Journal of the Korea Academia-Industrial cooperation Society
Vol. 20, No. 5 pp. 492-501, 2019

https://doi.org/10.5762/KAIS.2019.20.5.492
ISSN 1975-4701 / eISSN 2288-4688

빅데이터를 활용한 국내 보안솔루션 시장 동향 분석

박상천*, 박동수
국방기술품질원

Analysis of Domestic Security Solution Market Trend using Big Data

Sangcheon Park*, Dongsoo Park
Defense Agency for Technology and Quality

요 약 사이버 공간에서 안전하게 시스템을 사용하기 위해서는 상황에 적합한 보안 솔루션을 사용해야 한다. 사이버 보안을 강화하기 위해 과거부터 현재까지 보안의 흐름을 정확히 파악하고 미래의 다양한 위협에 대비해야 한다. 본 연구에서는 텍스트마이닝을 이용하여 신뢰도가 높은 네이버 뉴스의 보안/해킹 뉴스의 정보보안 단어들을 수집 후 분석하였다. 첫 번째는 지난 7년의 연도별 보안 뉴스 기사수를 확인하고 추이를 분석하였다. 두 번째는 보안/해킹 관련 단어 순위를 확인 후 매년 주요 관심사를 확인하였다. 세 번째는 보안 솔루션별 단어를 분석하여 어느 보안 그룹의 관심도가 높은지 확인하였다. 네 번째는 보안 뉴스의 제목과 본문을 분리 후 보안 관련 단어를 추출 후 분석하였다. 다섯 번째는 세부 보안 솔루션별 추이 및 동향을 확인하였다. 마지막으로 연도별 매출액과 보안 단어 빈도수를 분석하였다. 이러한 빅데이터 뉴스 분석을 통해 보안 솔루션에 대한 전반적인 인식 조사를 수행하고 많은 비정형 데이터를 분석하여 현재 시장 추세를 분석하고 미래를 예측할 수 있는 정보를 제공하는 데 기여하고자 한다.

Abstract To use the system safely in cyberspace, you need to use a security solution that is appropriate for your situation. In order to strengthen cyber security, it is necessary to accurately understand the flow of security from past to present and to prepare for various future threats. In this study, information security words of security/hacking news of Naver News which is reliable by using text mining were collected and analyzed. First, we checked the number of security news articles for the past seven years and analyzed the trends. Second, after confirming the security/hacking word rankings, we identified major concerns each year. Third, we analyzed the word of each security solution to see which security group is interested. Fourth, after separating the title and the body of the security news, security related words were extracted and analyzed. The fifth confirms trends and trends by detailed security solutions. Lastly, annual revenue and security word frequencies were analyzed. Through this big data news analysis, we will conduct an overall awareness survey on security solutions and analyze many unstructured data to analyze current market trends and provide information that can predict the future.

Keywords : Big Data, Text Mining, Information Security, Cyber Security, Security Solution

*Corresponding Author : SangCheon Park(Defense Agency for Technology and Quality)
Tel: +82-51-750-5143 email: wibrostar@dtaq.re.kr
Received March 28, 2019 Revised April 12, 2019
Accepted May 3, 2019 Published May 31, 2019

1. 서론

최근 인터넷이 일상생활 속에 자리 잡고 4차 산업혁명과 인공지능, IoT, 빅데이터 등이 급속도로 발전해 가면서 사이버 보안의 중요성은 점점 더 커지고 있다[1]. 사이버 보안은 세계적으로 주요한 관심사로 대두되고 있다. 중국의 경우 '정보화 영도조소'라는 기구를 신설하여 국가 사이버보안을 총괄하고 있으며, 미국의 경우 '17년 12월 '블록체인 사이버 보안연구'법안에 대통령이 서명함으로써 사이버보안 응용 프로그램 조사를 촉구한 바 있다[2]. 이처럼 국제적으로 사이버 보안 영역은 기업이나 개인의 영역이 아닌 국가적 차원에서 빠르게 변화하고 있다[3].

이에 현 정부도 '사이버 위협에 대응하기 위한 국가 차원의 사이버 안보 대응 역량 강화'라는 국정 과제를 가지고 사이버 공간의 안전한 보호 및 사이버전 수행 능력을 확보하고 있다[4]. 정부는 국가안보실 중심의 사이버안보 컨트롤타워 강화와 체계적인 사이버 안보 수행 체계의 정립을 통해 선진국 수준의 사이버 안보 대응 역량을 갖추기 위해 노력 중에 있다. 사이버 공격이 점차 지능화되고 대규모화 되면서 경제적 피해를 넘어 사회적 혼란야기 및 국가 안보 위협 수준까지 이르렀기 때문이다. 게다가 4차 산업혁명에 따른 다양한 기술간의 융합이 보안의 수요와 업무를 계속 증가시키고 있다. 반면 보안업계에 종사하는 정보보호 인력은 정부 및 공공기관, 일반업체를 포함하여 약 12만명이지만, 이중 정보보호 전공자는 겨우 750여명에 불과하다. 이처럼 보안인력의 수요와 공급의 불일치현상에서 정보보안 인력이 현저하게 부족한 것을 확인할 수 있다. 이에 대한 대안으로는 적절한 보안 솔루션 도입으로 인력부족현상을 보완하는 방법이 있다[5].

최근 4차 산업혁명의 흐름에 맞춘 학제간 융합연구의 중요성이 부각되면서 소셜미디어 빅데이터 분석과 관련된 연구가 지속적으로 증가하고 있다. 그리고 지난 10년간 문화, 디자인, 화학, 나노, 바이오, 로봇, IT, 정보통신 분야에서 빅데이터를 통한 시장동향 연구가 이루어 졌다[6].

보안 분야에서의 빅데이터 분석기법은 시장 동향에 초점을 맞추기 보다는, 해킹 공격 분석시에 예전에 보이지 않았던 다양한 패턴을 발견하는 방향으로 주로 활용되어 왔다. 그리고 지금까지의 보안솔루션 시장 동향 분석은 소셜미디어와 뉴스의 정보에 의존하는 형태가 아닌 IBM, McAfee, HP, Splunk와 같은 주요 IT 업체의 정보에 주로 의존하고 있었다[7,8]. 다양한 분야 및 주제에 대하여 텍스트마이닝을 통해 문서의 핵심 키워드 추출하고 분석하는 연구들이 존재하지만 뉴스정보로부터 보안솔루션의 동향을 파악하는 연구는 없었다[9,10].

따라서 본 연구에서는 각종 뉴스 속의 보안 솔루션의 단어 출현 빈도를 분석하고 이를 보안 솔루션의 매출 현황과 비교분석하여 보안시장동향을 분석한다. 추가로 과거부터 현재까지 솔루션 시장 동향분석 데이터를 토대로 미래 보안시장 예측을 시도하였다.

2. 본론

2.1 연구 방법 및 분류 정의

본 연구에서는 보안 솔루션 시장분석을 위한 자료수집을 위해 R 프로그램을 사용하였다. 크롤링(crawling) 기능으로 수집한 데이터를 전처리 과정으로 데이터를 정제 후 분석하였다. 분석 절차는 그림 Fig. 1에서 보는 바와 같다. 활용 데이터는 신뢰도 높은 한국정보보호산업협회(KISIA: Korea Information Security Industry Association)의 '정보보호산업 실태조사 보고서(Survey for Information Security Industry in Korea : Year 2018)'와 네이버(NAVER) 뉴스의 보안/해킹 기사를 활용하였다. INTERNET TREND에 의해 작성된 국내 포탈 사이트 점유율 순위에 의하면 2012년도 5월부터 2018년도 12월까지 네이버가 평균 80.23%의 점유율을 기록하여 신뢰도 높은 자료를 제공할 것이라 판단하였다[11,12].

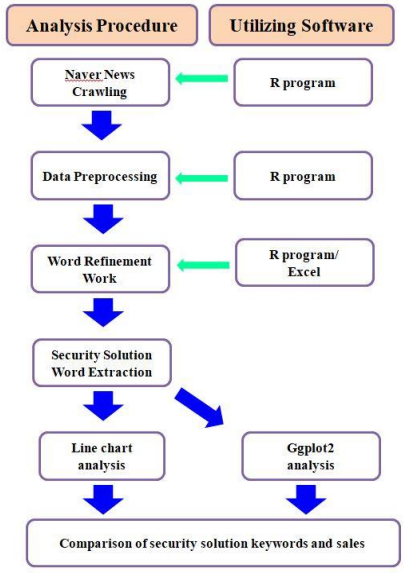

Fig. 1. Analysis procedure

본 논문에서는 상기 실태조사 보고서의 보안 솔루션에 관한 매출액 자료를 기반으로 빅데이터 텍스트마이닝(textmaining)을 이용하여 매출액과 웹기사의 솔루션 단어 빈도수 수치를 확인 후 향후의 국내 보안 솔루션 시장을 예측하고자 한다.

예측에 앞서 보안 솔루션의 정의는 KISIA 보고서의 분류에 따라 5가지(네트워크 보안, 시스템(단말) 보안, 콘텐츠(데이터)/정보유출방지 보안, 암호/인증, 보안관리)로 정의하였다[13].

2.2 분석 도구

네이버 '보안/해킹' 카테고리 뉴스에서 보안 솔루션 데이터 수집을 위하여 오픈소스 R 프로그램 3.4.3 버전를 사용하였다. 기사 수집을 위해 N2H4 패키지의 R source를 수정하여 사용하였다[14]. 수집한 기사를 tm 패키지, dplyr 패키지, useSejongDic 사전 패키지, stringr 패키지, KoNLP 패키지를 사용하여 텍스트마이닝을 진행하였다.

2.2.1 자료의 수집

네이버 '보안/해킹'뉴스 기사를 대상으로 R 프로그래밍을 통한 뉴스의 제목 및 본문 기사를 크롤링(crawling)하였다.

네이버 뉴스는 총 76개 언론사에서 뉴스를 제공 받고 있다. 그 중 보안 솔루션은 IT분야인 보안/해킹 기사에서 다룬다. Table 1에서 보는 바와 같이 "아이뉴스24, 전자신문, 디지털데일리, 디지털타임스" 등 다양한 신문사에서 보안/해킹 관련기사를 상당수 다루고 있으며, 보안/해킹관련 네이버 뉴스에서 97% 정도를 차지하고 있다. 네이버에서 '보안/해킹' 기사 게재를 시작한 2012년 5월 3일부터 2018년 12월 31일까지 약 7년간의 기사를 수집 및 정제하였다.

Table 1. Security/Hacking Articles(Unit: number)

Year / News	2012	··	2017	2018	sum
Inews 24	1812	··	2812	3016	15108
Electronic newspaper	683	··	797	1204	8301
Digital Daily	795	··	1388	946	5789
Digital Times	555	··	141	591	5074
Total	4100	··	5219	5896	35065

총 35,065기사를 수집하였고, 수집된 단어의 종류는 119,667개, 분석된 단어의 수는 제목(Title) 단어 153,051개, 본문(Body) 단어 4,497,341개로 총 4,650,392개 이다.

2.2.2 데이터 전처리

수집한 기사에 대하여 텍스트 전처리 과정을 시행하였다. R 프로그램의 dplyr, stringr, KoNLP 패키지를 이용하여 전처리 과정을 진행하고, 370,957 단어를 보유하고 있는 useSejongDic 세종 사전 패키지를 사용하였다[15]. 의미있는 단어를 추출하기 위해 다음과 같은 정제 작업을 진행하였다. 1차는 '수', ' 및', '등', '있는', '고'와 같은 조사 제거 작업을 진행하였다. 2차는 'nkc', 'advance', 'mou', 'adc', 'access'와 같은 외래어 제거 작업을 진행하였다. 3차는 '차단한다', '상황이다', '차지했다', '확인됐다', '평가했다'와 같은 동사 단어를 제거하였다. 4차는 '` ∀', '〃', '〃s', '〃易云音樂', '☎02'와 같은 특수 문자 및 한자를 제거하였다. 5차는 '따르면', ' 때문에', '신속하게', '기존에', '높이기'와 같은 형용사, 부사형, 접속어를 제거하였다. 6차는 '흑자전환', '흐름', '휴지', '휴스턴', '휘발유'와 같은 의미없는 단어를 제거하였다. 7차는 '방화벽어슈어런스', '방화벽뿐만아니라', '방화벽뿐만', '방화벽마다' 단어를 '방화벽'으로 단어에 붙어 있는 조사 및 단어 통합 작업을 진행하였다. 추가적으로 str_replace() 함수와 gsub() 함수를 사용하여 유사 단어 통합 및 보안 외 의미없는 단어 삭제 작업을 진행하였다.

2.2.3 보안 솔루션 단어 추출

전처리 과정을 마치고 수집한 결과는 Table 2에서 보는 바와 같다. 세부항목의 기준은 한글명, 영문명, 약어를 기준으로 작성하였다. 추가적으로 일반적으로 부르는 제품명도 포함하였다. 예를 들어 방화벽은 방화벽, 네트워크방화벽, 시스템방화벽, Firewall, Network Firewall, System Firewall로 세부항목을 분류하였다. Table 2는 영문 키워드만 작성하였다.

2.3 분석 결과

본 연구에서는 보안/해킹 기사분석 결과를 여러 가지 Chart로 도출하였다. 첫 번째는 연도별 기사수의 증감 분석과 변동 사유에 대하여 확인하였다. 두 번째는 보안/해킹 기사의 키워드를 제목과 본문으로 구분하여 분석하였다. 세 번째는 보안 솔루션 그룹별 집합의 추이를 확인

하였다. 네 번째는 보안 솔루션의 제목과 본문의 키워드를 중심으로 빈도순위를 분석하였다. 다섯 번째는 매년 보안 솔루션별 그룹의 추이 분석을 하였고, 마지막으로 보안 솔루션의 매출액과 솔루션의 키워드 빈도수를 대조 분석함으로서 추후 동향분석을 하였다.

Table 2. Security Solutions Classify

NS	SS	CILPS	PA	SM
Web Firewall	SAC	DB security	OP	ISM
Firewall	Malware	DB password	PKI	TMS
IPS	Anti -Malware	Secure USB	IAM	PMS
DDoS	Spam Protection	DCM	IATM	BMS
ISS	SOS	DLP		MLS
VPN	APT -Response	IPMS		VAS
NAC	Mobile Security	DB access control		DFS
WNS	PC firewall			RM
NSe	SPSW			Log analysis
NF	SecureOS			VA
SF				

*NS: Network Security, SS: System (Terminal) Security, CILPS: Content(data)/Information leakage prevention security, PA: Password/Authentication, SM: Security Management, ISS: Integrated Security System, VPN: Virtual Private Network, NAC: Network Access Control, WNS: Wireless Network Security, NSe: Network Separation, NF: Network Firewall, SF: System Firewall, SAC: System Access Control, SPSW: Spam protection SW, SOS: Security Operating System, DCM: Digital copyright management, IPMS: Individual Privacy Management System, OP: One-time password, PKI: Public Key Infrastructure, IAM: Integrated Access Management, IATM: Integrated account management, ISM: Integrated security management, TMS: Threat Management system, BMS: Backup Management system, MLS: Manage logs system, VAS: Vulnerability Analysis system, DFS: Digital forensics system, RM: Recovery Management, VA: Vulnerability Analysis

2.3.1 연도별 뉴스(보안/해킹) 기사수

Fig. 2을 보면 2015년도가 기사수가 가장 적고, 2016년도에 가장 많은 것을 알 수 있다. 2015년도는 5월 2일부터 뉴스 게재를 시작한 것 감안하면 2012년도부터 2015년도 까지는 전반적으로 기사수가 감소한다는 것을 알 수 있다. 정부의 지식정보사회 구현(국가정보화기본법 제6조) 추진으로 2012년도 정보통신서비스의 대표적인 기기인 스마트폰 가입자는 2,756만명을 돌파하고, 모바일 IPTV 서비스를 제공하기 시작하였다. 2012년 개인정보보호기술지원센터 개소를 시작으로 2013년 정보보호 산업 종합대책을 발표하면서 정보보호의 안정화를 찾아가는 듯 보였지만, 민간 해킹사고 접수처리 1,444건, 국내 피싱사이트 6,944건 차단으로 안심 할 수 있는 상황은 아니였다.

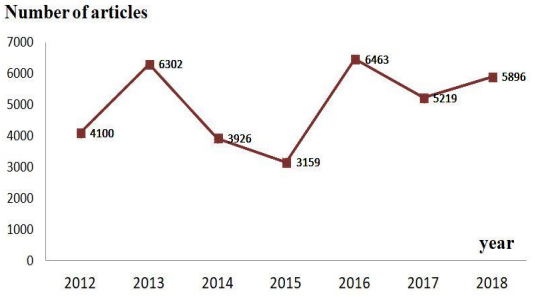

Fig. 2. Number of security/hacking articles by year

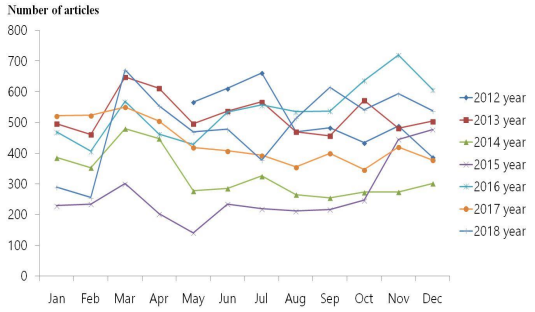

Fig. 3. Monthly Security/Hacking Articles

그래서 정부는 2015년 후반부터 '개인정보 비식별 조치 가이드라인' 시행, 2016년 '클라우드 컴퓨팅 서비스 정보보호에 관한 기준 고시', 2017년 '정보보호 제품 성능평가 제도' 시행으로 지속적으로 정보보호를 위해 노력하고 있다. 2015년 후반부터 'K-ICT 전략 2016' 발표를 기점으로 보안/해킹 기사수가 증가하였다[16].

Fig. 3을 보면 전반적으로 1월에서 2월에는 감소하다가 3월에는 상승하고 5월까지는 감소하다가 다시 5월부터 7월까지는 상승하는 것을 볼 수 있다. 8월과 9월을 평이하게 가다가 10월부터는 다시 상승하는 경향을 보인다. 주목 할 만한 점은 1월과 9월 설 혹은 추석 같은 명절 전·후 연휴 관련 택배 사칭 문자와 정치적 이슈를 활용하여 문자를 통한 해킹 사고가 많아진다는 특징이 있다.

2.3.2 보안/해킹 키워드 분석

Table 3. Security/Hacking News Title Keyword Frequency Rank

Year	1st	2nd	3rd	4th	5th
2012	Sec	Sol	Ent	Service	Tec
2013	Sec	Sol	Ahn Lab	Cyber	Mal
2014	Sec	Sol	IPr	Ent	Cyber
2015	Sec	IPr	Sol	Cyber	Ent
2016	Sec	Sol	Cyber	Cloud	Ent
2017	Sec	Sol	Cyber	Ransom ware	Cloud
2018	Sec	Block Chain	Virtual	Sol	Money

*Sec : Security, Sol: Solution, Ent: Enterprise, Tec: Technology, IPr: Information Protection, Mal: Malignity

Table 4. Security/Hacking News Body Keyword Frequency Rank

Year	1st	2nd	3rd	4th	5th
2012	Ahn Lab	Ent	Privacy	Mal	Ent
2013	Sol	Attack	Ser	Mal	Ent
2014	Ent	Sol	Ser	Int	Tec
2015	Ent	Ser	Int	Sol	Tec
2016	Ser	Sol	Inf	Tec	Cyber
2017	Sec	Sol	Cyber	Ransom ware	Cloud
2018	Sec	Ent	Inf	Tec	Ent

*Sec : Security, Sol: Solution, Ser: Service, Ent: Enterprise, Tec: Technology, Inf: Information, IPr: Information Protection, , Mal: Malignity, Int: Internet

수집한 보안/해킹 뉴스 기사를 가지고 연도별 키워드 빈도수를 뉴스의 제목과 본문으로 분리하여 각 빈도수를 추출하였다. Table 3를 보면 '보안'이라는 단어 빈도수가 가장 많다. 그 다음 순위로 많이 나오는게 '솔루션' 단어이다. 보안 분야에서 솔루션에 대한 관심이 상당히 높다는 것을 확인 할 수 있다. 그 다음으로 5위까지 나오는 단어는 '정보보호', '사이버', '클라우드', '가상', '기업'이다. 기업이 사이버영역인 가상화 공간인 클라우드를 고려하고 정보보호에 힘쓴다는 걸 알 수 있다. 2017년 랜섬웨어 공격으로 전국민을 긴장하게 하였고, 2018년도 가상화폐로 '랜섬웨어'와 '블록체인' 단어가 상위권에 진입하였다. Table 4에서 보는 바와 같이 본문에서는 1위는 2017년도부터 '보안'이지만 그 전에는 '안랩', '솔루션', '기업', '서비스'이다. 2012년도 최다 빈도수 단어는 '안랩'으로 보안 솔루션 분야에서 안랩이 차지하는 비중 및 기업에 미치는 영향도가 크다는 것을 알 수 있다. 5위까지 상위권 단어는 보면 '개인정보보호', '악성', '기술', '인터넷', '랜섬웨어', '클라우드', '정보', '사이버 공격'이다.

2.3.3 보안솔루션 그룹별 키워드 수집량 분석

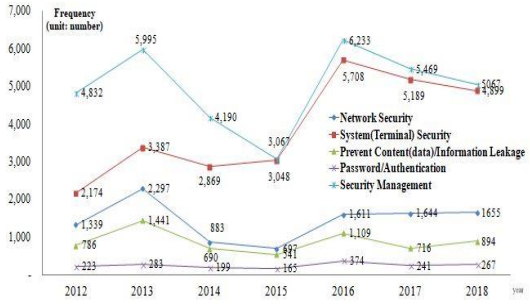

Fig. 4. Frequency by Monthly Information Security Solution

5가지 보안 솔루션 분류에 대한 키워드 빈도수를 분석한 결과는 Fig. 4와 같다.

네트워크 보안 솔루션의 경우 1000~2000개의 빈도수를 보이며, 2013년도에 빈도수가 상승함을 확인할 수 있다. 이중 빈도수를 가장 많이 차지하는 단어는 디도스(DDoS), 방화벽, IPS, NAC, VPN 등이 있다. 2013년의 주요 키워드는 디도스, 2016년도에는 차세대방화벽(NGFW)이 상당수의 빈도수를 보였다. 과거 방화벽과 IPS와 같은 각각의 기능 영역이 확실한 솔루션들이 활용했던거와는 반대로 최근 추세를 보면 차세대방화벽이라는 진화된 솔루션으로써 방화벽, IPS, VPN, 안티바이러스, URL, 필터링 등 네트워크 보안에 필요한 모든 기능이 통합되고 있으며, 차세대 네트워크 보안 플랫폼으로 확장되고 있다[17,18].

시스템(단말) 보안 솔루션 관련 단어는 약 1000개의 빈도수를 보였으며, 2013년과 2016년도에 약간 상승세를 보이고 있다. 관련 단어는 APT, 스팸, 안티멀웨이, 모바일 보안이 많은 빈도수를 보인다. 2013년도에 특히 APT공격 관련 기사가 많았는데 이는 2011년 4월 농협 전산망 마비사고, 7월 SK커뮤니케이션즈의 3500만여건의 고객 개인정보 유출로 시작된 APT공격이 점차 진화되어 2013년도 KBS, MBC, YTN, 농협, 신한은행, 제주은행 등 방송·금융 6개사의 전산망이 마비되는 3·20 사이버테러 발생하였기 때문이다. 그 후 전산망 안전성 확보를 위해 많은 기관 및 회사에서 망분리 사업이 추진되었다[19]. 그러나 시스코가 2017년도에 발표한 '연례 사이버보안 보고서'에 따르면, 물리적 망분리 후에도 스팸메일을 통하여 내부 시스템에 침입 하였고, 2016년 조사된 바에 따르면 전체 이메일의 65%를 스팸메일이 차지하며 2010년 이후 최고 수준을 보였다고 한다[20,21].

이 시점을 기점으로 국내에서는 불필요하고 무작위적인 메일들을 차단하기 위한 '안티 스팸(Anti-Spam)'기능, 메일 첨부파일에 대한 바이러스 감염 여부 식별과 차단/제거를 위한 '안티 바이러스(Anti-Virus)'기능, 첨부파일 중 악의적인 목적으로 제작된 해킹 파일이 유입되는 것을 방지하기 위한 '안티 멀웨이(Anti-Malware)'기능을 보완하기 시작한다.

보안관리 솔루션 관련 단어는 2016년도에 급격하게 증가하는 것을 확인하였으며, 취약점분석, TMS, PMS, 포렌식, 로그분석 등의 단어들이 주를 이루었다. 2016년 한 해 동안 총 6,400개가 넘는 취약점이 발견 되었다. 대표적인 취약점 공격으로 시스템의 모든 권한을 장악 할 수 있는 Dirty Cow취약점 (CVE-2016-5195), PHP에서 광범위하게 사용되는 이메일 발송 라이브러리를 공격하는 PHPMailer RCE 취약점 (CVE-2016-100 33), ImageMagick이 https형식의 문서를 처리하는 과정에서 임의의 명령을 실행 할 수 있는 Image Tragick 취약점 (CVE-2016-3714) 등 많은 기관 및 회사들은 2016년도부터 시스템 보안 취약점 패치 관리에 노력하고 있다[22].

암호/인증 관련 솔루션 단어는 2012년부터 2018년도까지 솔루션 단어 빈도수가 많지 않다. 솔루션 단어 빈도수를 가장 많이 차지하는 단어는 OTP, SSO, PKI 등이 있었다. 대부분의 단어들이 보안사고에 직접적인 영향을 받는 단어들이 아니고, 2차 보안 방어 및 보안 관리에 주로 사용하는 단어들로 뉴스에서 단어 빈도수가 적다.

콘텐츠(데이터)/정보 유출 방지 솔루션 관련 단어는 2013년과 2016년도에 높은 단어 빈도수를 보이고 있다. 솔루션 단어 빈도수를 가장 많이 차지하는 단어는 DLP, DRM, PIMS, DB보안 등이 있다. DLP(Data Loss Prevention) 기술이 2013년도부터 보안 이슈로 부상했다. 잊을만 하면 발생하는 개인정보 유출 사고로 인해 기업의 중요정보 및 개인정보 보호를 위한 솔루션에 대한 관심이 2016년을 기점으로 증가하기 시작하였다. 개인정보 유출과 같은 정보보안사고가 기업가치에 부정정인 영향을 미치고, 정보보안사고 발생 시 그 사고 유출규모에 따라 천문학인 피해보상을 해야 하는 경우가 발생하기 때문에 기업에서는 상황에 맞는 보안 솔루션 도입이 필요하다[23].

2.3.4 보안 솔루션별 키워드 수집량 분석

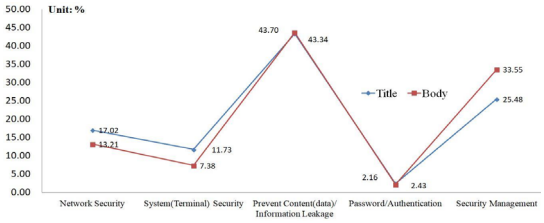

Fig. 5. Security Solution Articles (Title/Body) ratio analysis

보안 솔루션별 기사 제목(Title)를 비율로 분석해 보면 콘텐츠(데이터)/정보유출방지 보안(43.34%), 보안관리(25.48%), 네트워크 보안(17.02%), 시스템(단말) 보안(11.73%), 암호/인증(2.43%) 순으로 나타났다. 본문(Body)을 비율로 분석해 보면 콘텐츠(데이터)/정보유출방지 보안(43.70%), 보안관리(33.55%), 네트워크 보안(13.21%), 시스템(단말) 보안(7.38%), 암호/인증(2.16%) 순으로 나타났다. Fig. 5에서 보는 바와 같이 정보보안 솔루션별 기사(제목/본문)를 비율로 분석 해 본 결과 제목과 본문의 점유율이 비슷함을 확인할 수 있다. 콘텐츠(데이터)/정보유출방지 보안은 43%, 암호/인증은 2%데로 거의 똑같고 나머지 분야도 크게 벗어나지 않는다. 이처럼 기사의 제목을 참고하면 본문의 내용에 대한 분석이 가능하다는 점을 알 수 있다. 세부적인 정보보안 솔루션 빈도수 순위는 Table 5와 Table 6에서 보는 바와 같이 확인 할 수 있다.

Table 5. Security solution title Frequency ranking

year	1st	2nd	3rd	4th	5th
2012	IPMS	Cipher	APTR	TMS	DDoS
2013	IPMS	APTR	TMS	DDoS	Firewall
2014	TMS	IPMS	VA	APTR	Firewall
2015	TMS	VA	APTR	IPMS	DDoS
2016	TMS	IPMS	VA	DDoS	APTR
2017	TMS	IPMS	VA	DDoS	Firewall
2018	TMS	IPMS	VA	DDoS	Firewall

*IPMS: Individual Privacy management system, TMS: Threat Management system, VA: Vulnerability Analysis, APTR: APT Response

Table 6. Security solution body Frequency ranking

year	1st	2nd	3rd	4th	5th
2012	IPMS	TMS	VA	APTR	DDoS
2013	TMS	IPMS	APTR	VA	DDoS
2014	TMS	VA	IPMS	APTR	DDoS
2015	TMS	VA	IPMS	APTR	Firewall
2016	TMS	VA	IPMS	APTR	DDoS
2017	TMS	VA	IPMS	DDoS	Firewall
2018	TMS	VA	IPMS	DDoS	Firewall

2.3.5 보안 솔루션별 빈도수(연도별)

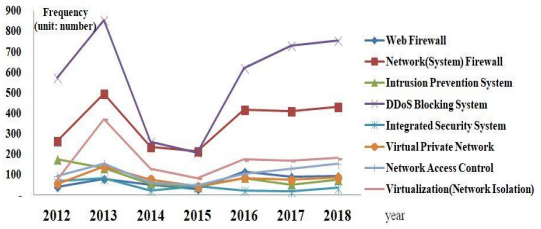

Fig. 6. Frequency of network security solutions

Fig. 6에서 보는 바와 같이 네트워크 보안 솔루션 중 웹방화벽, 침입방지시스템(IPS), 통합보안시스템(UTM), 가상사설망(VPN), 네트워크접근제어(NAC), 가상화(망분리)는 약 200개 이내에서 빈도수를 유지하고 있지만, 네트워크(시스템)방화벽과 DDoS차단 시스템은 2013년도 급증과 2014년도 감소 후 다시 2015년도부터 급증하고 있다.

Fig. 7에서 보는 바와 같이 시스템(단말) 보안 솔루션 중 Anti 멀웨어, 스팸차단 SW, 모바일 보안은 약 200개 이내에서 빈도수를 유지하고 있지만, APT 대응은 2013년도와 2016년에 급증하였다.

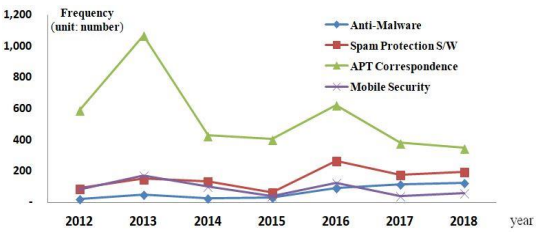

Fig. 7. Frequency of system(terminal) security solution

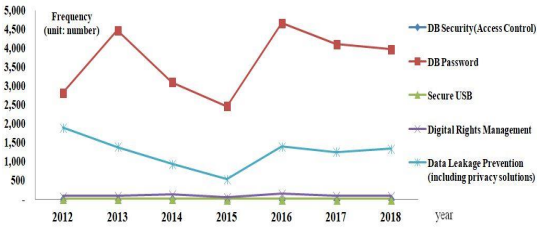

Fig. 8. Frequency of content(data) / information leakage prevention solution

Fig. 8에서 보는 바와 같이 콘텐츠(데이터)/정보유출방지 솔루션 중 DB보안(접근통제), 디지털저작권관리(DRM), 보안USB는 약 100개 이내에서 빈도수를 유지하고 있다. DB암호는 2013년도와 2016년도 상승 후 다시 하강 곡선을 그리고 있고, 데이터유출방지(DLP)는 2015년까지 하락 후 2016년도 상승 하였지만 2017년는 다시 하락 추세이다.

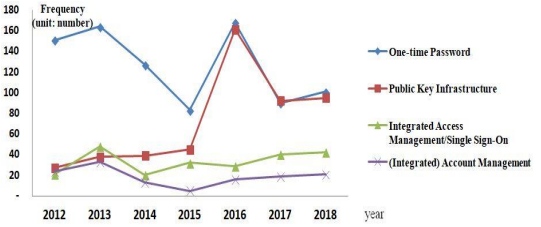

Fig. 9. Frequency of password/authentication solutions

Fig. 9에서 보는 바와 같이 암호/인증 솔루션 중 통합접근관리(EAM)/싱글사이온(SSO)과 통합계정관리(IM/IAM)는 약 40개 이내에서 빈도수를 유지하고 있지만, 공개키기반구조(PKI)는 2015년도 까지 꾸준히 증가 후 2016년도 급증하였다가 2017년도 하락하는 모습니다. 일회용비밀번호(OTP)는 2013년도부터 2015년도까지 하락 곡선을 보인 후 2016년도에 급증하였지만 2017년도에 다시 급하락 하였다.

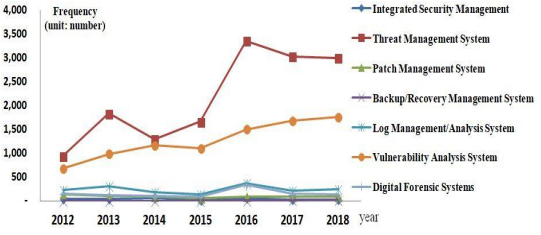

Fig. 10. Frequency of security management solution

Fig. 10에서 보는 바와 같이 보안관리 솔루션 중 통합관리보안(ESM), 패치관리시스템(PMS), 백업/복구관리시스템, 로그관리/분석시스템, 디지털포렌식 시스템은 약 400개 이내에서 빈도수를 유지하고 있지만, 취약점분석시스템(VAS)은 꾸준한 증가를 보이고 있다. 위협관리시스템(TMS)도 2013년도와 2016년도에 상승점이 있지만 전반적으로 증가 추세이다.

2.3.6 보안 솔루션 매출액-기사 빈도수 비율 분석

정보보안 솔루션 매출액 비율과 보안 솔루션 단어 빈도수 비율을 확인 결과 Table 7에서 보는 바와 같다.

Table 7. Information security solution sales - Solution word frequency

Year / Information / Security Solution(%)	'12	'13	'14	'15	'16	'17	'18
Network security sales	40.9	39.2	37.4	37.0	35.9	32.1	31.9
Number of network security words	14.3	17.1	10.0	9.3	9.1	9.0	8.9
System Security sales	14.7	18.6	16.5	16.9	17.9	21.9	22.1
System security word count	23.2	25.3	32.5	40.5	38.0	39.1	39.3
Security Management Sales	10.3	8.5	14.7	14.1	13.7	15.7	16.3
Security management word count	51.7	44.7	47.4	40.8	41.5	41.2	39.6
Content information leakage prevention Sales	24	23	24	25	26	25	24
Content information leak prevention word count	52	45	47	41	41	41	43
Password/ Certification sales	10	11	8	7	7	7	6
Password/ Verification Words	2	2	2	2	2	2	2

5가지 보안 솔루션 분야 중 네트워크 보안, 시스템(단말) 보안, 보안관리 분야는 빈도수와 향후 추이 동향 분석이 가능하였지만, 콘텐츠(데이터)/정보유출방지 보안와 암호/인증 솔루션 분야는 단어수와 매출액 비율의 정보를 가지고 추이를 분석하기에는 자료가 충분하지 않았다.

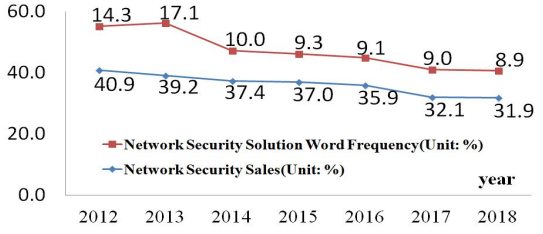

Fig. 11. Security Solutions Sales and Word Frequency (Network Security)

Fig. 11에서 보는 바와 같이 네트워크 보안 솔루션 단어수 비율은 2018년까지 꾸준히 감소하고, 매출액 비율도 단어수와 비례해서 감소한다. Fig. 5에서 보는 바와 같이 2015년도 망분리 사업 시점을 기준으로 네트워크 단어의 빈도수는 증가하였으나, 다른 분야의 솔루션 도입의 증가로 매출액 비율에는 큰 영향을 주지 않았서 시장에서 네트워크 보안의 관심도는 감소중이다. 2018년도 이후 단어수 비율이 감소 될 것으로 예상되고, 이와 함께 매출액 비율도 함께 감소 할 거라고 예상 할 수 있다.

Fig. 12에서 보는 바와 같이 시스템(단말) 보안 단어수 비율은 2015년도 까지 꾸준히 상승 후 2016년도 부터는 일정하게 유지중이다. 보안 솔루션 매출액 비율은 2013년도 상승 후 2015년도 하락을 겪었지만 2018년도까지 소폭 상승하고 있다. 2018년도 이후는 단어수 비율과 매출액 비율 비율은 일정하게 유지 할 것으로 예상된다.

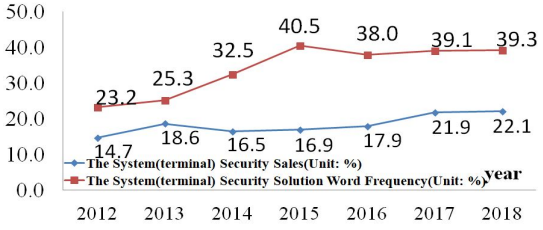

Fig. 12. Security Solutions Sales and Word Frequency (System (Terminal) Security)

네트워크 보안은 이미 강화가 많이 되어 있고 최근 인적사고 발생 비율이 높아지고 있는 상황으로 Fig. 6에서 보는 바와 같이 시스템(단말) 보안 솔루션 중 Anti 멀웨어, 스팸차단 소프트웨어 보안을 더 강화해야 한다.

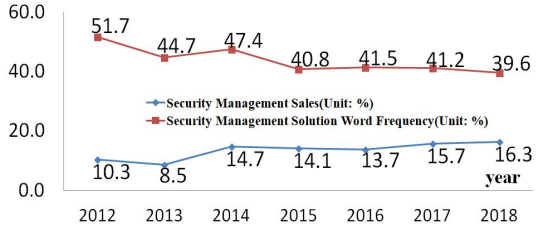

Fig. 13. Security Solutions Sales and Word Frequency (Security management)

Fig. 13에서 보는 바와 같이 에서 보안관리 보안 솔루션 단어수는 2014년도 기점으로 꾸준히 감소 중이다. 매출액 비율은 2018년도까지 조금씩 상승하였다. 2018년도 이후 단어수 비율은 소폭 감소 될 것으로 예상하고, 매출액 비율은 상승 될 것으로 예상된다. Fig. 9에서 보는 바와 같이 보안관리 솔루션 중 취약점분석시스템(VAS)과 위협관리시스템(TMS) 보완에 더 많은 관심이 있을 것이라 판단한다.

3. 고찰

본 논문은 R 프로그램을 통하여 2012년 5월부터 2018년 12월까지의 네이버 보안/해킹 기사를 수집하고 텍스트마이닝하였다. 연도별 보안/해킹 관련 기사 건수, 월별 기사 건수, 종류별 보안 솔루션의 빈도수, 보안 솔루션별 빈도수를 제목과 본문 비율의 Chart 분석 진행하였고, 그 결과를 종합적으로 정리함으로서 보안 솔루션 시장의 과거부터 현재까지의 동향과 추후에 어떠한 함의를 가질 수 있는가를 알아 알아보았다.

24시간 사이버공격에 노출되어 있는 현재에 부족한 인원을 가지고 사이버보안 업무를 수행하기 위해서는 적절한 보안 솔루션 사용이 필요하다. 보안 솔루션 시장 동향을 보면 솔루션들은 점차 지능화되고 단일 기술과 기능의 보안 솔루션이 아니라 다각적인 통합 보안 솔루션으로 발전하고 있다. 앞으로 단순 예방과 탐지만으로는 늘어나는 침해사고에 대처하기는 어렵게 될 것이다. 이제는 방어적인 차원의 보안 솔루션에서 능동적인 보안 솔루션 개발에 힘써야 한다.

본 논문에서 소개한 분석 방법 외에 추가적 연구가 필요하다. 사이버공격은 전 세계적으로 이슈가 되는 상황으로 국내 정보보안 솔루션 시장분석으로 제한하였다는 점과 기술보안 분야 뿐만 아니라 보안 분야를 확장 시켜 물리적 보안과 관리보안 분야도 포함 시켜 고려해야 한다. 그리고 보안 솔루션 시장 동향을 분석 결과를 가지고 보안 담당자는 향후 보안 트렌드를 파악 후 솔루션 구매 대상 선정에 도움을 받을 수 있다. 보안 업체는 시장 대응 및 신제품 개발 시 정보 활용이 가능하다.

References

[1] Defense Agency for Technology and Quality,“Future defense technology linked with the fourth industrial revolution”, 2017

[2] S. J. Yun,“[Military china - Yoon Seok-joon's 'Chaimil'] Already started US-China version "US naval ship accident, is not it a coincidence?”, 2017, [cited 2017 November 24], Avaliable From: https://blog.naver.com/china_lab/221147405605

[3] S. W. Jang,“President Trump's signature 'Block Chain Cyber Security Research' bill”, 2017,[cited 2017 December 14], Avaliable From: http://bigdata.getnews.co.kr/view.php?ud=20171214100217418 4d48e16fff2_23

[4] Blue house,“Enhanced capacity to respond to asymmetric threats such as North Korea's Nuclear Issue No.85 (Ministry of National Defense)”, 2017, Avaliable From: http://www1.president.go.kr/governmentprojects

[5] Korea Information Security Industry Association, “National Survey on Information Protection Industry”, 2017.

[6] Y. H. Noh, T. Y. Kim, D. K. Jeong, K. H. Lee, "Trend Analysis of Convergence Research based on Social Big Data" JOURNAL OF THE KOREA CONTENTS ASSOCIATION, Vol. 19, No. 2, pp. 135-146, Feb, 2019. DOI: https://10.5392/JKCA.2019.19.02.135

[7] M. Song, "Technology Trend Analysis Using Text Mining: Newspaper Article Data," Proceedings of the Korea Information Management Society, pp. 9-30, Nov, 2017.

[8] J. H. Kim, S. H. Kim, I. K. Cho, H. S. Cho, B. K. Noh, "Cyber Security Technology Trends Using Big Data", ETRI, pp. 19-29, 28, 2013.

[9] J.A. Park, Im, H. B. Im, "Understanding Customer's Mind using Big Data Analytics," Korean Psychological Association Conference, pp. 200-200, Aug, 2018.

[10] M. J. Kim, C. J. Kim, "Trend Analysis of News Articles Regarding Sungnyemun Gate using Text Mining," JOURNAL OF THE KOREA CONTENTS ASSOCIATION, Vol. 17, No. 3, pp. 474~485, Mar, 2017. DOI: https://doi.org/10.5392/JKCA.2017.17.03.474

[11] INTERNET TREND, May 3, 2012 - December 31, 2018, Avaliable From: http://trend.logger.co.kr/trendForward.tsp

[12] Naver,“Security/Hacking News”, May 3, 2012 - December 31, 2018, Avaliable From: http://news.naver.com/main/list.nhn?mode=LS2D&mid=shm&sid1=105&sid2=732

[13] Korea Information Security Industry Association, “National Survey on Information Protection Industry”, 2018.

[14] C. Y. Park, Naver News Crawlong, N2H4 R package, Avaliable From: https://github.com/mrchy park/N2H4

[15] A. Alexandra, C. Paulo, R. Paulo, M. Sergio, “Research trends on Big Data in Marketing: A text mining and topicmodeling based literature analysis”, Vol. 24, Issue 1, pp. 1-7, 2018

[16] "K-ICT Strategy 2016", July 1, 2016, Avaliable From: https://blog.naver.com/0mikyoung0/220750945 661

[17] S. J. Kim ,S. W. Lee, 2018, "Social Engineering based Security Requirements Recommendation Framework to Prevent an Advanced Persistent Threat," Journal of KIISE, Vol. 45, No. 10, pp. 1015-1028. DOI: http://dx.doi.org/18-510

[18] B. H. Kim, D. W. Cho, "Network security technology trends and prospects", The Journal of The Korean Institute of Communication Sciences, Vol. 31, No. 4, pp. 99~106, 2014

[19] Digital Story,“SK Communications Nate Hack, 2011”,

2011, Avaliable From: https://blog.naver. com /hahaseyo/10115157336

[20] Cisco Korea Blog,"How many spam do you receive?",February 28, 2017, Avaliable From: http://www.ciscokrblog.com/1169

[21] National Institute of Standards and Technology, Framework for Improving Critical Infrastructure Cybersecurity, 2017

[22] "Top 10 vulnerabilities in 2016", January 24, 2017, Avaliable From: http://blog.alyac.co.kr/942

[23] J. H. Eom, M. J. Kim, 2016, "Effect of Information Security Incident on Outcome of Investment by Type of Investors: Case of Personal Information Leakage Incident," Journal of the Korea Institute of Information Security & Cryptology, Vol. 26, No. 2, pp. 463~474. DOI: http://dx.doi.org/10.13089/JKIISC.2016.26.2.463

박 상 천(Sang-Cheon Park) [정회원]

- 2018년 8월 : 경상대학교 경영정보학과 (경영학석사)
- 2013년 12월 ~ 현재 : 국방기술품질원 연구원

〈관심분야〉
정보보안, 정보경영

박 동 수(Dong-Soo Park) [정회원]

- 2013년 2월 : 숭실대학교 컴퓨터학부(공학사)
- 2018년 2월 : 경상대학교 경영정보학과 (경영학석사)
- 2012년 12월 ~ 현재 : 국방기술품질원 선임연구원

〈관심분야〉
정보보안, 정보경영, 품질경영