IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 March 2024, accepted 7 March 2024, date of publication 11 March 2024, date of current version 21 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3375910

RESEARCH ARTICLE

A Scientometrics Analysis of Cybersecurity Using e-CSTI

KAZUMASA OMOTE[1,2], YOKO INOUE[1], YOSHIHIDE TERADA[3], NAOHIRO SHICHIJO[3], AND TOSHIYUKI SHIRAI[1]

[1]Department of Cabinet Office, Government of Japan, Tokyo 100-8914, Japan
[2]Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba 305-8573, Japan
[3]National Graduate Institute for Policy Studies, Tokyo 106-0032, Japan

Corresponding author: Kazumasa Omote (omote@risk.tsukuba.ac.jp)

ABSTRACT The research area of cybersecurity covers a wide range of fields from networking, software, and hardware to cryptography, authentication, and cyberattack countermeasures. Meanwhile, few cybersecurity experts are familiar with all areas of cybersecurity research. So, an evidence-based analysis covering all fields is very important to understand research trends in cybersecurity without bias. Such a field of study is called scientometrics, and there are many studies of scientometrics in the area of cybersecurity. However, in the analysis of these existing studies, the overall structure of the cybersecurity research is not clarified, and it is difficult to grasp the overall picture of cybersecurity research. In this study, we focus on cybersecurity as a research area covered by scientometrics analysis and conduct a detailed analysis of research trends in cybersecurity using e-CSTI (evidence data platform constructed by Council for Science, Technology and Innovation), and Dimensions bibliographic data for the 10 years from 2010 to 2019. Especially, we identify four representative research clusters (cyberattack, cryptography, authentication, and blockchain) and 55 research sub-clusters in the area of cybersecurity. We analyze the research trends in each country and the trends of the research topics of interest at the cluster level. For example, our results show that there are differences in research topics of emphasis between the U.S. and China. e-CSTI assists policymakers and researchers in getting a comprehensive understanding of global research trends and topics of interest in cybersecurity research.

INDEX TERMS Cybersecurity, scientometrics, research cluster, literature map, e-CSTI.

I. INTRODUCTION

A. EVIDENCE-BASED POLICY MAKING (EBPM)

A method that relies on collecting the personal views of a small number of experts in a field as a means of making policy regarding an important scientific or technological field may present a biased view. This is where the concept of EBPM is important. EBPM refers to evidence-based policymaking, where policy planning is not based on ad hoc episodes, but is evidence-based with clearly defined policy objectives [22]. The promotion of EBPM, which utilizes information and statistics that have important relevance to the measurement of policy effects, enhances the effectiveness of policies and contributes to ensuring public trust in government.

EBPM is promoted in many countries around the world. In the U.K., the ''Modernizing Government'' published in 1999 indicated a policy of promoting evidence-based policymaking with a long-term perspective, rather than responding only to short-term demands [31]. In the U.S., the Commission on Evidence-based Policymaking (CEP) was established in 2016 to conduct research for efficient policymaking through evidence.[1] In Japan, the ''Statistical Reform Promotion Council'' announced its final recommendation in May 2017 indicating the basic policy for promoting EBPM. In conjunction with this, the ''EBPM Promotion Committee'' was launched in August 2017 [13]. In particular, the Cabinet Office of Japan has been making various efforts

The associate editor coordinating the review of this manuscript and approving it for publication was S. K. Hafizul Islam.

[1]Evidence-Based Policymaking Commission Act of 2016, Pub. L. No. 114-140, 130 Stat. 317 (Mar. 30, 2016).

TABLE 1. Cybersecurity threat rankings by the organization (2020).

	Kaseya (U.S.)	ENSIA (EU)	IPA (Japan)
1	Phishing attacks	Malware	Confidential information theft by APT
2	Remote worker endpoint security	Web-based attacks	Information leakage by Internal fraudulent acts
3	Cloud jacking	Phishing	Financial loss by business e-mail compromise
4	IoT devices	Web application attacks	Attacks exploiting supply chain weaknesses
5	Sophisticated and targeted ransomware attacks	Spam	Financial loss by ransomware
6	Deep fakes	DDoS	Suspension of business due to unexpected IT infrastructure failure
7	Mobile malware	Identity theft	Careless information leakage
8	5G-to-WiFi security vulnerabilities	Data breach	Personal information theft from services on the Internet
9	Insider threats	Insider threat	Unauthorized use of IoT devices
10	API vulnerabilities and breaches	Botnets	Business service outage caused by DoS Attacks

to promote EBPM and has constructed e-CSTI (evidence data platform constructed by Council for Science, Technology and Innovation),[2] which is a platform for analyzing data related to science, technology, and innovation. e-CSTI is one of the concrete tools of scientometrics.

B. SCIENTOMETRICS

Scientometrics is a field of study concerned with the measurement and analysis of academic literature [16]. It helps to understand various aspects of the growth of a particular topic or field and allows quantitative evaluation of research trends as well as research institutions and researchers. The study of scientometrics is based largely on the work of Garfield [7], who was the first to propose the quantification of citations and is highly regarded as a foundational achievement in scientometrics. It is a bibliometric method for the effective study of various scientific fields and aims to provide researchers and policymakers with metrics and visualization methods to measure the latest research trends and aspects of the scientific and technological literature. Quantitative data on science and technology and its analysis are an indispensable basis for the formulation of science and technology policies, and also play an important role in understanding the status of a wide range of science and technology activities and in analyzing the effects and impacts of such policies.

The "Scientific and Technology Indicators" are used to measure and analyze the progress of science and technology using scientometrics methods [21], [32]. They are indispensable for evaluating the progress of science and technology in a country, region, industry, or research field. Data on research expenditures, the number of researchers, the number of publications and citations, and the number of patent applications are used as specific science and technology indicators. These indicators play an important role in helping policymakers, researchers, and companies evaluate scientific and technological progress and determine future directions.

C. CYBERSECURITY AND THREAT ANALYSIS

Cybersecurity is a fundamental research area of science and technology that provides security and safety for systems and services. The results of threat analyses in cybersecurity are published in many countries around the world. Table 1 shows the ranking of cybersecurity threats by research organizations in the U.S., Europe, and Japan. Kaseya is a U.S. company that provides cloud-based software platforms and security monitoring and publishes information on the Top 10 cybersecurity threats [11]. ENISA is a Greek company that shares the latest technology and information on cybersecurity in Europe, evaluates cybersecurity measures across the EU, and publishes information on the top 15 cybersecurity threats [6]. IPA is a Japanese independent administrative agency that promotes cybersecurity, provides information on cybersecurity measures, collects and analyzes information on threats and attacks, and publishes information on the Top 10 cybersecurity threats [9]. Such threat information enables each organization to understand the cybersecurity threats that have an impact on their business.

However, the results in Table 1 are based on information collected by each organization from their perspective, and there is a possibility of bias in the information and data sources used in the analysis. For example, ransomware is ranked 5th by Kaseya, unranked (13th) by ENSIA, and 3rd by IPA, showing a gap in ranking. In addition, phishing attacks are ranked first by Kaseya and third by ENSIA, but not by IPA. Therefore, such information is not suitable for a global and comprehensive analysis of the cybersecurity area and does not capture the overall structure of the cybersecurity area.

D. RELATED WORK

There are many studies on scientometrics, among which Lee [14] was the first to conduct a study on scientometrics in the area of cybersecurity. Lee used the SCI (Science Citation Index) database to extract important keywords and used Co-Word analysis to clarify trends and patterns in the cybersecurity area. For example, this study mentions the rapid transformation of cybersecurity topics. Olijnyk [23] used Scopus bibliographic data to analyze the profiles, dynamics, and structure of the cybersecurity area to clarify the intellectual structure of the cybersecurity area. For example, this study identifies institutions and authors that are highly influential in cybersecurity. Rai et al. [24] conducted a trend analysis of 2,720 research literature on cybersecurity using Scopus bibliographic data. This study includes an analysis of

[2]https://e-csti.go.jp/en

TABLE 2. Comparison of existing studies and our analysis.

Study	Data set	Period	Literature	Extraction way	Details
Lee [14]	SCI	1980–2003	all	text-base	co-word analysis, domain map
Olijnyk [23]	Scopus	1965–2015	all	text-base	co-word analysis, term map, notable literature, sources of publication, country-wise contributions, author productivity, institution productivity
Rai et al [24]	Scopus	2001–2018	all	text-base	collaboration index, author productivity, citations, country-wise contributions, sources of publication, institution productivity, funding agencies
Dhawan et al [5]	Scopus	1998–2019	all	text-base	citations, funding agencies, country-wise contributions, international collaboration, institution productivity, author productivity, sources of publication
Loan et al [17]	Web of Science	2011–2020	all	text-base	collaboration, country-wise contributions, common keyword, author productivity
Our analysis	Dimensions	2010–2019	top 10%	co-citation	cluster / sub-cluster map, country-wise contributions, word cloud, top conference literature

countries, institutions, collaborations, and research funding. Dhawan et al. [5] used 22 years of Scopus bibliographic data from 1998 to 2019 to analyze trends and developments in countries, institutions, and authors. For example, they list institution productivity, author productivity, and sources of publication. Loan et al. [17] analyzed the cybersecurity literature over 10 years (2011-2020) using Web of Science bibliographic data. Specifically, they search for literature using terms such as ''cybersecurity'', ''cyber-security'', ''web security'', ''information security'', and ''computer security'', and analyze related keywords, countries, and authors. Table 2 summarizes the existing studies on the scientometrics of cybersecurity. All of these previous analyses are based on keywords to extract security literature, various analyses are performed, and various indicators are computed.

However, the analysis of these existing studies does not clarify the overall structure of the cybersecurity research area. Specifically, there is no analysis of how many research clusters there are in the cybersecurity research area and what kind of research topics these clusters are composed of. Although existing studies often analyzed cybersecurity research by keywords, to the best of our knowledge, there were no security cluster maps generated using co-citation information.[3] Cluster maps generated by keywords may have the risk of picking up less relevant literature. There is also a possible risk of not being able to distinguish between studies if keywords are inadequate in the title and abstract. This is because it simply picks up keywords used in titles and abstracts, ignoring the relationship between two pieces of literature. Our analysis uses co-citation information, which reduces the risk of including irrelevant literature.

Several other analyses focused on specific areas of cybersecurity have been conducted. Makawana and Jhaveri [20] analyzed 149 research literature from January 2015 to December 2016 to analyze research trends specifically in the area of machine learning for cybersecurity, while Abbas et al. [1], in addition to analyzing trends and activities in countries, institutions, and authors, also provides a visual analysis of artificial intelligence (AI) applications using heat maps. Xu et al. [33] analyzed the number of literature by country and keywords, specializing in malware research, based on bibliographic data obtained from the China National Knowledge Internet, while Kappi et al. [10] analyzed the trends in the number of literature and research fields based on Scopus bibliographic data, with a special focus on the blockchain area, using information such as country, period, and author.

Note that, although there are many research survey papers on cybersecurity [4], [15], [18], [28], [29], these are different from scientometrics and are analyses based on the authors' subjectivity in their field of expertise.

E. CONTRIBUTION

In this study, we focus on cybersecurity as a research area handled by scientometrics analysis and conduct a detailed analysis of research trends in the cybersecurity area using e-CSTI and Dimensions bibliographic data for the 10 years from 2010 to 2019. Our study contributes to the global body of knowledge on cybersecurity by providing a holistic view of the development of cybersecurity area. The following is a list of specific contributions.

- e-CSTI automatically selects four representative research clusters (cyberattacks, cryptography, authentication, and blockchain) in the area of cybersecurity based on co-citation information, and identifies the structure of each research cluster. Each cluster is composed of several sub-clusters. This cluster/sub-cluster-based analysis helps policymakers and researchers gain a comprehensive understanding of global research trends and topics of interest in cybersecurity research.
- In all four cybersecurity-related clusters, we found that the U.S. was initially the leader in the number of Top 10% literature and was later overtaken by China. This implies that the U.S. is quickly initiating fundamental research in new areas. Furthermore, by focusing on the blockchain field, which is an emerging area in cybersecurity, and analyzing trends in the U.S. and China, it became clear that the U.S. was the first to start, followed by China overtaking the U.S.

[3]When references A and B are included in the bibliography of a single reference, A and B are in a co-citation relationship.

- Using the Top 10% literature in the analysis provides a new angle on the literature that is highly influential in the field and addresses important issues. For example, our results imply that the Top 10% literature may capture the decrease signs in the cryptography topic earlier.

F. ORGANIZATION

The remainder of the paper is structured as follows: Section II describes in detail e-CSTI, a research field analysis tool developed by the Cabinet Office of Japan, Section III describes in detail the dataset, methodology, and results of the analysis in the area of cybersecurity using e-CSTI, Section IV provides a discussion of the analysis results, and finally, Section V concludes the paper.

II. E-CSTI

The Cabinet Office of Japan has developed various analytical functions to collect evidence on the research, educational, and fundraising capabilities of universities and other research institutions and to visualize the relationship between inputs and outputs. Based on these functions, the Cabinet Office has constructed an evidence data platform constructed by Council for Science, Technology, and Innovation (e-CSTI), a platform for sharing analysis functions and data to relevant parties at the governments, national universities, and research and development agencies. Japan's ''Sixth Science, Technology, and Innovation Basic Plan'' (cabinet decision on March 26th, 2021) and ''Integrated Innovation Strategy 2022'' (cabinet decision on June 3rd, 2022) indicate that e-CSTI will be used to identify and analyze important science and technology fields and to revise sector-specific strategies and develop new national strategies.

e-CSTI assists policymakers and researchers in getting a comprehensive understanding of global research trends and topics of interest in cybersecurity research. This tool can visualize bibliographic information of all research fields by arranging them into one relevant research cluster based on co-citation relationships, and users, including policymakers, can use it without knowledge of SQL or Python. Furthermore, by analyzing the research clusters containing the literature and technologies of interest, we can obtain the number of literature related to the technology, the share of literature in each country, the degree of fusion of research fields, the degree of citation to patents, international research networks, and notable researchers and their budget execution data. In this study, we present the results of our analysis in the area of cybersecurity, focusing on results that can be made publicly available. Naturally, the same analysis can be performed for other research areas that have attracted attention in recent years, such as AI, quantum computers, and biotechnology.

III. ANALYSIS OF THE CYBERSECURITY AREA USING E-CSTI

A. DATA

Bibliographic data was collected for the Top 10% literature in the 10-year Dimensions bibliographic data (2010-2019) provided by Digital Science.[4] The Dimensions database is said to have better coverage of articles than other databases such as Scopus and Web of Science [25]. The number of included literature is 2,224,645 (with 3,201,598 authors). The following six types of literature are included.[5]

- Article: Article from a scientific journal or trade magazine, including news and editorial content
- Book: Edited book or volume comprised of chapters usually written by different authors and harmonized by one or more editors
- Chapter: Individual part of an (edited) book, including individual entries in an encyclopedia
- Monograph: Book on a single subject or an aspect of a subject, often by a single author
- Preprint: Non-peer-reviewed version of a scholarly or scientific paper
- Proceeding: Individual paper published in conference proceedings, including editorial content

The research literature is classified into 1,076 clusters based on the co-citation relationship among literature and is further classified into 12,445 sub-clusters. The data include the name of the publisher, type of literature, literature title, abstract, reference, author, institutional affiliation, country of affiliation, and research field. Research fields extracted as research clusters/sub-clusters mean that the number of literature is large and structured, and the areas to be clustered are of high importance. In addition, we use the ANZSRC (Australian and New Zealand Standard Research Classification) 2008 FoR (Fields of Research) as the research fields. The ANZSRC 2008 FoR can be classified at three levels: 22 divisions, 157 groups, and 1340 fields. In the literature map, it is possible to color-code at two levels: divisions and groups.

In this analysis, we focus on the Top 10% literature data, which is mainly used in the analysis of scientific and technological indices and represents important discoveries and advances in the field. The top 10% literature is generally considered to be highly influential research because such literature is cited more frequently by other literature [3]. Therefore, by analyzing the characteristics and trends of the Top 10% literature, it is expected to identify important research topics and directions in a certain research field. However, literature that is published for a short period of time often does not have a sufficient number of citations, making it difficult to select accurate citation relationships and the Top 10% literature. Therefore, in this study, we target literature that has been published for more than two years and employ the Top 10% literature from 2010 to 2019, measured as of November 2022.

To identify representative research clusters in the area of cybersecurity, the keywords such as ''security'', ''attack'',

[4]Data sourced from Dimensions, an inter-linked research information system provided by Digital Science (https://www.dimensions.ai).

[5]https://plus.dimensions.ai/support/solutions/articles/23000018866-which-publication-types-are-available-in-dimensions-

"cryptography", "encryption", "authentication", and "privacy" were used for the titles and abstracts of the literature. As a result, we identified the main four research clusters (cyberattack, cryptography, authentication, and blockchain) that are closely related to cybersecurity. The cyberattack research cluster contains 4,252 kinds of literature, the cryptography research cluster contains 2,908 kinds of literature, the authentication research cluster contains 1,412 kinds of literature, and the blockchain research cluster contains 2,449 kinds of literature, and a total of 11,021 kinds of literature are included in the analysis. Other research clusters related to cybersecurity include hardware security, privacy, and biometric authentication.

Self-citations, though can be used genuinely to credit someone's work, can play a significant role in the artificial manipulation of scientific impact. Amjad et al. [2] analyzed the impact of self-citations and showed that self-citations, if removed from total received citations, negatively influence the author ranking metrics. Thus, excluding self-citations could bias the results of the analysis. It is also very difficult to distinguish good self-citations from bad self-citations. We refer to this study and decide not to remove the self-citation literature.

B. ANALYSIS METHOD

There are three ways to express the relationship between two pieces of literature: direct citation, bibliographic linkage [12], and co-citation [26]. A direct citation considers two pieces of literature in a citation relationship to be related. A bibliographic linkage considers two pieces of literature that cite the same literature to be related. A co-citation considers two pieces of literature that are cited by the same literature to be related. The co-citation is based on the assumption that literature with related content is cited together. In this section, we conduct 12 different analyses using co-citation for the cybersecurity area. The analysis method is based on the method of Small et al. [27] for analyzing emerging fields, where literature is clustered using co-citation information. Specifically, after creating a network of literature based on co-citation relations, the analysis tool converts the combination and frequency of co-cited literature into a vector for each literature using the Node2Vec algorithm [8] and then converts them into a two-dimensional map using the t-SNE algorithm [19]. Furthermore, research clusters are generated by clustering the bibliographic data using the Leiden algorithm [30]. The number of research sub-clusters in each research cluster is automatically determined by this algorithm. The software "Tableau" is used to visually display bibliographic data.

The main analysis is as follows.

1) Literature map: Research clusters are grouped by co-citation information and each piece of literature is mapped to a single circle. In the literature map, each piece of literature can be visualized by color-coding by research sub-cluster, research field, literature type, and country of institutional affiliation. The size of the circle in the literature map is determined by the number of citations.
2) Trends in the number of literature: It visualizes the 10-year trends in the number of literature by research cluster, research sub-cluster, and country of institutional affiliation.
3) Ranking by the number of literature: It is displayed in order of the number of references to the country of the international conference and the institution to which the author belongs.
4) Word cloud: To see changes in research trends, keywords (including multiple words) appearing in titles and abstracts of literature are extracted and visualized.
5) The degree of fusion and spread: In order to see the interdisciplinary nature of the literature, we quantitatively evaluate the research clusters in terms of two aspects: fusion and spread.

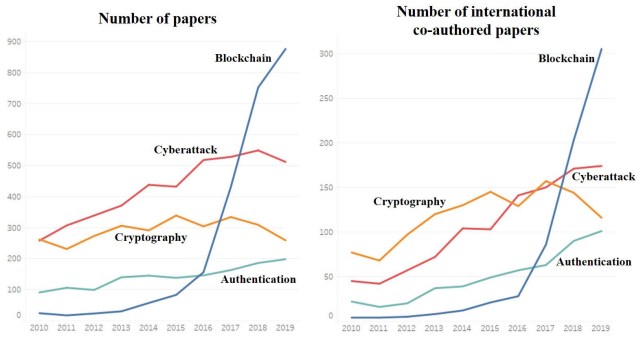

FIGURE 1. Number of literature and international co-authored literature in four cybersecurity clusters for Top 10% literature (2010-2019).

C. ANALYSIS RESULTS

1) FOUR MAJOR RESEARCH CLUSTERS RELATED TO CYBERSECURITY

Four research clusters (cyberattack, cryptography, authentication, and blockchain) were identified as representative research clusters in cybersecurity. The first two clusters, cyberattack and cryptography, are major research areas in cybersecurity, so their emergence is quite appropriate. Authentication technology is also important in cybersecurity, which can be divided into cryptography and applied authentication technology. Since cryptography such as digital signatures is included in the cryptography cluster, the authentication research cluster is considered to be composed of applied authentication technologies. Furthermore, the blockchain cluster is emerging as a security area. This cluster is considered to be a new cluster that has emerged due to the rapid expansion of blockchain research in recent years.

2) NUMBER OF LITERATURE AND INTERNATIONAL CO-AUTHORED LITERATURE

Figure 1 shows the number of literature and international co-authorship in the four research clusters related to

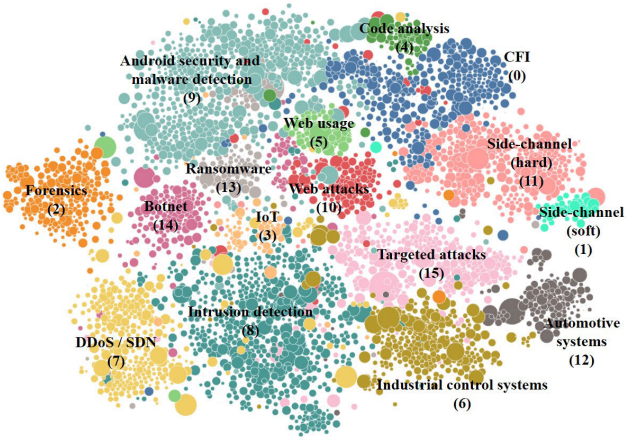

FIGURE 2. Literature map (Top 10%) of cluster on cyberattack by 16 sub-clusters.

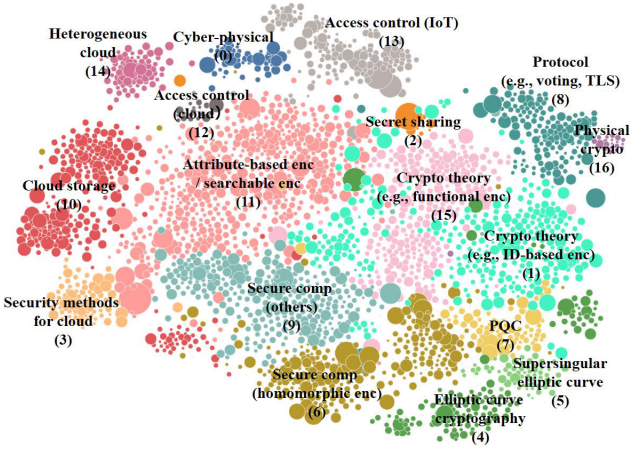

FIGURE 3. Literature map (Top 10%) of cluster on cryptography by 17 sub-clusters.

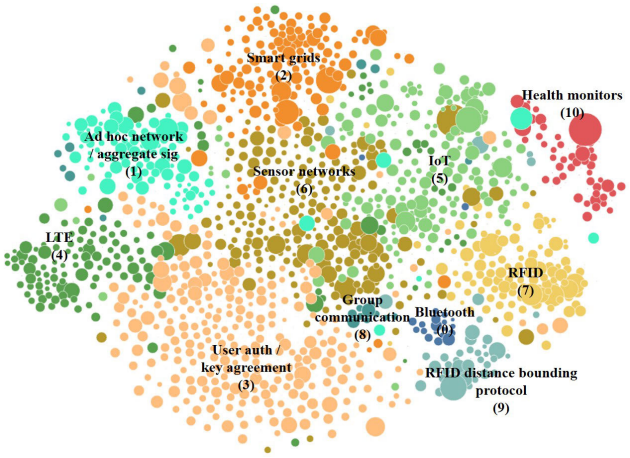

FIGURE 4. Literature map (Top 10%) of cluster on authentication by 11 sub-clusters.

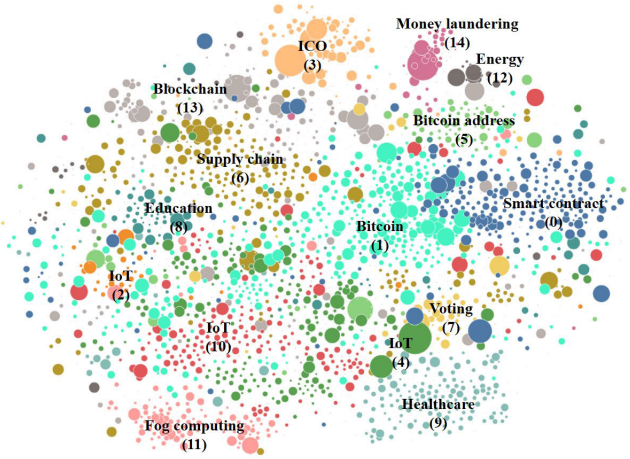

FIGURE 5. Literature map (Top 10%) of cluster on blockchain by 15 sub-clusters.

cybersecurity over 10 years (2010-2019). All four research clusters show an increasing trend in both the number of literature and the number of international co-authorships. The cyberattack research cluster has become more important in recent years, with not only an increasing number of literature but also an increasing number of international co-authored literature. The cryptography research cluster continues to have a certain number of literature, and the authentication research cluster has an increasing number of literature. Blockchain is a new technology that has shown potential and promise for application in a variety of fields, and the blockchain research cluster has seen a rapid increase in both the number of literature and international co-authorship since around 2016.

3) LITERATURE MAP BY RESEARCH SUB-CLUSTER

Figure 2-5 show a literature map of the four research clusters related to cybersecurity for the 11,021 literature, color-coded by research sub-cluster.[6] Because points that are mapped in a space of several hundred dimensions are forcibly reduced to a two-dimensional plane, clusters and literature in close proximity can be highly related. Since the research sub-clusters are grouped by co-citation information, the research sub-clusters include not only the studies of the labeled keywords but also their related studies and related technologies. An overview of the literature maps shows that the three clusters of cyberattack, cryptography, and authentication have more clearly differentiated sub-clusters, whereas the blockchain cluster has blurred sub-cluster boundaries due to the fact that blockchain is a new field of study. The labels of each sub-cluster are assigned by the authors based on the analysis using the word cloud and the content check of individual literature. The top, bottom, left and right of the literature map are meaningless, and the sub-cluster numbers are also meaningless.

The cyberattack research cluster consists of 16 sub-clusters, as shown in Figure 2, characterized by attack targets, methods, and countermeasures. The Top 10% literature map

[6]Research sub-cluster numbers mean nothing more than numbering.

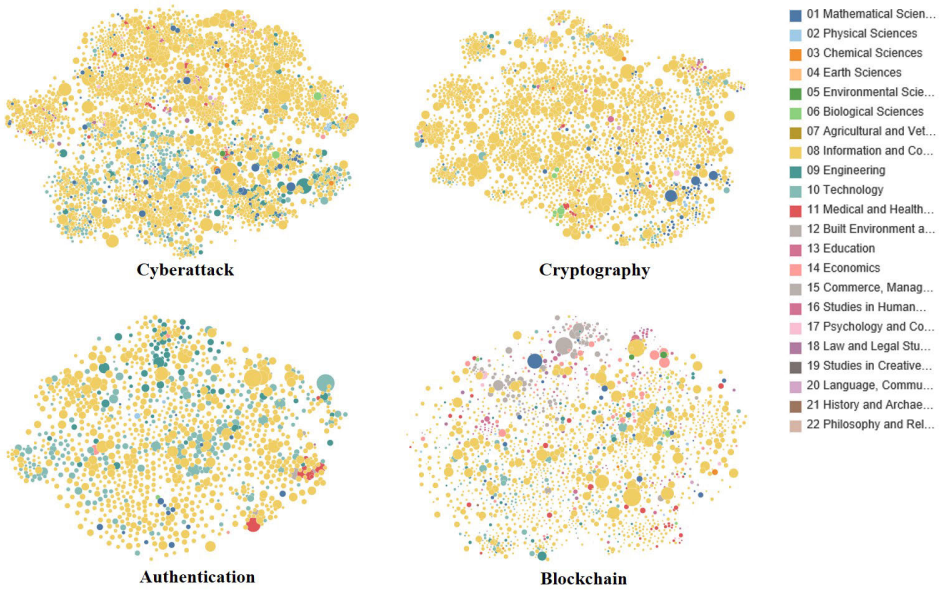

FIGURE 6. Research field distribution of ANZSRC 2008 FoR divisions in the literature map.

of cyberattacks shows the emergence of several convergence research sub-clusters, such as (12) Automotive systems, (11) Side-channel (hard), and (3) IoT, capturing a wide range of related fields. Among these, the increase in the number of literature is particularly remarkable in (8) Intrusion detection.

The cryptography research cluster consists of 17 sub-clusters as shown in Figure 3, which are mainly characterized by method and subject. In the literature map, there is an overlap between cloud-related research on the left and theory-centered research on the right, and (11) Attribute-based encryption / searchable encryption is located in this overlap. This research is important for fine-grained access control and service diversity in the cloud, and there is a large amount of both theoretical and applied research in this area. Among these, the increase in the number of literature was particularly notable for (1) and (15) Cryptography theory, which are the theoretical foundation of cryptographic techniques, (6) and (9) Secure computation, which are a fundamental technology that enables data analysis while maintaining the confidentiality of data, and (11) Attribute-based encryption / searchable encryption.

The authentication research cluster consists of 11 sub-clusters as shown in Figure 4, which are characterized by methods and targets. In this research cluster, we see a lot of applied research such as IoT (Internet of Things), smart grids, and health monitors. In particular, the number of literature on (5) IoT and (6) sensor networks related to IoT increases significantly.

The blockchain research cluster consists of 15 sub-clusters as shown in Figure 5, which are characterized by the application areas of blockchain and related technologies. Research is conducted for applications in various areas such as finance, supply chain, healthcare, energy, and education. The number of literature related to a blockchain is increasing in all research sub-clusters, and especially (1) Bitcoin, (6) Supply chain, and (10) IoT are rapidly increasing. Note that there were not many differences among the three IoT sub-clusters appearing in this research cluster. Reviewing the four research clusters related to cybersecurity, the keyword "IoT" is included in all four clusters, indicating that IoT is a very important keyword in the area of cybersecurity. In addition, "cyber-physical" for the cryptography cluster, "sensor networks" for the authentication cluster, and "fog computing" for the blockchain cluster appear as other keywords related to IoT.

4) VISUALIZATION OF LITERATURE AREAS

Figure 6 shows the four literature maps color-coded by the 22 different divisions in the ANZSRC 2008 FoR. Cybersecurity is the main research area in "08 Information and Computing Sciences", followed by "10 Technology". In the following paragraphs, we will describe the other areas except for 08 and 10. In the cyberattack research cluster, the literature of "01 Mathematical Sciences" and "11 Medical and Health Sciences" is sparsely placed. In particular, "11 Medical and Health Sciences" is frequently found in the sub-clusters (9) Android security and malware detection, (13) Ransomware, and (15) Targeted attacks, indicating the importance of countermeasures against cyberattacks on medical information. In the cryptography research cluster, "01 Mathematical Sciences" stands out on the right side of the map. This may be due to the fact that algebra is often used in cryptography theory. In particular, "01 Mathematical Sciences" stands out in the sub-clusters (4) Elliptic curve

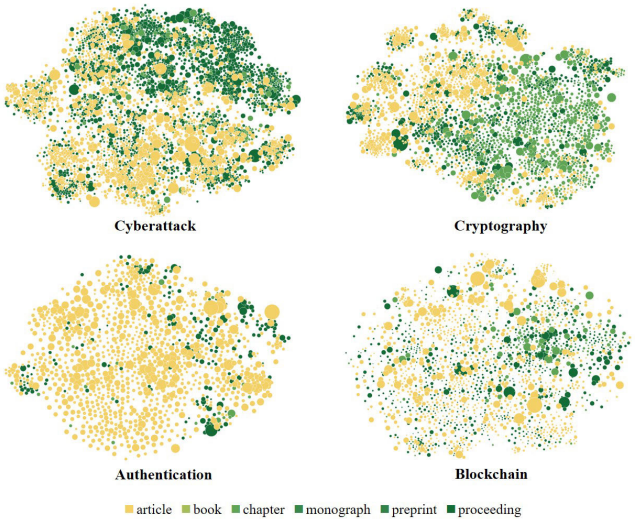

FIGURE 7. Literature type distribution in the literature map.

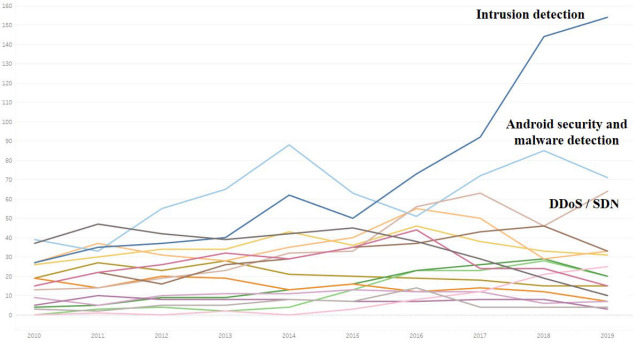

FIGURE 8. Number of literature per sub-cluster in a cyberattack.

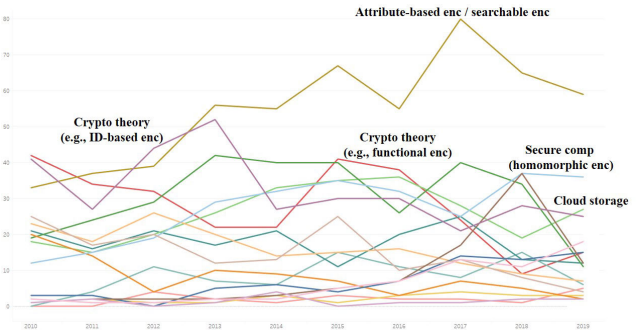

FIGURE 9. Number of literature per sub-cluster in cryptography.

cryptography, (5) Supersingular elliptic curve, and (7) Post-quantum cryptography (PQC). ''06 Biological Sciences'' is relatively common in (6) Secure computation (homomorphic encryption) and (9) Secure computation (others), where research such as genome analysis is conducted. In the authentication research cluster, ''06 Biological Sciences'' is often found in the sub-cluster (7) RFID, indicating that many eHealth-related researchers use RFID. In the blockchain research cluster, ''15 Commerce, Management,

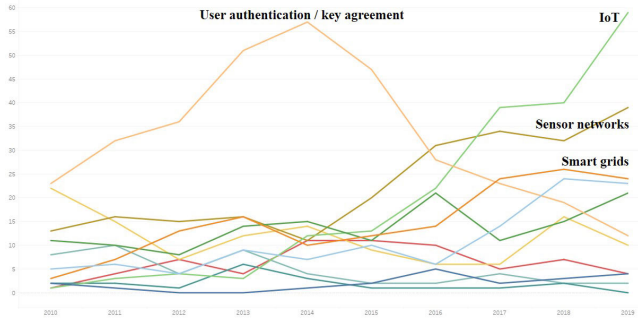

FIGURE 10. Number of literature per sub-cluster in authentication.

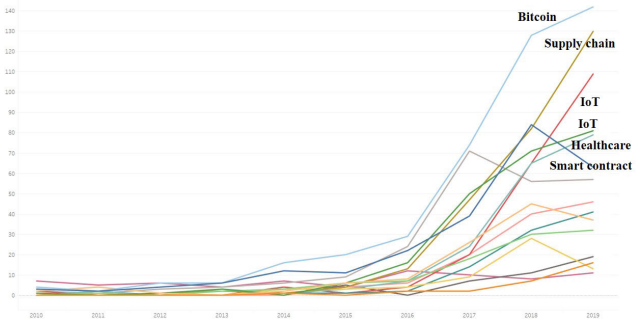

FIGURE 11. Number of literature per sub-cluster in the blockchain.

Tourism, and Services'' is found in (3) ICO, ''16 Studies in Human Society'' is found in (14) money laundering, ''14 Economics'' is found in (12) Energy, and ''13 Education'' is found in (8) Education, which indicate that this research cluster includes a variety of research areas.

5) VISUALIZATION OF LITERATURE TYPE

Depending on the research field, not only articles are important, but also chapters and proceedings of international conference papers are important. Some international conferences are more difficult to accept than articles. For example, four international conferences (IEEE S&P, ACM CCS, CRYPTO, and EUROCRYPT) are the top security conferences and have a reputation comparable to articles. Preprints are also important literature. Preprints are published as soon as they are submitted because there is no peer review period, and the latest research tends to be published earlier. The four literature maps in Figure 7 show that the cybersecurity research area has a relatively large number of chapters and proceedings in the literature. For example, in cryptography, the right half of the figure shows a large percentage of chapters and proceedings in the area of theoretical research.

6) NUMBER OF LITERATURE IN RESEARCH SUB-CLUSTERS

Figures 8-11 show the 10-year trend of the number of literature in the sub-clusters for each research cluster in the area of cybersecurity. In the cyberattack research cluster shown in Figure 8, intrusion detection shows the strongest

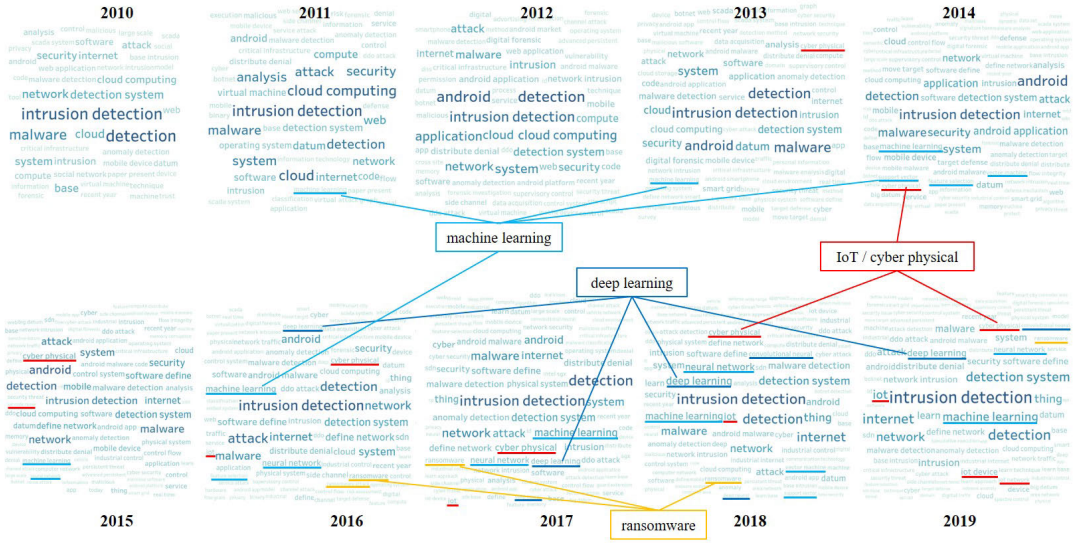

FIGURE 12. Time-series analysis of research trends by changing keywords (cyberattack).

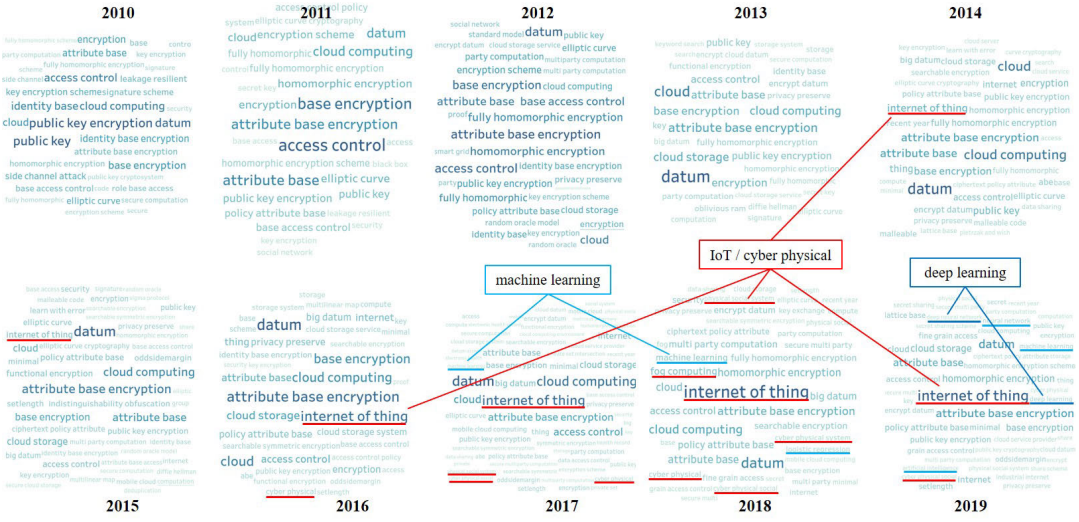

FIGURE 13. Time-series analysis of research trends by changing keywords (cryptography).

increase, followed by "Android security and malware detection" and DDoS / SDN. "Android security and malware detection" peaked in 2014, and the number of literature has been maintained in recent years. The reason why DDoS and SDN are in the same sub-cluster is that the literature on countermeasures against DDoS attacks on SDN systems has been increasing in recent years. With the recent remarkable development of networks such as 5G, it can be inferred that research on cyberattacks is more and more focused on network-centric research topics such as intrusion detection, Android, malware detection, DDoS, and SDN.

In the cryptography research cluster shown in Figure 9, the largest number of studies is on attribute-based encryption / searchable encryption to ensure fine-grained security in cloud systems. Secure computation is also on the rise, indicating its growing importance in cryptography research. Meanwhile, both sub-clusters of the two cryptography theories are decreasing. From the above, it can be inferred that there has been a shift in cryptography research from theory to two areas: cloud applications and secret computation. Both of these two areas are related to trends such as increasing data volumes, the emergence of cloud computing, and the use of AI.

In the authentication research cluster shown in Figure 10, there is a very large amount of literature related to IoT and sensor networks. Meanwhile, "User authentication / key agreement" peaked in 2014 and has decreased since then. In recent years, we have found that devices that require

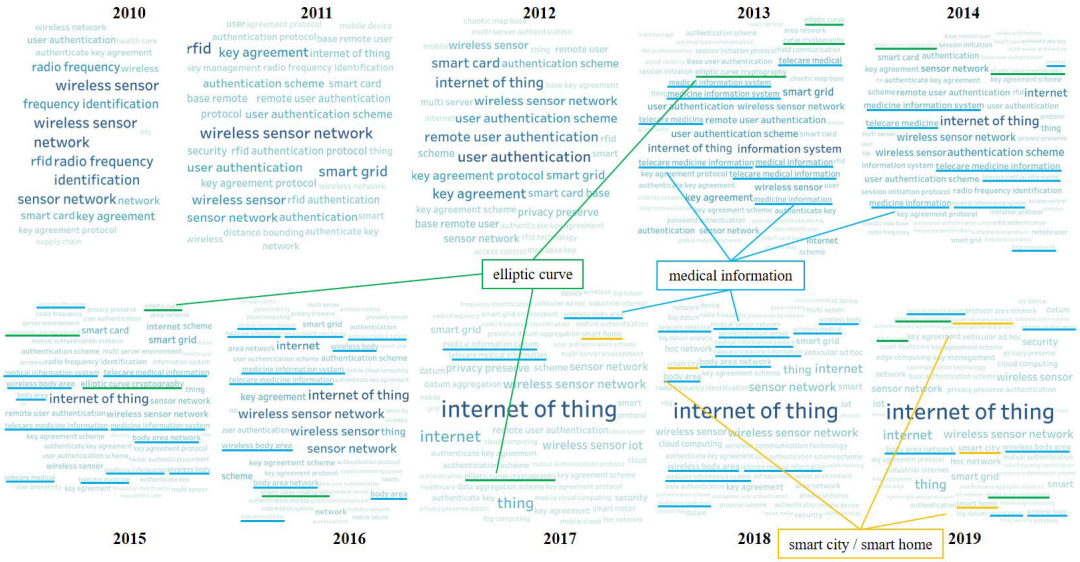

FIGURE 14. Time-series analysis of research trends by changing keywords (authentication).

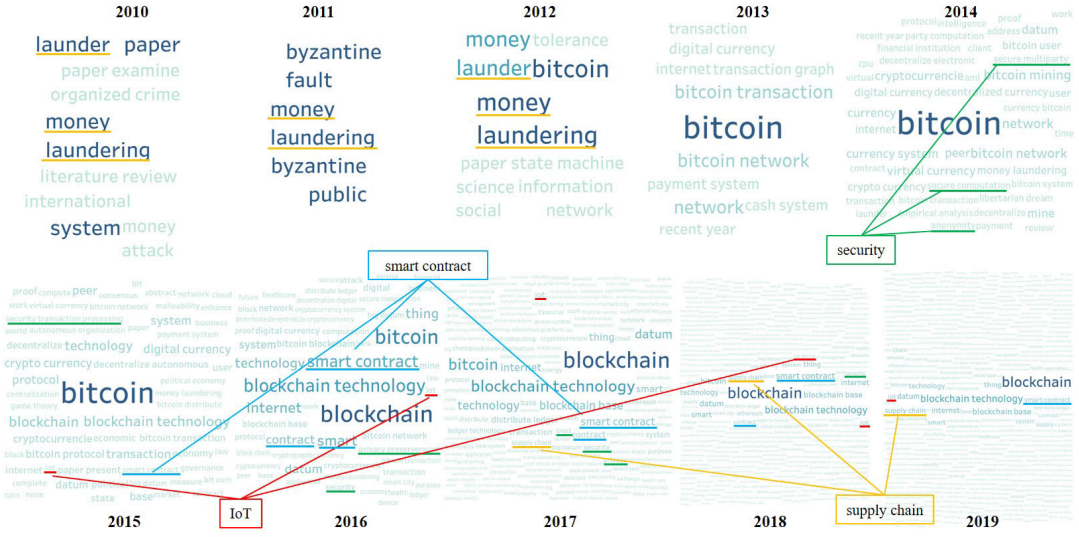

FIGURE 15. Time-series analysis of research trends by changing keywords (blockchain).

authentication are not computationally rich, but IoT devices and sensors that have few computational resources. From the above, it can be inferred that research on authentication has shifted to applied research on IoT-based authentication. We believe that research is shifting from user-based authentication to device-based one.

In the blockchain research cluster shown in Figure 11, "Bitcoin" has the largest number of literature, followed by "Supply chain", "IoT", and "Healthcare". The literature on "Smart contracts" is also on the rise. "Supply chain", "IoT", and "Healthcare" are all related to smart contracts. The invention of smart contracts around 2013 expanded the scope of their application, and research in these areas is expected to have increased rapidly around 2016.

7) TIME-SERIES ANALYSIS OF RESEARCH TRENDS BY CHANGING KEYWORDS

A word cloud is a visualization method that changes the font size according to the frequency of occurrence of words in a sentence or text. Figures 12-15 show a time-series analysis of research trends by the change of keywords in each research cluster using the word cloud. Each word cloud displays the frequently appearing keywords for each year.

In the cyberattack research cluster depicted in Figure 12, the keywords "intrusion detection", "malware", and "android" appear throughout. Machine learning-related topics have appeared frequently since around 2011, and deep learning-related topics have increased since around 2016, making them highly relevant to AI. Thus, we can

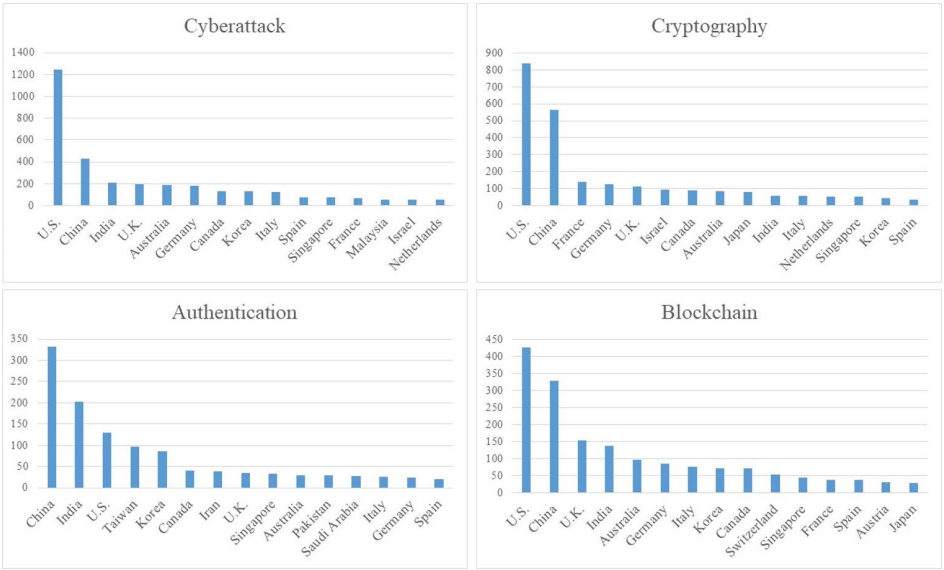

FIGURE 16. Number of literature by country in four literature clusters related to cybersecurity.

see that research on AI-based cyberattack countermeasures is highly active. In addition, IoT and cyber-physical-related keywords have emerged since around 2013. Since the IoT and cyber-physical keywords are increasing in the cryptography research cluster (see Figure 13) and the authentication research cluster (see Figure 14) at the same time (around 2013), we believe that IoT and cyber-physical keywords are similarly increasing in the cyberattack research cluster. Thus, we can see that research related to IoT cyberattacks started even before the emergence (around 2016) of the disruptive malware ''Mirai'', which targets IoT devices. Also, since ransomware-related keywords have appeared since around 2016, we believe this is related to CryptoLocker, a new type of ransomware that appeared around 2013 and shook the world.

In the cryptography research cluster shown in Figure 13, cryptographic keywords such as attribute-based encryption, homomorphic encryption, and multi-party computation, as well as application keywords such as access control and cloud computing appear throughout. The invention of attribute-based cryptography was in 2005 and homomorphic encryption was in 2009, and we can see that these two studies are growing due to these influences. Since attribute-based cryptography is closely related to access control and cloud computing, those keywords are also expected to increase. Keywords related to IoT and cyber-physical have emerged since around 2014, and keywords related to machine and deep learning since around 2017. This also explains the general trend that big data is gathered through the use of IoT and cloud computing, the need for cryptography arises from the handling of private information, and secure computation, which performs AI computation while protecting privacy, has been attracting attention. Thus, we can see that there is a lot of research on not only achieving better confidentiality of cloud data but also securing the use of data.

In the authentication research cluster shown in Figure 14, the keywords IoT and sensor network appear throughout. This indicates that authentication in IoT and sensors is very important. Keywords related to medical information and elliptic curves have been increasing since around 2013, and keywords related to smart cities and smart homes have been increasing since around 2017. Medical information, smart cities, and smart homes are areas in which authentication is important and indispensable, and the application of authentication technology is considered to be actively conducted. Meanwhile, elliptic curves are used in the key-agreement protocol, and since the key size of elliptic curve cryptography is shorter than those of general cryptographic techniques, elliptic curve-based authentication is attracting attention for its application to devices with poor performance, such as IoT and sensors. Therefore, we can see the importance of authentication techniques not only in IoT-based systems/applications but also in critical systems such as medical information and smart cities.

In the blockchain research cluster shown in Figure 15, the keywords change rapidly throughout. Around 2010-2012, there were many keywords related to the money laundering of crypto-assets, and around 2012-2017, there were many keywords related to bitcoin. Security-related keywords increased around 2014, and smart contract and IoT keywords increased around 2015. Furthermore, keywords related to the supply chain increased around 2017. These indicate a shift from research on crypto-assets and money laundering in Blockchain 1.0 to research on smart contracts and their applications in Blockchain 2.0. In particular, the emergence of Ethereum smart contracts in 2014 has expanded the scope of application and led to a rapid increase in applied research. For example, smart contracts are an important technology for IoT and supply chains, and these keywords are increasing.

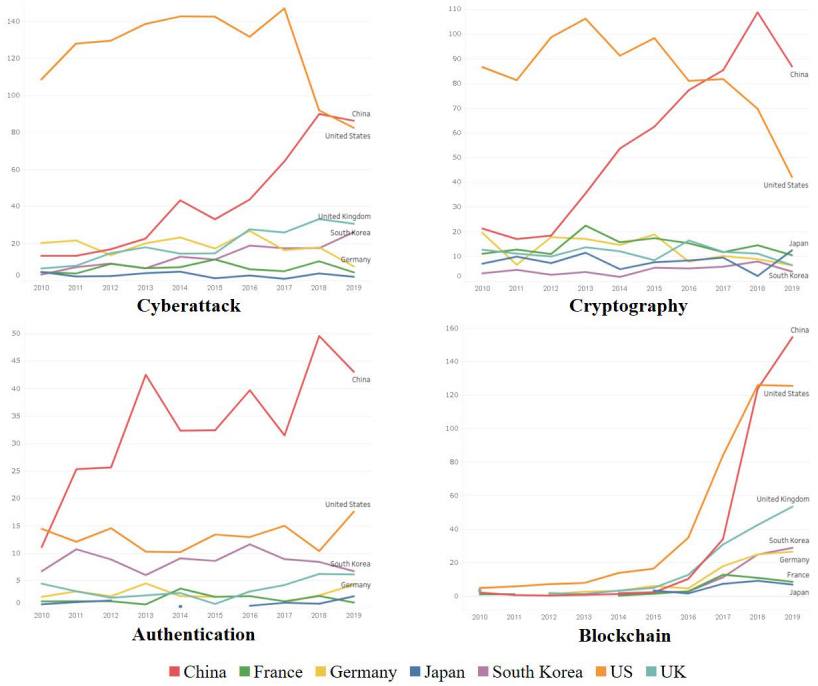

FIGURE 17. Number of literature in each country (seven selected countries) in the research cluster on cybersecurity.

Furthermore, the literature on IoT has increased rapidly not only in the blockchain research cluster but also in the authentication research cluster, indicating that IoT is growing in secure application research.

8) NUMBER OF LITERATURE BY COUNTRY IN EACH RESEARCH CLUSTER

From here, we will analyze the country. Figure 16 shows the top 15 ranking of the number of literature by country for the four research clusters related to cybersecurity. The numbers on the vertical axis show the literature counted in fractional counts by country. First, as an overall trend, the number of literature from the U.S. and China is high, and the number of literature from the U.S. is particularly prominent for the cyberattack cluster. Other than the U.S. and China, the countries ranked in all four cybersecurity-related clusters are, in descending order of the number of literature, India, the UK, Germany, Australia, Korea, Canada, Italy, Singapore, and Spain. As for the other countries, Malaysia (cyberattack), Japan (cryptography), Taiwan (authentication), and Switzerland (blockchain) have a relatively high number of Top 10% literature, indicating that these countries have strong research capabilities in their respective areas.

9) NUMBER OF LITERATURE IN SEVEN SELECTED COUNTRIES IN THE RESEARCH CLUSTERS

We analyze seven selected countries (China, France, Germany, Japan, South Korea, the U.S., and the U.K.). Figure 17 shows the trend in the number of literature in each country in the four research clusters on cybersecurity. In all four cybersecurity-related clusters, the U.S. was initially the leader in the number of Top 10% literature and was later overtaken by China. This implies that the U.S. is quickly initiating fundamental research in new areas. In the three clusters of cyber attacks, cryptography, and authentication, the timing of the rise in the number of literature has already passed, while in the blockchain cluster, the timing for the number of literature to start up exactly appeared in 2010. Focusing on the blockchain cluster, we see that the U.S. has led in the Top 10% literature since 2010 but was overtaken by China in 2019.

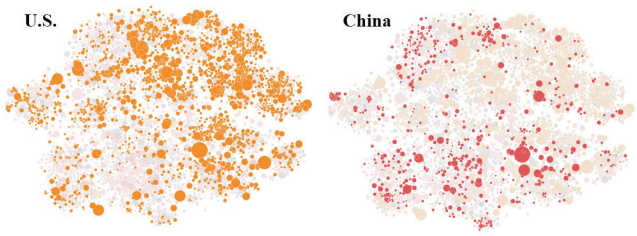

FIGURE 18. Literature map of cyberattack research clusters in the U.S. and China.

In the cyberattack research cluster, the literature from the U.S. has decreased since around 2017, while the literature from China has increased, and the difference between them has almost disappeared. It can be inferred that China is the main contributor to the increase in the number of literature in the cyberattack cluster in Figure 1. In addition, the U.K. and South Korea have steadily increased their literature. In the

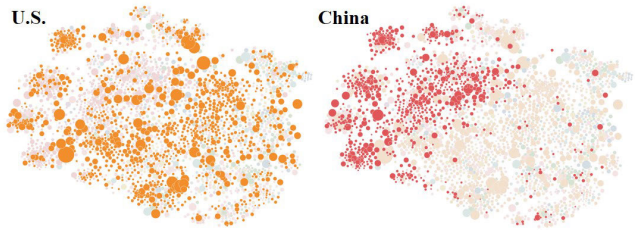

FIGURE 19. Literature map of cryptography research clusters in the U.S. and China.

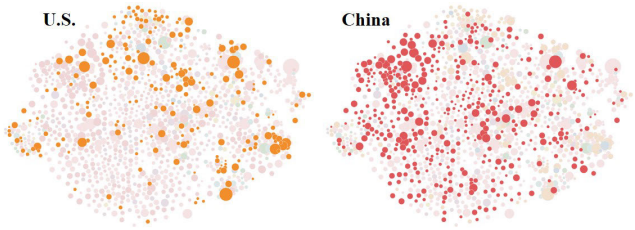

FIGURE 20. Literature map of authentication research clusters in the U.S. and China.

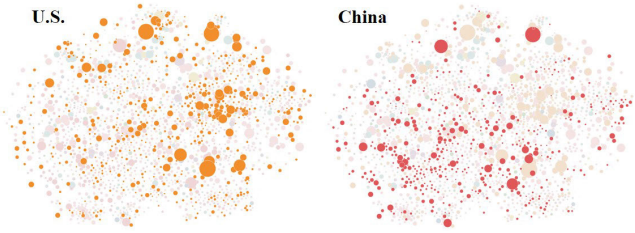

FIGURE 21. Literature map of blockchain research clusters in the U.S. and China.

cryptography research cluster, while the overall number of literature has decreased since around 2013, the number of literature from China has increased significantly. It can be inferred that the number of literature between China and the other countries cancels each other out, resulting in the flattening of the graph in Figure 1. An important point of view from this is that the number of cryptography literature is decreasing in the world except for China. Looking only at the total number of literature does not reveal global trends. In the authentication research cluster, China has the largest number of literature by far. In the blockchain research cluster, the number of literature has increased rapidly in many countries since around 2015, and both the U.S. and China are outstanding. Many countries consider blockchain to be important, especially the governments of the U.S., China, and the U.K., which have declared national blockchain strategies.

10) TRENDS IN THE U.S. AND CHINA

The results so far show that the U.S. and China are strong in the area of cybersecurity. In Figures 18-21, only U.S. and Chinese literature is colored in the literature map for comparison. Thus, we explore the trend analysis of the U.S. and China to understand their respective strengths and other trends in the literature maps of the four research clusters. We see that the U.S. is strong in cyberattack and cryptography, China is strong in authentication, and both the U.S. and China are strong in blockchain. In cyberattacks, the U.S. is strong in the side channel (hardware) on the right side of the map, while China is strong in the network-related area on the lower left side of the map. From these facts, we believe that the U.S. tends to regard hardware attacks as a threat, while China tends to regard network attacks as a threat. In cryptography, the U.S. is strong overall, including theoretical research, while China is strong in the cloud-related technologies on the left side of the map. In authentication, we can see that China is focusing on IoT networking, including ad hoc networks, aggregated signatures, IoT, and sensor networks. In the blockchain, we can see that the U.S. is focusing on Bitcoin in the center of the map and ICOs in the upper part of the map, while China is focusing on IoT and fog computing in the lower left part of the map.

As an overall trend, the U.S. was found to be strong in all four areas, with particular emphasis on CPU security and theoretical research. China, in contrast, was found to be more focused on applied technologies, including IoT and cloud computing.

11) RELATIONSHIP WITH AI

AI is used in various research fields of science and technology. Figure 22 shows a literature map of the four research clusters on cybersecurity, coloring the literature in the ANZSRC 2008 FoR group ‘‘0801 Artificial Intelligence and Image Processing’’. The results show that AI is relevant in all four research clusters, and cyberattack in particular is closely related to AI. The AI-related literature in the cyberattack research cluster includes (6) Industrial control systems, (8) Intrusion detection, (9) Android security and malware detection, (12) Automotive systems, (13) Ransomware, and (14) Botnet. Meanwhile, in the cryptography research cluster, AI-related literature was particularly found in the (6) Secure computation (homomorphic encryption). In this sub-cluster, research is focused on realizing machine learning algorithms using secure computation such as homomorphic encryption. In the authentication research cluster and blockchain research cluster, AI-related research is widely distributed.

12) DEGREE OF FUSION AND SPREAD

Research that involves findings from a variety of fields is likely to generate value, and research that is cited in the literature of various fields is also considered to be highly valuable. Figure 23 shows the results of calculating the degree of fusion and spread of the four research clusters in cybersecurity concerning the literature. The degree of fusion is defined by the Herfindahl-Hirschman Index (HHI) of each research field (groups of FOR) in all the kinds of literature cited by a given literature. The higher the value, the more influences from various areas of the literature have. The degree of spread is also defined by the HHI for each research field in all the kinds of literature cited by a given literature. The higher the value, the greater the spread across

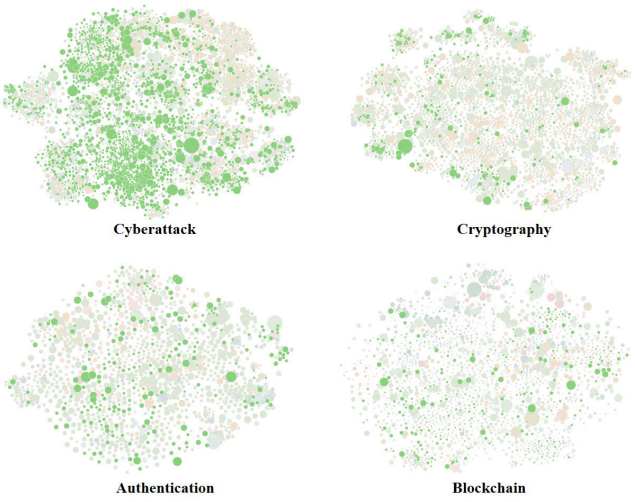

FIGURE 22. Literature map related to AI in major four cybersecurity clusters.

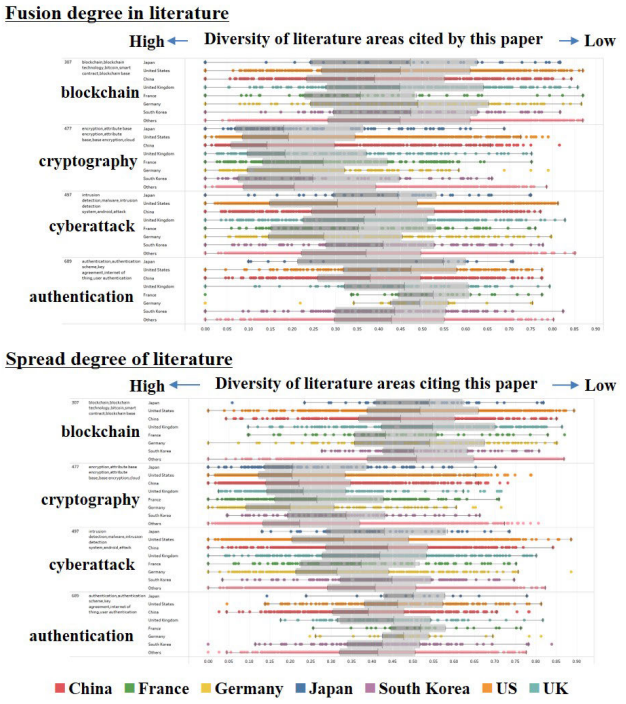

FIGURE 23. Degree of fusion and spread in the cybersecurity literature.

diverse research areas. The results show that among the four research clusters, the blockchain research cluster has both a high degree of fusion and spread, while the cryptography research cluster has both a low degree of fusion and spread. In particular, the blockchain research cluster, as a growing area, is utilizing knowledge and influencing literature from a variety of areas.

In general, theoretical research tends to be limited to a specific area and is unlikely to have an impact on different areas. In addition, theoretical research often assumes basic knowledge and skills in the area and is often difficult for experts in other areas to understand. This is why the degree of fusion and spread of cryptography research clusters is likely to be low. Meanwhile, applied research tends to have a greater impact on different areas because it focuses on practical problems. Applied research tends to impact different areas because it is often conducted to solve practical problems in society and industry, and often requires cooperation and knowledge sharing with experts in different research areas. Therefore, blockchain and authentication research clusters are expected to have a high degree of fusion and spread, as many applied types of research are found in such literature maps shown in Figure 6.

IV. DISCUSSION

A. RECENT TRENDS IN THE AREA OF CYBERSECURITY

Our analysis targets the Top 10% literature, which makes it difficult to analyze recent literature because a certain period of time is needed for the number of co-citations and citations in the literature to accumulate. Thus, we perform an additional analysis of the most recent ''all literature'' to check for the latest trends at the keyword level for the area of cybersecurity. Specifically, we use ''SciVal'', an analysis tool in Scopus, another bibliographic resource by Elsevier, to identify broad trends for several topics in cybersecurity for the Scopus dataset (2013-2022). Figure 24 shows the time series of the number of literature related to three keywords of ''Blockchain'', ''Cryptography'', and ''Authentication'' in the Scopus topic cluster TC.84 (Cryptography; Authentication; Data Privacy). Note that We have not included cyberattacks because TC.84 contains little literature related to cyberattacks. The latest data as of February 14, 2024, includes data from 2013 to 2022, displaying results that are three years newer than our 10% literature analysis. Figure 24 shows roughly similar trends on the left side of Figure 1 from 2013 to 2019. Recent trends show that the number of literature in the blockchain area is rising, while the number of literature in cryptography and authentication is slightly increasing. In particular, the analysis results in the area of cryptography show that the number of literature using SciVal is on an increasing trend, while the number of Top 10% literature using e-CSTI has already reached a plateau. This means that the Top 10% literature may capture the decrease signs in the cryptography area earlier. In fact, as shown in 9) of ''III-C Analysis Results'', the number of literature in the area of cryptography has reached a plateau worldwide, except for China.

B. TRENDS IN BLOCKCHAIN AREAS IN DIFFERENT COUNTRIES

Figures 1 and 24 show that blockchain research is growing rapidly in the cybersecurity area, and we can observe the emergence of this area in our results. Therefore, we check which countries have increased the literature for the blockchain area. Figure 25 displays the number of literature for the seven selected countries in the blockchain subgroup

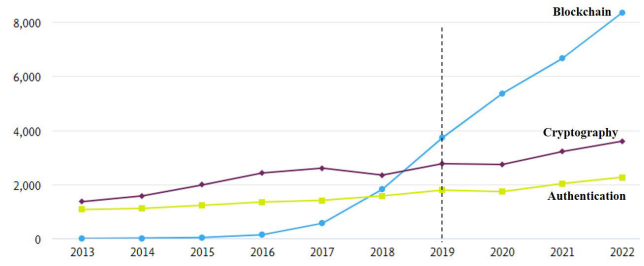

FIGURE 24. Recent trends in the area of cybersecurity using SciVal tool in Scopus.

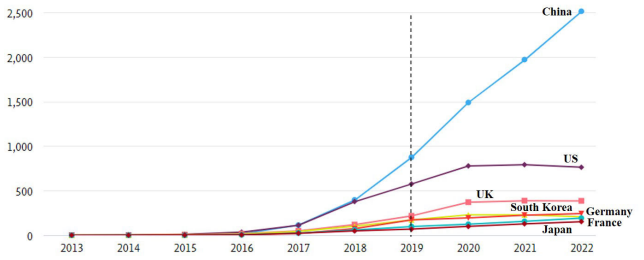

FIGURE 25. Recent trends in the area of blockchain using SciVal tool in Scopus.

within TC.84 of SciVal tool. The results show that the bottom right of Figure 17 and Figure 25 from 2013 to 2019 are roughly similar. For example, both rankings in 2019 are China, the U.S., the U.K., South Korea, Germany, France, and Japan, all in descending order of number of literature.[7] Recent trends in the total literature related to blockchain show that the gap between China and the other six countries is growing, and the U.S. has been overtaken by China in the process. More precisely, the U.S. was overtaken by China in 2017 in SciVal's analysis of all literature, while the U.S. was overtaken by China in 2019 in our analysis of the Top 10% literature. As a result, we can see that the U.S. initially increased the number of Top 10% of literature or all literature, while China started to increase the number of such literature in the middle of the period. This shows that the U.S. was leading the blockchain area. Thus, it may be important to check the trends in the U.S. for emerging research areas. Furthermore, these latest trends have not changed significantly since 2020, confirming that the trend is continuing.

C. OTHER RESEARCH CLUSTERS ON CYBERSECURITY

In the area of cybersecurity research, we have analyzed four major research clusters (cyberattack, cryptography, authentication, and blockchain). In addition to these research clusters, there are several other clusters related to cybersecurity, such as hardware security, privacy, and biometrics authentication. These three clusters are not major research clusters because of their small number of literature compared to the four major research clusters, but they include important cybersecurity technologies. In this section, we discuss each of these clusters.

1) HARDWARE SECURITY

The research sub-clusters with the highest number of literature in the Top 10% in this cluster are hardware trojans, Physical Unclonable Function (PUF), side-channel attacks, and lightweight block ciphers. The overall number of literature in this research cluster has been decreasing in recent years. Many public-key cryptography belongs to the cryptography research cluster, while symmetric-key cryptography, such as AES, belongs to the hardware security research cluster. Since symmetric-key cryptography is capable of efficient computation, the focus is likely to be on hardware implementation and its security. In the analysis of the seven selected countries, the U.S. is the strongest country for research on hardware trojans and PUFs, France and Germany are the strongest countries for research on side-channel attacks, and France and China are the strongest countries for research on lightweight block ciphers. Also, Research on side-channel attacks and lightweight block ciphers further produces the Top 10% literature worldwide.

2) PRIVACY

The research sub-clusters with the highest number of literature in the Top 10% in this cluster are differential privacy, location privacy, and social networking service (SNS) privacy. Differential privacy is a privacy-preserving technique that makes it impossible to determine whether a particular individual is included in aggregated personal information. Regarding user privacy, it is important to protect location privacy and SNS privacy, in which personal information is collected. With the widespread use of smartphones, a large amount of data containing personal information is being collected, and privacy protection of such data has become a target of research. Furthermore, the research sub-cluster of federated learning has seen a rapid increase in the number of literature since around 2018. Federated learning is a method for learning models in a distributed environment without sharing data. However, user privacy protection is also important in federated learning because various user data are used for learning. Of the seven selected countries, the U.S. is the strongest country for research on differential privacy and SNS privacy, while the U.S. and China are the strongest countries for research on location privacy and federated learning.

3) BIOMETRICS AUTHENTICATION

The research sub-clusters with the highest number of literature in the Top 10% in this cluster are face recognition, vein authentication, iris recognition, and gait recognition. Face recognition has by far the largest number of literature. The number of literature on gait recognition has been increasing in recent years and has become the second-largest sub-cluster in 2019. In addition, vein authentication and iris recognition are important technologies that have been

[7]To be precise, Figure 25 had the same number of literature for South Korea and the U.K. in 2019.

TABLE 3. Structure of research clusters (Top 10% literature) in the area of cybersecurity extracted by e-CSTI.

Research cluster	Research sub-cluster
Cyberattack	(0) Control flow integrity, (1) Side-channel (soft), (2) Forensics, (3) IoT, (4) Code analysis, (5) Web usage, (6) Industrial control systems, (7) DDoS / SDN, (8) Intrusion detection, (9) Android security and malware detection, (10) Web attacks, (11) Side-channel (hard), (12) Automotive systems, (13) Ransomware, (14) Botnet, (15) Targeted attacks
Cryptography	(0) Cyber-physical, (1) Cryptography theory (e.g., ID-based cryptography), (2) Secret sharing, (3) Security methods for cloud, (4) Elliptic curve cryptography, (5) Supersingular elliptic curve, (6) Secure computation (homomorphic encryption), (7) Post-quantum cryptography, (8) Protocol (e.g., voting, TLS), (9) Secure computation (others), (10) Cloud storage, (11) Attribute-based encryption / searchable encryption, (12) Access control (cloud), (13) Access control (IoT), (14) Heterogeneous cloud, (15) Cryptography theory (e.g., functional encryption), (16) Physical cryptography
Authentication	(0) Bluetooth, (1) Ad hoc network / aggregate signature, (2) Smart grids, (3) User authentication / key agreement, (4) LTE, (5) IoT, (6) Sensor networks, (7) RFID, (8) Group communication, (9) RFID distance bounding protocol, (10) Health monitors
Blockchain	(0) Smart contract, (1) Bitcoin, (2) IoT, (3) Initial coin offering, (4) IoT, (5) Bitcoin address, (6) Supply chain, (7) Voting, (8) Education, (9) Healthcare, (10) IoT, (11) Fog computing, (12) Energy, (13) Blockchain, (14) Money laundering

put into practical use in various areas. In addition, face recognition, iris recognition, and gait recognition are superior from the viewpoint of infection control because they can be used for non-contact authentication. Face recognition has been attracting attention due to the widespread use of camera-equipped cell phones, and the gait recognition system is useful for identifying criminals using security cameras. Of the seven selected countries, China and the U.S. are the strongest countries for research on face and iris recognition, and China has the strongest country for research on vein authentication and gait recognition. In addition, research on facial and iris recognition further produces the Top 10% literature worldwide.

D. LIMITATIONS OF OUR ANALYSIS

As mentioned in section III-A, a certain period of time is needed for the number of co-citations and citations in the literature to accumulate in order to make the analysis more accurate using the co-citation information and Top 10% literature. If the period of time after publication is short, there are few opportunities for the literature to be cited, and sufficient citation relationships and citation counts cannot be obtained. Therefore, our analysis tool is not suited for real-time analysis and has the limitation that the results of the current analysis cannot be immediately obtained. However, our analysis allows us to analyze the Top 10% literature in terms of their quality, rather than a real-time analysis focused on the quantity of literature. We believe that there is value in capturing such long-term trends while there will be some delay.

Since this study uses a co-citation relationship, the research cluster of a given literature changes depending on how the references are formed. In other words, two pieces of literature with similar research contents can be classified into different research clusters if they are not cited in the same literature. Furthermore, references that cover multiple research fields may not form a large cluster because they are divided into multiple clusters. As a result, some important areas cannot be identified at the research cluster level. For example, we focus on IoT security, which is important in the area of cybersecurity. Although no research cluster has been formed for IoT security, it is an important research area in cybersecurity because IoT appears in all four major research clusters in cybersecurity. However, its importance is not visible at the research cluster level.

V. CONCLUSION

In this study, we focus on cybersecurity as a research area that is handled by scientometrics analysis and analyze four representative research clusters (cyberattack, cryptography, authentication, and blockchain) and 55 research sub-clusters in the cybersecurity area using e-CSTI. The analysis results revealed the following specific findings.

- The research area of cybersecurity has a structure of 4 research clusters and 55 research sub-clusters as a result of the analysis of the Top 10% literature in the Dimensions bibliography for the period 2010-2019.
- In the cybersecurity area, the number of literature from the U.S. and China tends to be high overall. The other countries in order of the number of literature are India, the U.K., Germany, Australia, South Korea, Canada, Italy, Singapore, and Spain. In addition, the U.S. has a significant lead in the number of literature among the four top security conferences. In other countries, Germany, China, and the UK are the strongest in security in general, and Israel, France, China, and Germany are the strongest in cryptography.
- The U.S. and China focus on different research areas. For example, in the cyberattack research cluster, the U.S. tends to regard hardware attacks as a threat, while China tends to regard network attacks as a threat.
- Research on the blockchain is generating a major cluster within the cybersecurity area. In this context, we see a shift from Blockchain 1.0 research on crypto-assets and money laundering to Blockchain 2.0 research on smart contracts and their applications.
- IoT is a very important keyword in the cybersecurity area and appears in all four research clusters.
- The cybersecurity research area is closely related to AI. In particular, the cyberattack research cluster is extremely closely related to AI.

A future task would be to conduct a similar analysis using the entire literature by a method with less time lag, and

compare the results with the present results using the Top 10% literature.

APPENDIX

Table 3 organizes the four research clusters and 55 research sub-clusters of Top 10% literature related to cybersecurity into a table.

ACKNOWLEDGMENT

The authors thank Prof. Takahiro Ueyama, a full-time Executive Member of the CSTI of the Cabinet Office for initiating and promoting of the overall project, invaluable advice and discussion for this study. They also thank Dr. Hiroaki Nagai (Cabinet Office, Government of Japan), Yutaka Kase, and Shota Shimizu (National Graduate Institute for Policy Studies) for data prepossessing and analysis assistance.

REFERENCES

[1] N. N. Abbas, T. Ahmed, S. H. U. Shah, M. Omar, and H. W. Park, "Investigating the applications of artificial intelligence in cyber security," *Scientometrics*, vol. 121, no. 2, pp. 1189–1211, Nov. 2019.
[2] T. Amjad, Y. Rehmat, A. Daud, and R. A. Abbasi, "Scientific impact of an author and role of self-citations," *Scientometrics*, vol. 122, no. 2, pp. 915–932, Feb. 2020.
[3] L. Bornmann and H. Daniel, "What do citation counts measure? A review of studies on citing behavior," *J. Documentation*, vol. 64, no. 1, pp. 45–80, Jan. 2008.
[4] K.-F. Cheung, M. G. H. Bell, and J. Bhattacharjya, "Cybersecurity in logistics and supply chain management: An overview and future research directions," *Transp. Res. E, Logistics Transp. Rev.*, vol. 146, Feb. 2021, Art. no. 102217.
[5] S. M. Dhawan, B. M. Gupta, and B. Elango, "Global cyber security research output (1998–2019): A scientometric analysis," *Sci. Technol. Libraries*, vol. 40, no. 2, pp. 172–189, Apr. 2021.
[6] ENISA. (2020). *15 Top Threats in 2020*. [Online]. Available: https://www.enisa.europa.eu/topics/cyber-threats/threats-and-trends/enisa-threat-landscape-2020-top-15-threats
[7] E. Garfield, "Citation indexes for science: A new dimension in documentation through association of ideas," *Science*, vol. 122, no. 3159, pp. 108–111, Jul. 1955.
[8] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 855–864.
[9] IPA. (2020). *10 Major Security Threats 2020*. [Online]. Available: https://www.ipa.go.jp/files/000084114.pdf
[10] M. Kappi, C. Sab, V. Bagalkoti, and B. Dr, "Scientometrics dimensions of world Bitcoin research a study based on scopus database," *Int. J. Inf. Dissemination Technol.*, vol. 10, no. 2, pp. 82–87, 2020.
[11] Kaseya. (2020). *Top 10 Cybersecurity Threats in 2020*. [Online]. Available: https://www.kaseya.com/blog/2020/04/15/top-10-cybersecurity-threats-in-2020/
[12] M. M. Kessler, "Bibliographic coupling between scientific papers," *Amer. Documentation*, vol. 14, no. 1, pp. 10–25, Jan. 1963.
[13] Y. Kobayashi. (2018). *Effect of Shadow Education Vouchers and the Implications for Evidence-Based Policymaking*. RIETI Rep. [Online]. Available: https://www.rieti.go.jp/en/rieti_report/220.html
[14] W. H. Lee, "How to identify emerging research fields using scientometrics: An example in the field of information security," *Scientometrics*, vol. 76, no. 3, pp. 503–525, Sep. 2008.
[15] R. Leszczyna, "Review of cybersecurity assessment methods: Applicability perspective," *Comput. Secur.*, vol. 108, Sep. 2021, Art. no. 102376.
[16] L. Leydesdorff and S. Milojevic, "Scientometrics," 2012, *arXiv:1208.4566*.
[17] F. A. Loan, B. Bisma, and N. Nahida, "Global research productivity in cybersecurity: A scientometric study," *Global Knowl., Memory Commun.*, vol. 71, nos. 4–5, pp. 342–354, May 2022.
[18] Y. Lu and L. D. Xu, "Internet of Things (IoT) cybersecurity research: A review of current research topics," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2103–2115, Apr. 2019.
[19] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008.
[20] P. R. Makawana and R. H. Jhaveri, "A bibliometric analysis of recent research on machine learning for cyber security," in *Intelligent Communication and Computational Technologies*, vol. 19. Cham, Switzerland: Springer, 2018, pp. 213–226.
[21] *Main Science and Technology Indicators*, OECD, Paris, France, 2022, p. 87.
[22] *Evidence in Education: Linking Research and Policy*, OECD, Paris, France, 2007, p. 182.
[23] N. V. Olijnyk, "A quantitative examination of the intellectual profile and evolution of information security from 1965 to 2015," *Scientometrics*, vol. 105, no. 2, pp. 883–904, Nov. 2015.
[24] S. Rai, K. Singh, and A. K. Vama, "Global research trend on cyber security: A scientometric analysis," *Library Philosophy Pract. (e-J.)*, vol. 3339, p. 3769, Nov. 2019.
[25] V. K. Singh, P. Singh, M. Karmakar, J. Leta, and P. Mayr, "The journal coverage of web of science, scopus and dimensions: A comparative analysis," *Scientometrics*, vol. 126, no. 6, pp. 5113–5142, Jun. 2021.
[26] H. Small, "Co-citation in the scientific literature: A new measure of the relationship between two documents," *J. Amer. Soc. Inf. Sci.*, vol. 24, no. 4, pp. 265–269, Jul. 1973.
[27] H. Small, K. W. Boyack, and R. Klavans, "Identifying emerging topics in science and technology," *Res. Policy*, vol. 43, no. 8, pp. 1450–1467, Oct. 2014.
[28] P. J. Taylor, T. Dargahi, A. Dehghantanha, R. M. Parizi, and K.-K.-R. Choo, "A systematic literature review of blockchain cyber security," *Digit. Commun. Netw.*, vol. 6, no. 2, pp. 147–156, May 2020.
[29] K. Thakur, M. Qiu, K. Gai, and M. L. Ali, "An investigation on cyber security threats and security models," in *Proc. IEEE 2nd Int. Conf. Cyber Secur. Cloud Comput.*, Nov. 2015, pp. 307–311.
[30] V. A. Traag, L. Waltman, and N. J. van Eck, "From Louvain to leiden: Guaranteeing well-connected communities," *Sci. Rep.*, vol. 9, no. 1, p. 12, Mar. 2019.
[31] *Modernising Government*, U.K. Cabinet Office, London, U.K., 1999.
[32] *UNESCO Science Report*, UNESCO, Paris, France, 2010.
[33] B. Xu, Y. Li, and X. Yu, "A scientometric analysis of malware detection research based on CiteSpace," in *Machine Learning for Cyber Security*, vol. 12486. Cham, Switzerland: Springer, 2020, pp. 100–110.

KAZUMASA OMOTE received the Ph.D. degree in information science from Japan Advanced Institute of Science and Technology (JAIST), in 2002. He was with Fujitsu Laboratories Ltd., from 2002 to 2008, and was engaged in research and development for network security. He was a Research Assistant Professor and an Associate Professor with JAIST, from 2008 to 2011 and from 2011 to 2016, respectively, and an Associate Professor with the University of Tsukuba, from 2016 to 2022. Since 2022, he has been a Professor with the University of Tsukuba. His research interests include applied cryptography, network security, and blockchain security. He received the WISTP 2019 Best Paper Award. He was the General Co-Chair of the ACNS 2023 International Conference.

YOKO INOUE received the Bachelor of Laws degree from Chuo University, Japan, in 2018. She was engaged in promoting the utilization of e-CSTI, from 2020 to 2023. She is currently the Staff of the Department of Cabinet Office, Government of Japan.

YOSHIHIDE TERADA received the B.A. and M.A. degrees in economics from Keio University. He was a Project Research Associate with Keio University, from 2019 to 2020, with a focus on sustainable management research. Since 2020, he has been a Professional Staff with the National Graduate Institute for Policy Studies (GRIPS) and has held part-time lecturer positions with Hitotsubashi University and Keio University. His research interests include evidence-based science, technology, and innovation policy at GRIPS.

NAOHIRO SHICHIJO received the Ph.D. degree in computational material science from The University of Tokyo, in 1999. He was a Research Associate with the Research into Artifacts, Center for Engineering (RACE), The University of Tokyo, from 1999 to 2000. He was an Assistant Professor, a Project Associate Professor, and an Associate Professor with the Interfaculty Initiative for Information Studies, The University of Tokyo, from 2000 to 2004, from 2004 to 2007, and from 2007 to 2010, respectively. He was an Associate Professor with Waseda Institute for Advanced Study, from 2010 to 2012, an Senior Research Fellow with the National Institute for Science and Technology Policy, from 2012 to 2016, and a Professor with Tokyo University of Technology, from 2016 to 2020. Since 2020, he has been a Professor with Hitotsubashi University. His research interests include science policy and social data science.

TOSHIYUKI SHIRAI received the B.S. and M.S. degrees from The University of Tokyo, in 1996 and 1998, respectively, and the M.B.A. degree from Georgetown University, in 2004. He has been working with the Ministry of Economy, Trade and Industry, while serving as the Director for evidence-based policy making with the Secretariat of Council for Science, Technology and Innovation, Department of Cabinet Office, from 2021 to 2023, where he directed analysis of research performance in Japan based on funding data and attributes of researchers in universities.