8, T9	T7, T9	T1 T2	T4, T6, T7, T9 T3	T2, T4, T9, T4, T7	T3, T6	T4, T9, T8, T10
	13. Security vulnerability	T6	T4, T6, T7, T2, T9	T8, T10	T1, T2, T4, T5, T1, T4, T7	T1, T2, T4	T5	T5, T4, T6	T9 T4	T1, T5, T2, T5	T2, T6, T8, T6	T1, T6, T4
	14. System security	T8	T1, T4, T6		T1, T2	T6	T6					
	15. Malware	T1, T9	T1, T3, T4	T2, T6, T7, T9	T4, T6 T3	T2	T6 T8	T2, T4 T1, T4, T7, T9	T2, T7, T9, T2, T7	T3, T8	T1, T8, T10	T2, T5, T1, T4
	16. Threat analysis and control	T8, T9	T2, T10	T7, T6, T10	T8, T8	T5, T7	T10, T6	T7, T9, T3	T5, T1, T5	T8, T5, T7	T7, T4, T5, T10	T10, T4
IoT security	17. Home city security	T9, T2	T8, T7	T6, T1, T9	T1, T3, T4, T6, T8	T2, T3, T5, T10	T8, T1	T9, T7	T6, T7	T2, T5	T4, T2	T2, T7
	18. Industrial control system security			T8		T3		T8	T6			
	19. Vehicle security	T1				T9, T10		T4, T5	T1, T1, T5, T8	T6	T4, T1, T5	T7, T3
	20. Ship, ocean and air security		T5			T9						T10
	21. Healthcare, Medical security	T6	T8, T6	T5, T6, T10, T7	T1, T5	T1, T2, T6	T3, T2, T3, T8	T2, T6	T1		T1, T6	T1, T3
	22. Other ICT security						T8	T3	T8		T5	T5, T9

As shown in Table 7, cybersecurity areas are largely divided into the following five main categories: network security, data and application service security, physical security, system and password security, and IoT security. The papers are mostly centered on middle categories that fall under network security, data and application service security, and system and password security, while the patents are mostly centered on physical security and IoT security. While the wired network security area has relevant topics each year, the wireless network security area has been showing more vivid technology growth than the wired network security area since 2012. In addition, it has undergone steady technological development from 2010 to 2020, along with the area of cloud security. Table 7 below shows how the LDA results are visualized in the form of a network as well as the convergence areas to identify major convergence areas by year. Moreover, connection keywords between topics were used to identify cyber convergence fields.

Table 7. Networks of major cybersecurity convergence fields by year

Year	Connection keyword	Convergence area
2016	data	application security, cloud security, malware
	device	healthcare, medical security, malware
	security	application security, cloud security, malware, security vulnerability, wireless network security
2018	data	cloud security, e-money, fin-tech security, security vulnerability, threat analysis and control
	device	certification/authorization technique, cloud security, e-money, fin-tech security, security vulnerability
	malware	cloud security, data security, e-money, fin-tech security, malware
	method	cloud security, e-money, fin-tech security
	security	e-money, fin-tech security, security vulnerability
	user	certification/authorization technique, e-money, fin-tech security
2020	attack	application security, certification/authorization technique, cloud security, data security, e-money, fin-tech security, home city security, human/bio recognition, malware, security vulnerability, wired network security
	data	application security, cloud security, data security, healthcare, medical security
	device	application security, cloud security, digital forensics, e-money, fin-tech security, healthcare, medical security, home city security, human/bio recognition, malware, wired network security
	network	Malware, security vulnerability, wireless network security
	security	application security, cloud security, data security, healthcare, medical security, malware wireless network security
	server	Malware, wired network security
	system	application security, malware, security vulnerability
	vulnerability	application security, cloud security, malware, security vulnerability

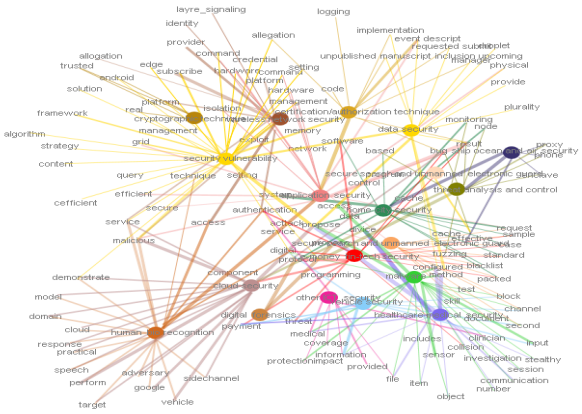


Figure 2. Changes in the average number of papers and patents for each cybersecurity field by year

Figure 2 visualizes the correlation between keywords that occurred in 2020.

In this way, convergence relationships between cybersecurity areas for each year can be analyzed based on connection keywords. However, because the annual topics extracted with LDA do not have connections by year, there is a limitation to understanding the flow of convergence areas.

5.3 Analysis of Trends in Cybersecurity Areas and Convergence Fields

The existing deduplication technique keeps the file-to-user mapping structure as metainformation. In that mechanism, the administrator with viewing rights can collect the metainformation, which makes the system extremely susceptible to insider attacks. In other words, the cloud server administrator/manager can easily access the information including the uploader lists of certain files and conversely the file lists of certain users.

The second approach applied for the purpose of analyzing changes over time in cybersecurity areas and convergence fields is the DTM algorithm to understand trends in the papers and patents for the entire period. With DTM, it is possible to identify trends in the topics as the algorithm extracts topics from each year by using an additional condition that the topics of the current year are similar to those of the previous year. Table 8 and Table 9 below show DTM keywords by topic and the results of mapping the topics according to cybersecurity areas using DTM.

Table 8. Examples of keywords by topic from papers and patents with DTM (for 2020)

Paper Topic	Keywords	Patent Topic	Keywords
Topic 1	computation, efficient, protocol, secure, cryptographic, encryption, practical, implementation, scheme, server, encrypted	Topic 1	associated, request, access, server, communication, device, network, second, client, authentication, wireless
Topic 2	adversary, channel, cache, attacker, defense, network, attack, mechanism, vulnerability, server, threat, authentication, vulnerable, system, demonstrate	Topic 2	communication, packet, system, mobile, device, network, node, includes, vehicle, computing, control
Topic 3	software, policy, control, tool, flow, memory, fuzzing, bug, implementation, vulnerability, application, program, kernel	Topic 3	memory, industrial, device, cryptographic, processor, includes, value, unit, operation, control
Topic 4	binary, detection, feature, detecting, accuracy, tool, classifier, malware, spam, model, detect, technique, domain, malicious	Topic 4	software, target, object, test, vulnerability, network, detection, malware, includes, source, sample, identifying, embodiment
Topic 5	password, website, user, payment, internet, network, account, site, content, online, service, transaction	Topic 5	medical, financial, patient, communication, encrypted, healthcare, device, document, transaction, record, secure, message
Topic 6	software, policy, author, challenge, technology, risk, issue, trust, secure, cybersecurity, development, article, community, threat, system	Topic 6	home, method, user, searching, system, smart, network, search, result, database, engine, provides, query
Topic 7	android, device, user, image, permission, mechanism, mobile, location, access, application, authentication	Topic 7	associated, threat, process, system, event, detection, second, includes, action, automation, control
Topic 8	enclave, intel, phone, attack, voice, manager, processor, hardware, trusted, system, audio	Topic 8	service, said, platform, access, customer, entity, container, environment, storage, cloud, policy, virtual
Topic 9	router, anonymity, advertisement, workshop, society, internet, network, networking, routing, edge, paper, node	Topic 9	merchandise, component, plurality, system, sensor, vulnerability, configured, includes, risk, score, item
Topic 10	storage, provenance, service, provider, process, memory, cloud, monitoring, access, cloud_ computing, system, overhead, event	Topic 10	image, display, camera, content, forensic, digital, position, portion, structure, apparatus, includes, circuit, capture

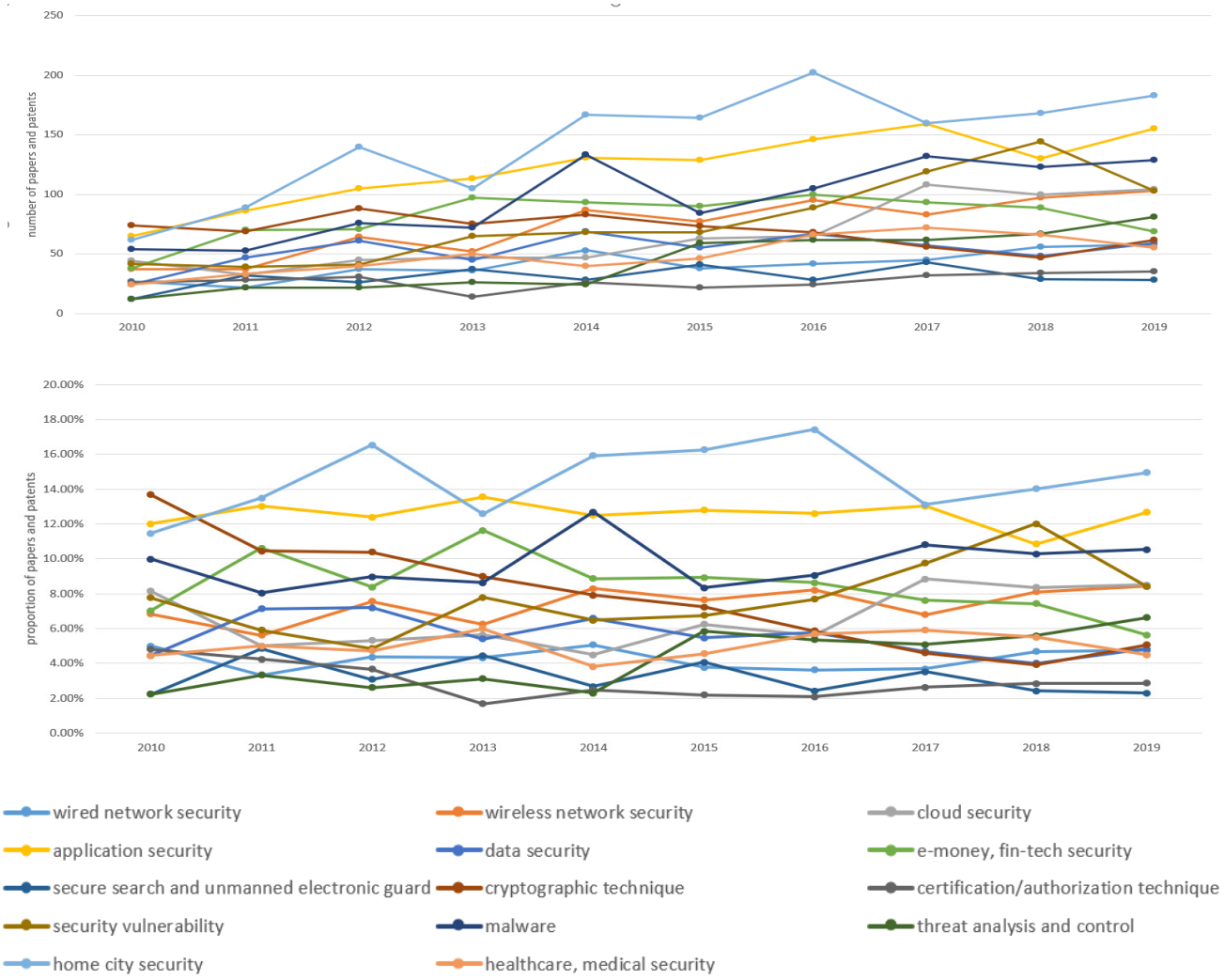


Figure 3. Cybersecurity field by year

(Changes in the average number of papers and patents for each (UP), Changes in the average proportion of papers and patents (DOWN))

As shown in Table 9, because most security areas have undergone continuous technological developments for 11 years, they are listed as DTM's topics. However, because topics from eight areas (digital forensics, human/bio recognition, CCTV surveillance/control, system security, industrial control system security, vehicle security, ship, ocean and air security, and other ICT security) appear either sporadically in certain years or intermittently at certain yearly intervals, as seen in the LDA results in Section 4.3, they were not shown in the DTM results. The cybersecurity fields with the largest number of papers and patents appearing lately are home city security (183), followed by application security (155), malware (129), cloud security (104), wireless network security (103) and security vulnerability (103). As seen here, both the academia and the business world are paying attention to diverse security fields to develop and operate cyber systems in several fields that utilize cyber systems in this era of the 4th industrial revolution. The cybersecurity fields that had consistently large numbers of papers and patents from 2010 to 2019 are home city security (144.00), application security (121.90), malware (96.10), and e-money/fin-tech security (81.00). Naturally, the average proportions displayed a similar pattern as the average numbers, with home city security taking up the biggest portion at 14.57%, followed by application security (12.54%), malware (9.73%), and e-money/fin-tech security (8.47%). Shown in Figure 3, respectively, changes in the average number and the average proportion of papers and patents for each cybersecurity field by year are useful in understanding trends in cybersecurity fields. Furthermore, the increased rates of the number and proportion of papers and patents for each cybersecurity field are effective indicators that can be used to estimate trends in cybersecurity technologies. Of these, the compound annual growth rate (CAGR) is particularly useful in monitoring trends in cybersecurity technologies. CAGR is calculated to identify the general index growth rate when the interval for index growth is one year. Therefore, it shows that the growth rates of several different cyber safety fields are noteworthy to identify trends.

Table 9. Mapping of cybersecurity areas with DTM

Main category	Middle category	2010year-2020year
Network security	1. Wired network security	T2
	2. Wireless network security	T2 T2
	3. Cloud security	T10 T8
Data and application service security	4. Application security	T9 T6
	5. Data security	T6
	6. E-money, Fin-tech security	T5 T5, T10
	7. Digital forensics	-
Physical security	8. Human/bio recognition	-
	9. CCTV surveillance/control	-
	10. Secure search and unmanned electronic guard	T9
System and password security	11. Cryptographic technique	T1
	12. Certification/authorization technique	T8 T1
	13. Security vulnerability	T3 T3
	14. System security	-
	15. Malware	T2, T4 T4
	16. Threat analysis and control	T6 T7
IoT security	17. Home city security	T7 T2, T6
	18. Industrial control system security	-
	19. Vehicle security	-
	20. Ship, ocean and air security	-
	21. Healthcare, Medical security	T5, T8
	22. Other ICT security	-

5.4 Connection Strength of Cybersecurity Convergence Areas

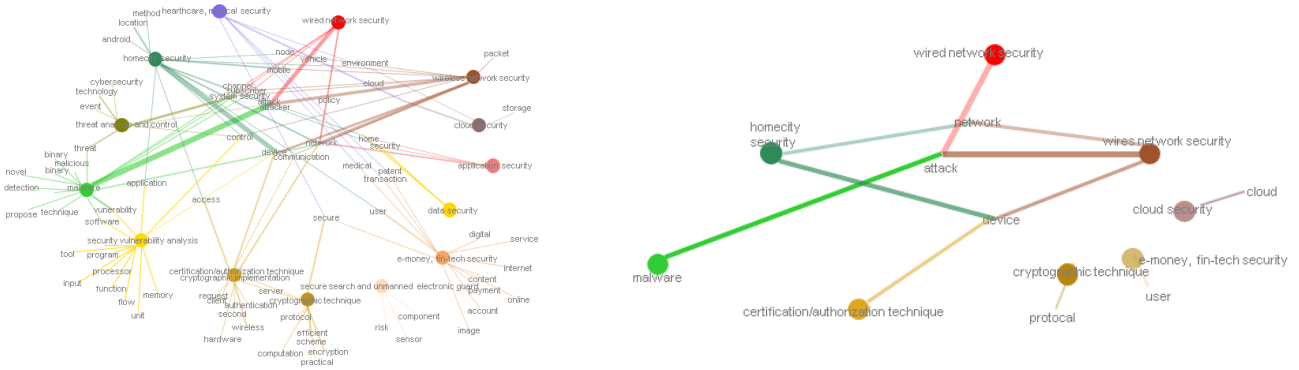
Keywords that occur concurrently among those from topics within each cybersecurity convergence field were used as connection links to extract cybersecurity convergence fields. Table 10 below shows the results of DTM network visualization, convergence areas, and connection keywords for the entire period according to the connection keyword strength.

Figure 4 shows a change according to the connection strength, and looking at (a), it is confirmed that there are many elements of connection between nodes. (b)In the case of setting a high connection strength as described above, only a connection having a high relationship may be confirmed.

The strength of connection keywords was expressed by first adding all the numbers of concurrences of the connection keywords for each year within cybersecurity areas that form the convergence areas, and then by normalizing the result as a value between 1 and 10. This means that even convergence areas with lower coherency can be expressed when the strength is low, whereas only the convergence areas with higher coherency are shown when the strength is high. It can be interpreted that more connection keywords indicate stronger coupling and fewer connection keywords indicate weaker coupling in the convergence relationship.

In addition, the prediction of highly influential research can be used as an indicator of investment, which can be a valid indicator not only from a corporate perspective but also from a national institutional perspective that requires effective development with limited capital. Therefore, it is judged that the research results of this paper can be used as investment indicators in the field of convergence security.

However, the results of the research that are predicted to have a high influence due to the results of this study may be different from those of the fusion field that will occur in the future. The prediction mechanism always has a limitation in that it cannot provide the same result as the future results, and it is required to conduct a study that increases the prediction rate by evaluating and modifying the accuracy of the fusion field proposed through continuous research.



(a) Relationship with connection strength of 1 or more (b) Relationship with connection strength of 4 or more

Figure 4. Relationship change according to connection strength

Table 10. Networks of cybersecurity areas and convergence fields for all years

Connection strength	Convergence area	Connection keyword
1	certification/authorization technique, home city security	access
	Malware, wired network security, wireless network security	Adversary, attack, attacker, channel, defense
	home city security, security vulnerability	application
	cloud security, healthcare, medical security	cloud
	certification/authorization technique, home city security, wireless network security	communication
	home city security, security vulnerability, threat analysis and control, wireless network security	Control, environment
	cryptographic technique, security vulnerability	Cryptographic, implementation
	certification/authorization technique, home city security, security vulnerability, wireless network security	device
	application security, data security, home city security	Home, search
	e-money, fin-tech security, healthcare, medical security	Medical, patient, transaction
	home city security, wireless network security	Mobile, node, vehicle
	application security, certification/authorization technique, home city security, malware, wired network security, wireless network security	network
	cloud security, healthcare, medical security, threat analysis and control	policy
	cryptographic technique, e-money, fin-tech security, healthcare, medical security	Secure
	certification/authorization technique, cryptographic technique	server
	Malware, security vulnerability	Software, vulnerability
	home city security, threat analysis and control, wired network security, wireless network security	system
	Malware, threat analysis and control	threat
	e-money, fin-tech security, home city security	user
4	Malware, wired network security, wireless network security	attack
	certification/authorization technique, home city security, wireless network security	device
	home city security, wireless network security	network

5.5. Research Analysis

The purpose of this paper is to explore studies with high influence in the future based on prominent journals in the field of security science. Therefore, journal magazines and patent information were collected, and learning was performed based on the collected information to analyze the fusion area that is expected to have high influence in the future. These research results are believed to be able to influence in various fields.

First of all, this study can be used as an academic reference indicator as one of the predictive studies of convergence technology with high influence in the future. In general, research on prediction can be evaluated based on the accuracy of future predictions that know the results through learning performance for previous years, and based on the accuracy of previously presented methodologies Therefore, research results on many predictions can be used as the basis for future new prediction mechanisms.

5.6 Comparative Analysis of Related Studies

In this study, an analysis of the convergence area that can occur in the cybersecurity field was performed, and some related studies were introduced. Since the field of convergence of technology is diverse, the analysis of related research was conducted focusing on the analysis of convergence technology without limiting it to the limited area of cyber security. As a result, similar research trends were confirmed in three fields: medical field, bioinformatics field, communication field and automobile field. Table 11 shows the comparative analysis of the research conducted and the proposed research contents.

Table 11. Comparative analysis of related studies

	Tahereh Saheb	Yuan Zhou	TaeSan Kim	Suggestion mechanism
Research purpose	Analysis of key factors in a specific area	Analysis of technology convergence process	Convergence area prediction	Convergence region analysis
Research area	Medical IoT	Bioinformatics	Telecommunications and automotive	Cyber security
Keyword auto extraction	X	O	O	O
Collection metadata area	Paper	Paper	Patent	Paper + Patent
Application technique	Qualitative Analysis, Quantitative Analysis	Newman Topology Clustering, LDA	Doc2Vec	DTM LDA

First, Tahereh Saheb's study analyzed the factors that influenced IoTBDA, collected papers in the medical IoT area, and analyzed the factors that had a great influence on the IoTBDA paradigm by performing standard analysis and quantitative analysis. Yuan Zhou collected papers in the field of bioinformatics and analyzed the changes in the fusion area over time using Newman topology clustering and LDA. TaeSan Kim's study collected patent data from communication and automobile areas, analyzed meaning using Doc2Vec, and attempted to predict fusion areas using Artificial Neural Networks, Support Vector Machine, Decision Tree, and Random Forest. In this paper, metadata of papers and patents in the cybersecurity area was collected, and convergence area analysis was performed according to time change by performing metadata-based topic classification and annual association analysis of classified topics using LDA.

6 Conclusion

This study aimed to analyze changes in development and evolution of relevant technologies from various aspects as an essential step in recognizing and understanding cybersecurity safety issues. This study's contribution comes from its monitoring of major cybersecurity convergence fields and technologies, which will provide a substantial insight into cybersecurity solution developers and experts. This is practically a first ever approach to identifying the types of cybersecurity technology convergence fields and to understanding changes in cybersecurity technology development.

Each cybersecurity technology is related to cybersecurity managers, and system operation and developers, both directly and indirectly. The highlight of this study is that understanding the changes and trends in the cybersecurity convergence fields is the most effective way to respond to cybersecurity threats and dangers in a preemptive manner and it also plays an important role in inducing new paradigms in the next generation cybersecurity technology development. Inducing a new paradigm of cybersecurity technology development can prevent the availability of computer systems from being compromised by cyberattacks.

Despite such contributions, additional studies will be needed to establish an even safer cyber system. While this study shows core cybersecurity technologies based on the keywords extracted from papers and patents, it does not mention detailed technology specifications for safety innovation. To offer more implications for cybersecurity managers and technology researchers, individual characteristics of the papers and patents, such as author information, number of citations, citation relationships, commercialization, organizations, and countries, need to be investigated. In the future, in this study, keywords are extracted through DTM and the convergence domain classified using LDA is used as an index of future convergence technology, based on the dataset from 2010 to 2019. Term Memory) Learning based convergence field prediction is performed, and trend prediction is performed in the classified convergence field by securing validity through cross validation. The prediction results of this paper may not accurately reflect future prediction trends. Such future studies will eventually provide detailed technology specifications and applicable technologies to resolve cybersecurity issues, while also presenting practical guidelines based on the analysis of papers and patents.

Acknowledgement

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5C2A04083374). The corresponding author is Namje Park.

References

[1] C. Abraham, D. Chatterjee, R. R. Sims, Muddling through cybersecurity: Insights from the U.S. healthcare industry, *Business Horizons*, Vol. 62, No. 4, pp. 539-548, July-August, 2019.

[2] M. M. Alani, M. Alloghani, Security Challenges in the Industry 4.0 Era, in: M. Dastbaz, P. Cochrane (Eds.), *Industry 4.0 and Engineering for a Sustainable Future*, Springer, Cham, 2019, pp. 117-136.

[3] I. Bongiovanni, The least secure places in the universe? A systematic literature review on information security management in higher education, *Computers & Security*, Vol. 86, pp. 350-357, September, 2019.

[4] N. Park, M. Kim, Implementation of load management application system using smart grid privacy policy in energy management service environment, *Cluster Computing*, Vol. 17, No. 3, pp. 653-664, September, 2014.

[5] H. S. Chen, G. Q. Zhang, D. H. Zhu, J. Lu, Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014, *Technological Forecasting and Social Change*, Vol. 119, pp. 39-52, June, 2017.

[6] A. Clark-Ginsberg, R. Slayton, Regulating risks within complex sociotechnical systems: Evidence from critical infrastructure cybersecurity standards, *Science and Public Policy*, Vol. 46, No. 3, pp. 339-346, June, 2019.

[7] S. Ryu, J. Kim, N. Park, Y. Seo, Preemptive Prediction-Based Automated Cyberattack Framework Modeling, *Symmetry*, Vol. 13, No. 5, Article No. 793, May, 2021.

[8] T. U. Daim, G. Rueda, H. Martin, P. Gerdsri, Forecasting emerging technologies: Use of bibliometrics and patent analysis, *Technological Forecasting and Social Change*, Vol. 73, No. 8, pp. 981-1012, October, 2006.

[9] F. Dotsika, A. Watkins, Identifying potentially disruptive trends by means of keyword network analysis, *Technological Forecasting and Social Change*, Vol. 119, pp. 114-127, June, 2017.

[10] B. Golembiewski, N. V. Stein, N. Sick, H. D. Wiemhofer, Identifying trends in battery technologies with regard to electric mobility: evidence from patenting activities along and across the battery value chain, *Journal of Cleaner Production*, Vol. 87, pp. 800-810, January, 2015.

[11] J. Jang-Jaccard, S. Nepal, A survey of emerging threats in cybersecurity, *Journal of Computer and System Sciences*, Vol. 80, No. 5, pp. 973-993, August, 2014.

[12] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, D. Siegel, Prognostics and health management design for rotary machinery systems—Reviews, methodology and application, *Mechanical Systems and Signal Processing*, Vol. 42, No. 1-2, pp. 314-334, January, 2014.

[13] D. Lee, N. Park, G. Kim, S. Jin, De-identification of metering data for smart grid personal security in intelligent CCTV-based P2P cloud computing environment, *Peer-to-Peer Networking and Applications*, Vol. 11, No. 6, pp. 1299-1308, November, 2018.

[14] S. Mendhurwar, R. Mishra, Integration of social and IoT technologies: architectural framework for digital transformation and cyber security challenges, *Enterprise Information Systems*, Vol. 15, No. 4, pp. 565-584, 2019.

[15] T. Saheb, L. Izadi, Paradigm of IoT big data analytics in the healthcare industry: A review of scientific literature and mapping of research trends, *Telematics and Informatics*, Vol. 41, pp. 70-85, August, 2019.

[16] Y. Zhou, F. Dong, D. Kong, Y. Liu, Unfolding the convergence process of scientific knowledge for the early identification of emerging technologies, *Technological Forecasting and Social Change*, Vol. 144, pp. 205-220, July, 2019.

[17] T. S. Kim, S. Y. Sohn, Machine-learning-based deep semantic analysis approach for forecasting new technology convergence, *Technological Forecasting and Social Change*, Vol. 157, pp. 1-10, August, 2020.

[18] A. Saravanan, S. S. Bama, A Review on Cyber Security and the Fifth Generation Cyberattacks, *Oriental Journal of Computer Science and Technology*, Vol. 12, No. 2, pp. 50-56, 2019.

[19] E. C. Silva, Accidents and the technology, *Journal of Loss Prevention in the Process Industries*, Vol. 49, pp. 319-325, September, 2017.

[20] L. A. Tawalbeh, F. Muheidat, M. Tawalbeh, M. Quwaider, IoT Privacy and security: Challenges and solutions, *Applied Sciences*, Vol. 10, No. 12, Article No. 4102, June, 2020.

[21] N. Park, The core competencies of SEL-based innovative creativity education, *International Journal of Pure and Applied Mathematics*, Vol. 118, No. 19, pp. 837-849, 2018.

[22] M. Vitunskaite, Y. He, T. Brandstetter, H. Janicke, Smart cities and cyber security: Are we there yet? A comparative study on the role of standards, third party risk management and security ownership, *Computers and Security*, Vol. 83, pp. 313-331, June, 2019.

[23] J. Kim, S. Ryu, N. Park, Privacy-Enhanced Data Deduplication Computational Intelligence Technique for Secure Healthcare Applications, *CMC-Computers, Materials and Continua*, Vol. 70, No. 2, pp. 4169-4184, 2022.

[24] Q. Chen, W. Wang, X. Huang, H. Liang, Attention-Based Recurrent Neural Network For Traffic Flow Prediction, *Journal of Internet Technology*, Vol. 21, No. 3, pp. 831-839, May, 2020.

[25] Y. Kim, H. K. Kim, Cluster-Based Deep One-Class Classification Model For Anomaly Detection, *Journal of Internet Technology*, Vol. 22, No. 4 , pp. 903-911, July, 2021.

Biographies

Sungwook Ryu received a Master's degree in Future Strategy from KAIST, Korea. in 2021. From 2021, he has been a doctoral course with KAIST, Korea. He is currently an researcher in future strategy. His research interests include knowledge evolution, future technology prediction and data processing and analysis etc.

Jinsu Kim graduated from Jeju National University, for the degree of Ph.D. Candidate. His research interests include privacy, network security.

Namje Park is a Professor of Department of Computer Education in Teachers College at Jeju National University since 2010. He has been serving as a Research Scientist of Arizona State University, UCLA, Electronics and Telecommunication Research Institute etc. (corresponding author: namjepark@jejunu.ac.kr).