Trend Analysis of Information Security Technology

Seung-Yeon Hwang[1], Jeong-Joon Kim[2*]

[1] Dept of Computer Engineering, University of Anyang, Anyang-si, Gyeonggi-do, Republic of Korea
[2] Dept of Software, University of Anyang, Anyang-si, Gyeonggi-do, Republic of Korea
Email: [1]syhwang@ayum.anyang.ac.kr, [2]jjkim@anyang.ac.kr

ABSTRACT

The amount of data generated by recent developments in bigdata and related technologies has been rapidly increasing, and the need to predict changes in future societies and present technologies to be realized has been continuously raised to lay the foundation for national scientific and technological planning. The existing methods of predicting future technologies have their respective advantages, but problems also exist. Thus, this paper newly establishes and applies the methodology to be used for predicting future technologies specialized in information security fields beyond the existing comprehensive prediction, and draws out innovative technologies that are expected to have high ripple effects in the future, and analyzes the technological diffusion points of each technology to predict future technological changes in the information security sector. It is expected that this will ensure reliability and objectivity of the forecast survey results and allow more sophisticated and multilayered predictions than the overall scientific and technological forecast surveys.

Keywords

Bigdata, Information Security, Prediction Methodology etc.
Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

Introduction

With the recent development of big data and related technologies, the amount of data generated has been rapidly increasing, and the need to lay the foundation for national science and technology planning has been continuously raised [1]. Accordingly, the Science and Technology Forecasting Survey has been conducted and published every five years since 1994, and the "5th Science and Technology Forecasting Survey" report was published by the Korea Institute of Science and Technology Planning (KISTEP) in 2017. In the above survey, we forecast and analyze future societies in consideration of internal and external environmental changes, and predict and analyze future technologies that are expected to emerge across all fields of science and technology by 2040, based on changes in future social demand [2].

Methods such as Delphi, Futures Wheel, Economic, and Horizon Scanning are mainly used as future technology prediction methods. These prediction methods have their own advantages, but the following limitations exist: In the case of Delphi surveys, which repeatedly conduct surveys to experts to establish an average image of the future, it is difficult to secure representation and objectivity of participating experts. Future wheels, which construct possible causal relationships around events such as major technological upheaval, may have weak grounds based on a few imaginative prospects. Existing statistical data-based economic metrics can have a narrow range of data because they are predicted based on given data. Horizon scanning, which identifies variation across science and technology, can be an overview of the microdynamics between more detailed technologies.

Therefore, this study newly establishes and applies methodologies to predict future technologies specialized in information security fields, deriving innovative technologies that are expected to have high ripple effects in the future, and analyzing technology diffusion points of each technology to predict future technology changes.

This paper describes the techniques required for predicting future technologies in the field of information security in Chapter 2, and Chapter 3 describes the research content and methods of this study. Chapter 4 interprets the data analysis and concludes in Chapter 5.

Related Work

2.1. Big Data

Big data includes structured data such as tables covered by existing databases, as well as semi-structured data such as xml and html documents and unstructured data such as text, images, and speech. Big data is treated as five stages of collection-storage-processing-analysis-

visualization, and through these processes, it is used in various ways, such as discovering or predicting significant patterns in diversified modern society. The characteristics of big data are 3V (Volume, Variety, Velocity), which means the amount of data, the type of data, and the speed of data, respectively [3, 4].

2.2. Text Mining

Text mining is the field of exploring advanced information hidden in unstructured documents. Applications of text mining include categorization, which analyzes the content of text to properly assign predefined categories, clustering text into multiple congregations according to content similarities, and summarization to extract content that can represent the entire content of the document.[5, 6]. In this work, we construct text with similar themes through clustering from collected text data into the same semantic network.

2.3. Web Crawling

Crawling is a technology that collects data, including numerous documents on the Internet, and automatically collects various information in a web environment. In addition, the importance of web crawlers is further highlighted by the increase in content from various media such as SNS, news, etc. [7]. In this work, we utilize web crawlers to collect abstract data from several journal and conference papers.

2.4. TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) is a statistical figure used to evaluate how important the keywords in a document are for text mining. The TF-IDF weight is multiplied by TF and IDF. The TF value means the frequency of the corresponding keyword within the document, and the IDF value means the total number of documents divided by the number of documents in which a specific keyword appears. This is likely to be a universal keyword, with the IDF value of keywords appearing in many documents becoming smaller. On the contrary, the IDF value of keywords that frequently appear in certain documents increases and are likely to be keywords that have important meanings in those documents. This means that keywords with large TF-IDF values are likely to relate to the topic of the document to which they belong [8, 9]. In this work, we use TF-IDF techniques to extract the importance ranking of keywords.

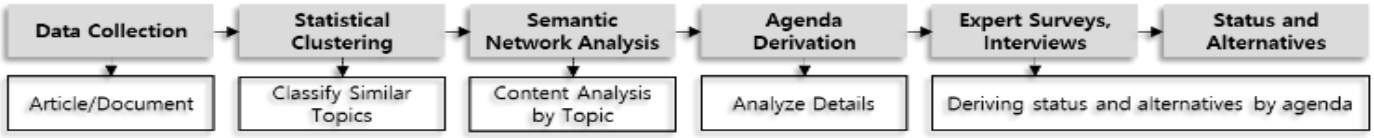

Fig 1: Research Procedures

Table 1: Data Source

Journal Lists	Total Count
ACMCCS(ACM Computer and Communications Security Conference) NDSS(The Network and Distributed System Security Symposium) USENIX security(ACM Computer and Communications Security Conference) ACM TAAS(ACM Transactions on Autonomous and Adaptive Systems) ACM TECS(ACM Transactions on Autonomous and Adaptive Systems) ACM TOCS(ACM Transactions on Computer Systems) ACM TODS(ACM Transactions on Database Systems) ACM TOIT(ACM Transactions on Internet Technology) ACM TOPS(ACM Transactions on Privacy and Security) ACM TOSN(ACM Transactions on Sensor Networks) IEEE TIFS(IEEE Transactions on Information Forensics and Security) IEEE IS(IET Information Security) IEEE SP(IEEE Security & Privacy) IEEE TDSC(IEEE Transactions on Dependable and Secure Computing) SD JISA(Science Direct Journal of Information Security and Applications) SD CS(Science Direct Computers & Security)	98,647

Table 2: Search terms in 7(1~7) security technologies

Class	Terms
1	Quality Management, mobile security, Secure Token, VPN, Secure Communication, Command Control, Cryptographic Implementation, Hardware architecture, Communication protocol, Key Management, Hardware Security Module, Cryptography System, Satellite Communication Security, System On-Chip Design, cryptography devices, system testing, telecommunication security, cryptographic protocols
2	Intrusion Detection, Threat Intelligence, Advanced Persistent Threat(APT), C&C traffic, Security Evaluation, Cyber Security, Malware Analysis, deobfuscation, Reverse Engineering, Security Automation, Cloud Security, Cyber deception, Cyber Incident Response, Mobile Malware, Cyber Security Training, Correlation Analysis, Behavior Analysis, Anti Analysis, Penetration Test
3	Steganography, Data Hiding, Digital Forensic, Anti-Forensic, Fuzzing, Vulnerability Analysis, Exploit, Reverse Engineering, Mobile Vulnerability, Hardware Vulnerability, Covert Channel, OS Vulnerability, Cyber Attack, Network Device Vulnerability, IoT Vulnerability, Biometric Security
4	Electromagnetic Pulse(EMP), Surge Protection, Electromagnetic Shielding, Direction Finding, Fake BTS, Vulnerability Analysis, Wireless Signal Detection, Wireless Signal Identification, Anti-drone, Cyber Physical System(CPS), Industrial Control System(ICS), Supervisory Control And Data Acquisition(SCADA), Smart City, Smart Factory, Smart Manufacturing, STIX/TAXII
5	Cryptographic Module Validation Program(CMVP), Tamper, Side-Channel Attack, Cryptographic Device Verification, Cryptographic Device Security, Verification Technique, Approximate Computing, Physical Layer Security, Hardware Testing, Supply Chain Security, Test Automation, Security Critical System Evaluation, Common Criteria(CC), Security Evaluation, Protection Profile(PP), Vulnerability Attack, Validation Scheme, security Assurance, Device Evaluation, security certification, security validation
6	authentication, block cipher, code based, crypto chip, cryptographic protocol, digital signature, fault injection, hash function, implementation, key exchange, lattice based, lightweight, multivariate equation, post quantum, power analysis, provable security, public key, quantum algorithm, quantum cryptography, quantum key distribution, security proof, side channel attack, stream cipher, symmetric key
7	Physical Security, Biometric, Bio Security, Human Identity, Video Monitoring, CCTV Security, Video Surveillance, Security Screening, Security System

Procedures

The research procedures in this study are shown in Figure 1. First, we collect data from the paper and construct text with similar themes into the same semantic network through statistical clustering, etc. based on TF-IDF. In addition, a comprehensive trend and meaning for each topic are identified through mathematical statistics and semantic network analysis for each security technology item. Finally, agenda items for each topic shall be derived through a qualitative detailed analysis, and alternatives shall be drawn after organizing the status of each agenda.

3.1. Data Collection

In this study, we seek to identify the factors of interest and research trends in security technology and R&D issues in major academic papers and conferences over the last four years (2016-2019). Therefore, 98,647 papers green data were collected from a total of 16 journals or conferences by confirming corresponding journals, conferences, and search terms in seven foreign (1-7) security technologies through expert panels, as shown in Table 1. The search terms for seven (1-7) security technologies are shown in Table 2.

3.2 Extract and Structure Important Information

We extract noun morphemes through the main steps of natural language processing from the abstract data collected during the data collection phase. We then unify singular and plural words and structure keywords through stopwords processing. In this work, TF-IDF techniques from text mining are utilized to extract core keyword rankings by reflecting the relative proportion of the frequency of keywords to the number of documents they contain. It then combines core keywords and concurrent associations to form a co-occurrence matrix, which connects keywords that co-occur in each paragraph unit to form a semantic network.

3.3 Representation of Core Keywords Network

We derive important keywords from a set of keywords derived through text mining by utilizing the value of network centrality. Network centrality is represented by various indicators depending on its topological function. Degree Centrality measures the frequency of connections on a particular node. Betweenness Centrality measures the degree to which a cluster is centered. Closeness Centrality measures the degree to which a particular node has the shortest distance over the other entire nodes. The Eigenvector weights the connectivity centrality of adjacent nodes to measure how much they are connected to influential nodes, and the Bonasich power is calculated by adding a combination of the connectivity centrality of the Eigenvector.

3.4 Statistical Prediction of Security Technology Trends

In this work, acceleration metrics are utilized along with quarterly keyword frequency variations to quantify changes in core security technology concepts and to predict future mobility. Acceleration is assumed to be an interval quadratic function that draws curves of frequencies, then regression is applied, and the coefficient value of order 2 is extracted. By representing the detailed descriptive keywords in each of the eight security technology categories as the quarterly average frequency (X-axis) and acceleration (Y-axis), we would like to identify detailed technologies with low current frequency but high recent mobility.

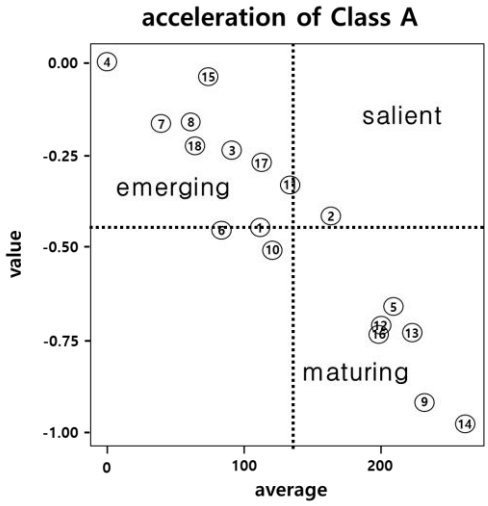

Fig 2: Research Procedures

As shown in Figure 2, we distinguish the technology in 1st quadrant into a salient area with both high frequency and acceleration. The 2nd quadrants are divided into the emerging area, where frequency is still lower than the overall mean, but acceleration is high. 4th quadrants are divided into maturing area with high frequency but low acceleration.

3.5. Deriving Core Mediated Technologies using Semantic Network Analysis

Detailed technologies or major concepts are derived for each category that has high mediated centrality in the co-occurrence network and is highly likely to be mediated between technologies. Through this, it is expected that it will be able to invest intensively for the development of security-related technologies in the future or use the information to derive agenda for the development of linked technology ecosystem.

Analyze Data

This chapter describes the analysis and results of Class 1 among the seven classes, and organizes and compares the derived results by class.

4.1. Class 1 Analysis Results

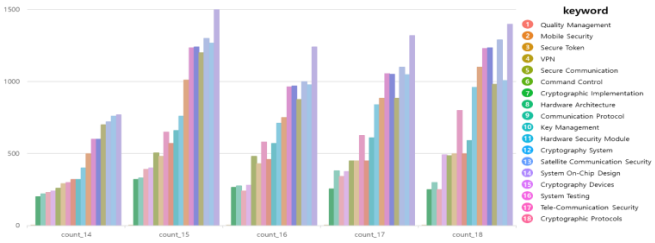

Fig 3: Annual Frequency Trends of Class 1

Figure 3 shows a graph of the frequency trend of each year of the Class 1 technology keyword. From the top of the graph, the system on-chip design, satellite communication security, and communication protocol have high annual frequencies, and are generally on the rise.

Table 3 shows the ranking of the mediastinities of the technologies in Class 1. Mobile security, secure communication, cryptographic protocols, cryptographic devices, and system on-chip design are ranked high, indicating that these technologies are the primary intermediaries in the formation of the Class 1 technology ecosystem.

Table 3: Class 1 security technology's Medianity ranking

Rank	Technology	Medianity
1	mobile security	121.1
2	Secure Communication	110.2
3	cryptographic protocols	88.7
4	cryptography devices	86.3
5	System On-Chip Design	70.7
6	Quality Management	69.4
7	Cryptographic Implementation	50.2
8	Communication protocol	49.8
9	System testing	48.5
10	VPN	41.3
11	Key Management	32.9
12	Hardware Security Module	20.1
13	Command Control	10.2
14	Satellite Communication Security	8.4
15	Cryptography System	6.3
16	Secure Token	3.4
17	telecommunication security	0.5
18	Hardware architecture	0.2

Figure 4 shows the results of the abstract network analysis in Class 1. The number of circles shown in red in the figure represents the medianity ranking in Table 3. Mobile security (1) and cryptography devices (4) are central to the entire network, and secure communication (2) is connected to the cloud, network, and Internet. With regard to Privacy, we can see that cryptographic protocols (3) are directly connected.

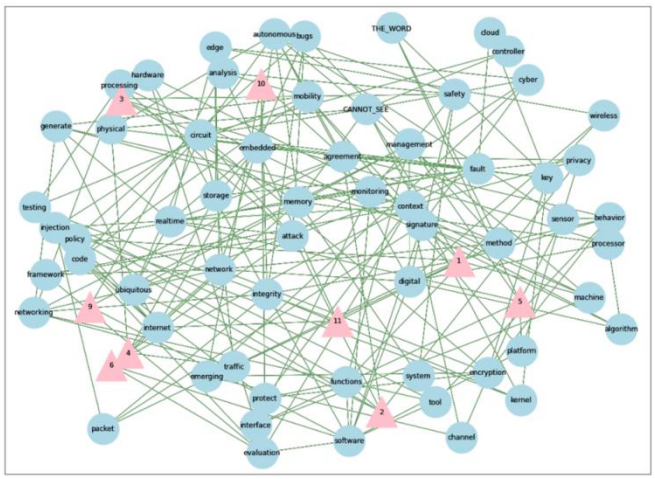

Fig 4: Keyword Network of Class 1

Table 4: Comparing the Medianity and PP of Class 1

Rank	Technology	Medianity	PP
1	mobile security	121.1	0.75
2	Secure Communication	110.2	0.63
3	cryptographic protocols	88.7	0.58
4	cryptography devices	86.3	0.56
5	System On-Chip Design	70.7	0.53
6	Quality Management	69.4	0.48
7	Cryptographic Implementation	50.2	0.44
8	Communication protocol	49.8	0.37
9	System testing	48.5	0.34
10	VPN	41.3	0.27
11	Key Management	32.9	0.24
12	Hardware Security Module	20.1	0.22
13	Command Control	10.2	0.17
14	Satellite Communication Security	8.4	0.13
15	Cryptography System	6.3	0.11
16	Secure Token	3.4	0.02
17	telecommunication security	0.5	0.02
18	Hardware architecture	0.2	0.01

Table 4 shows the comparison results of PP with the medianity of Class 1. Due to the high PP of System on-chip design, it is predicted that the possibility of future mediation may increase.

Figure 5 shows a graph of the quarterly average frequency and acceleration of the Class 1 technology keyword. This allows us to determine keywords that are relatively accelerated compared to frequency, which are likely to emerge in the future. In Class 1, mobile security from a emerging perspective has recently been dominant in motility, and hardware security modules have similar patterns. In the future, in particular, the rise of cryptography devices is predicted, and VPNs have room for observation as the average frequency is close to zero.

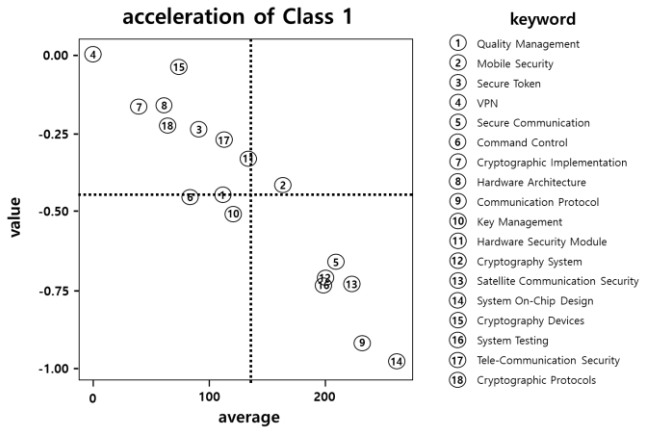

Fig 5: Average Frequency and Acceleration of Class 1 Keywords on Quarterly

4.2 organization and comparison each group

In this section, we compare future descriptive names and necessary element techniques derived by a panel of experts with technical keywords and associations derived by quantitative analysis results. In addition, we organize and compare trends implemented by qualitative methods of expert panels and quantitative methods of big data, respectively, by group.

Table 5: Results of Comparison and Analysis of Class 1

Class	Experts	Analysis
	Promising Technology	**Promising Technology**
1	Data security, IoT convergence system	Mobile security
	Blockchain hidden channel communication	secure communication
	IoT Key Agreement Protocol	cryptographic protocols
	5G RAN Security	system on-chip design
	BCI vulnerability analysis technology	hardware security mod.

Table 6: Results of Comparison and Analysis of Class 2

Class	Experts	Analysis
	Promising Technology	**Promising Technology**
2	Analysis of IoT Malicious Code Vulnerabilities	malware analysis
	Intelligent Digital Forensics	cyber security
	Security check automation technology	cyber security training
	Home IoT Intrusion Detection	threat intelligence
	Mobile Cloud Security	cloud security

Table 7: Results of Comparison and Analysis of Class 3

Class	Experts	Analysis
	Promising Technology	**Promising Technology**
3	Roadbot Vulnerability Diagnostics	IoT vulnerability
	Mobile Malicious Code Detection	mobile vulnerability
	Hidden channel steganography	steganography
	Digital forensics	biometric vulnerability
	5G Communication Stability	network dev. vulnerability

Table 8: Results of Comparison and Analysis of Class 4

Class	Experts Panel	Analysis
	Promising Technology	**Promising Technology**
4	5G Industrial Wireless Communication	SCADA
	Control System Instrument Vulnerabilities	STIX
	Machine Vision Security	wireless signal identification
	Production and Control Vulnerabilities	wireless signal detection
	Industrial Security Standards	cyber physical system

Table 9: Results of Comparison and Analysis of Class 5

Class	Experts Promising Technology	Analysis Promising Technology
5	Sub-channel Attack Analysis	validation scheme
	NFV virtualization network	physical layer security
	Crime Prediction AI Verification	security evaluation
	Firmware security verification	CMVP
	Software Evaluation	-

Table 10: Results of Comparison and Analysis of Class 6

Class	Experts Promising Technology	Analysis Promising Technology
6	Sub-channel analysis of high-end devices	quantum algorithm
	quantum-resistant password sub-channel analysis	quantum key distribution
	Quantum Computer Lightweight Symmetry Key	quantum cryptography
	Homomorphic password for machine learning	post quantum
	Validate Approximation Operations	-

Table 11: Results of Comparison and Analysis of Class 7

Class	Experts Promising Technology	Analysis Promising Technology
7	Ultra-precision UV transmittance video	biosecurity
	Terahertz Security Scan	security system
	Emotional Awareness Criminal Behavior Prediction	-
	Image Analysis for Robotics	-
	AI-based automatic reading of commodities	-

Conclusion

In this work, we compare future technical names and necessary element technologies from seven security technology fields derived through expert panels with technical keywords and associations derived by quantitative analysis results from this study. We experiment with indicators for future prediction through natural language processing, semantic network analysis, and statistical mathematical approaches to big data in paper surges, which previously consisted mainly of expert panels. As a result, we were able to identify which of the given technology categories was the main existing technology. And we were able to identify what are highly mediated and expanding the appearance of the technology ecosystem. In addition, even if the frequency, mediated centrality, etc. is not high in the current context, if it has high acceleration or PP values, it can be classified as a technology with high potential, giving room for further interpretation. Comparing the results of big data analysis with the results of future promising technologies from expert panels, most of them overlap similarly to ensure a certain degree of reliability and objectivity. However, in the association analysis of the network, there was a limit to showing discriminative value in deriving new concepts because everyday words were derived as the main association in the paper. Therefore, it is expected that if detailed information indexing in Phrase units is done for each technology group item, security technology semantic networks can be implemented to reflect that information and specifically to show changes in the technology ecosystem.

References

[1] Seung-Yeon Hwang, Ji-Hun Park, Ha-Young Youn, Kwang-Jin Kwak, Jeong-Min Park, Jeong-Joon Kim (2019) Big Data-based Medical Clinical Results Analysis. The Institute of Internet, Broadcasting and Communication 19(1):187-195

[2] Korea Institue of S&T Evaluation and Planning(KISTEP) (2017) 5th Science and Technology Forecasting.

[3] Seung-Yeon Hwang, Dong-Jin Shin, Kwang-Jin Kwak, Jeong-Joon Kim, Jeong-Min Park (2019) Real-time Processing of Manufacturing Facility

Data based on Big Data for Smart-Factory. The Institute of Internet, Broadcasting and Communication 19(5):219-227

[4] Seung-Yeon Hwang, Jin-Yong Moon, Jeong-Joon Kim (2019) Relationship Analysis between Fine Dust and Traffic in Seoul using R. The Institute of Internet, Broadcasting and Communication 19(4):139-149

[5] Jae-Young Chang (2013) A Study on Research Trends of Graph-Based Text Representations for Text Mining. The Journal of The Institute of Internet, Broadcasting and Communication 13(5):37-47

[6] Seung-Yeon Hwang, Kyung-Min Kwak, Dong-Jin Shin, Kwang-Jin Kwak, Young-J Rho, Kyung-won Park, Jeong-Min Park, Jeong-Joon Kim (2019) Analysis of Defective Causes in Real Time and Prediction of Facility Replacement Cycle based on Big Data. The Institute of Internet, Broadcasting and Communication 19(6):203-212

[7] Jong-Hwa Lee (2018) Building an SNS Crawling System Using Python. the Korea Industrial Information Systems Research 23(5):61-76

[8] Sungjick Lee, Han-joon Kim (2009) Keyword Extraction from News Corpus using Modified TF-IDF. Society for e-Business Studies 14(4):59-73

[9] Eun-Hee Jeong, Byung-Kwan Lee (2018) An Analysis Scheme Design of Customer Spending Pattern using Text Mining. Korea Institute of Information, Electronics, and Communication Technology 11(2):181-188