

빅데이터 기반의 사이버위험 측정 방법 및 사이버사고 예측모형 연구

2018. 9

이진무

머 리 말

인공지능, 빅데이터, 사물인터넷 등으로 대표되는 정보통신기술의 발전은 개인의 일상생활, 기업활동, 정부 행정기능 등 사회전반적으로 긍정적인 영향을 끼치고 있다. 하지만, 기술에 대한 의존도가 증가함에 따라 이를 악용하는 사이버공격도 급속하게 증가하고 있다. 이러한 사이버위험에 대한 적절한 대응의 부재 시 4차 산업혁명으로의 지속적 발전에 방해요소로 작용될 수 있다. 사이버보험이 디지털 경제의 피해구제를 위한 대안으로 제시되고 있는 시점에서, 시장 활성화를 위한 연구가 필요하다.

본 보고서는 사이버보험에 대한 이해를 높이기 위한 연구의 결과이다. 먼저, 사이버 위험 및 사이버보험에 대한 미국 및 유럽의 현황을 살펴본다. 또한, 사이버공격에 대한 빅데이터를 수집하고 분석함으로써 사이버위험 측정 방법론을 제시하고, 이를 국내기업 및 기관들에 적용하여 사이버위험에 대한 상황을 점검한다. 사이버위험에 대한 정보공시 정책이 해당 기관에 어떠한 영향을 끼치는지 필드실험을 통해 경험적으로 알아본다. 마지막으로, 사이버보험료의 정확한 측정을 위한 기초자료로서, 기계학습을 이용한 사이버사고 예측모형을 개발한다. 이는 빅데이터 기반의 사이버보험 데이터 표준체계 구축의 가능성을 보여준다.

본 보고서를 통하여 사이버보험에 대한 심도깊은 논의가 촉발되기를 기대한다.

마지막으로 본 보고서의 내용은 연구자 개인의 의견이며, 우리 원의 공식적인 의견이 아님을 밝혀둔다.

2018년 9월

보 험 연 구 원

원장 한 기 정

■ 목차

요약 / 1

I. 서론 / 6

1. 연구배경 및 목적 / 6

2. 연구범위 / 8

II. 사이버위협 및 사이버보험 현황 / 9

1. 사이버위협 및 인식 현황조사 / 9

2. 사이버보험 및 인식 현황조사 / 16

3. 소결 / 23

III. 사이버위협 측정체계 개발 / 24

1. 사이버위협 관련 개념 및 데이터 종류 / 24

2. 사이버위협 빅데이터 수집 및 처리 시스템 / 26

3. 사이버위협 공시의 효과 측정 실험 / 40

4. 소결 / 48

IV. 사이버사고 예측모델 개발 / 50

1. 기계학습 개념 및 예측 문제 정의 / 50

2. 데이터 수집 / 52

3. 예측 알고리즘 및 모델 평가 / 54

4. 소결 / 57

V. 결론 / 58

| 참고문헌 | / 59

■ 표 차례

〈표 II-1〉 미국 상장사들이 인식하는 사이버위험 키워드 / 14

〈표 III-1〉 아시아 6개국 기관 통계 / 42

〈표 III-2〉 실험 데이터 통계 / 45

〈표 III-3〉 실험효과 측정(이메일 수신 기관 기준) / 46

〈표 III-4〉 실험효과 측정(실험 전 사이버위험 감지된 기관 기준) / 47

〈표 III-5〉 실험효과 측정(이메일 열람한 기관 기준) / 47

〈표 III-6〉 실험효과 측정(실험 웹사이트 방문한 기관 기준) / 47

〈표 IV-1〉 예측모델 성능비교 / 57

■ 그림 차례

- 〈그림 II-1〉 정보유출사고 건수 및 유출량 추이(2005~2017년) / 10
- 〈그림 II-2〉 국가별 기업평균 사이버공격 피해금액(2017년 기준) / 11
- 〈그림 II-3〉 미국 상장사의 공시 데이터 수집 흐름도 / 12
- 〈그림 II-4〉 Latent Dirichlet Allocation 토픽모델링 개념 / 13
- 〈그림 II-5〉 미국 상장사의 위험 인식 키워드 / 14
- 〈그림 II-6〉 미국 상장사가 인식하는 사이버위험 언급 분포의 증가 / 15
- 〈그림 II-7〉 미국 상장사가 인식하는 IT 관련 위험 분포의 증가 / 15
- 〈그림 II-8〉 국가별 사이버보험 보험료 분포(2016년 기준) / 16
- 〈그림 II-9〉 전세계 시장의 연간 사이버보험료 추정액 추이 / 17
- 〈그림 II-10〉 사이버보험 보험료 기준 상위 보험사 리스트(2016년 기준) / 18
- 〈그림 II-11〉 사이버공격 종류별 보험 청구 통계 / 19
- 〈그림 II-12〉 사이버보험의 보장 종류 / 20
- 〈그림 II-13〉 사이버보험 보험 청구(2016년 기준) / 20
- 〈그림 II-14〉 사이버보험 구매의향 설문조사 결과(2011~2015년) / 21
- 〈그림 II-15〉 사이버보험 비구매 사유(2016년 기준) / 22
- 〈그림 II-16〉 사이버보험 가입거부 사유(2016년 기준) / 22
- 〈그림 III-1〉 봇넷의 개념도 / 25
- 〈그림 III-2〉 사이버위험 데이터 수집 시스템 구조 / 27
- 〈그림 III-3〉 인터넷 관리 기관 / 29
- 〈그림 III-4〉 인터넷 주소를 기관으로 매핑하는 과정 / 30
- 〈그림 III-5〉 한국표준산업분류(2017년 1월 10차 개정 기준) / 32
- 〈그림 III-6〉 한국 상장사들의 산업분류 분포(2018년 기준) / 33
- 〈그림 III-7〉 인터넷 주소 할당 받은 한국기관의 산업분류 분포(2018년 기준) / 34
- 〈그림 III-8〉 한국 사이버위험 랭킹 - 전체(2018년 5월 기준) / 35
- 〈그림 III-9〉 한국 사이버위험 랭킹 - 교육기관(2018년 5월 기준) / 36
- 〈그림 III-10〉 한국 사이버위험 랭킹 - 정부기관(2018년 5월 기준) / 37
- 〈그림 III-11〉 한국 사이버위험 랭킹 - 비영리기관(2018년 5월 기준) / 38
- 〈그림 III-12〉 한국 사이버위험 랭킹 - 제조업(2018년 5월 기준) / 38
- 〈그림 III-13〉 한국 사이버위험 랭킹 - 정보통신업(2018년 5월 기준) / 39

■ 그림 차례

〈그림 Ⅲ-14〉 한국 사이버위험 랭킹 - 금융/보험업(2018년 5월 기준) / 40

〈그림 Ⅲ-15〉 실험 설계 / 41

〈그림 Ⅲ-16〉 사이버위험 지수 공유 이메일 샘플 / 43

〈그림 Ⅲ-17〉 사이버위험 자료공유 웹사이트 - 메인페이지 / 44

〈그림 Ⅲ-18〉 사이버위험 자료공유 웹사이트 - 개별페이지 샘플(Alibaba) / 44

〈그림 Ⅲ-19〉 기관 반응별 실험효과 비교 / 48

Developing Cyber Risk Assessment Framework for Cyber Insurance: A Big Data Approach

As our society is heavily depending on the information and communication technology, the associated risk has also significantly increased. Cyber insurance has been emerged as a possible means to better manage such cyber risk. However, cyber insurance market is still in a premature stage due to the lack of data sharing and standards on cyber risk and cyber insurance. To address this issue, this research proposes a data-driven framework to assess cyber risk using externally observable cyber attack data sources such as outbound spam and phishing websites. We show that the feasibility of such approach by building cyber risk assessment reports for Korean organizations. Then, by conducting a large-scale randomized field experiment, we measure the causal effect of cyber risk disclosure on organizational security levels. Finally, we develop machine-learning models to predict data breach incidents, as a case of cyber incidents, using the developed cyber risk assessment data. We believe that the proposed data-driven methods can be a stepping-stone to enable information transparency in cyber insurance market.

Keywords: cyber risk assessment, cyber attack, cyber insurance, Big Data, machine learning, randomized field experiment

요약

I. 서론

- 최근 증가되고 있는 사이버위협 요소에 대한 관리 방안이 필요함
 - 기술적, 정책적 해결 방안들이 지속적으로 개발되고 있음
 - 동시에 사이버보험을 통한 사이버위협 관리에 대한 수요가 증가함
 - 하지만 사이버보험 관련 데이터 및 표준의 부재가 사이버보험 시장 확대의 방해 요소로 작용함
 - 본 연구에서는 빅데이터 기반의 사이버위협 측정체계 및 사이버사고 예측모델을 개발하고자 함

II. 사이버위협 및 사이버보험 현황

1. 사이버위협 및 인식 현황조사

- 사이버공격으로 인한 사이버위협은 지속적으로 증가하고 있는 추세임
 - 사이버위협은 정보통신시스템 사용에 방해되는 모든 요소를 의미하는데, 발생 원인은 단순실수나 해커의 의도된 공격 등으로 다양함
 - 정보유출과 같은 사이버공격의 피해가 최근 12년간 10배 이상 증가하였고, 사이버공격은 정보통신이 발달한 나라에 집중되고 있음
 - 정보통신시스템은 여러 시스템이 서로 유기적으로 연결되어 동작하는데, 이 때문에 사이버위협은 상호복합성(Interdependency) 및 높은 불확실성 때문에 모델링이 어려움

- 지속되는 사이버공격으로 인해, 기업들의 사이버위험에 대한 인식 또한 증가하고 있음
 - 미국 상장사의 연간보고서에 기재된 위험요소(Risk Factor) 정보를 수집하여 텍스트 마이닝한 결과, 사이버위험에 대한 언급량이 최근 12년간 3배 이상 증가한 것으로 확인됨

2. 사이버보험 및 인식 현황조사

- 사이버보험 시장은 미국 등 일부 국가들 중심으로 시작됨
 - 보험료 기준으로 미국이 전세계 시장의 90%를 차지함
 - 전세계 보험료 추정액은 선형적으로 증가하고 있음
 - 현재 보험상품들은 다양한 사이버위험에 대하여 보장함
 - 해킹에 의한 보험 청구 건이 가장 많이 발생함
- 기업들의 사이버보험 구매의향이 증가하고 있음
 - 구매의향 비율이 2011년 24%에서 2015년 64%로 크게 증가함
 - 수요자 측면에서는 보험상품의 복잡성이 비구매 사유로 거론됨
 - 공급자 측면에서는 잠재 피보험자의 낮은 사이버보안 수준이 가입거부 사유로 거론됨

Ⅲ. 사이버위험 측정체계 개발

1. 사이버위험 관련 개념 및 데이터 종류

- 사이버위험을 발생시키는 다양한 사이버공격에 대한 데이터가 존재함
 - 사이버공격의 종류는 스팸이메일, 피싱공격 등이 있음

- 이러한 사이버공격은 악성코드나 바이러스에 감염된 컴퓨터들의 네트워크 봇넷(Botnet)을 통해 수행됨
- 실제 공격자의 신원을 파악하는 것은 매우 어려움

2. 사이버위협 빅데이터 수집 및 처리 시스템

■ 스팸이메일, 피싱사이트 데이터를 통하여 사이버위협 측정이 가능함

- 스팸이메일은 CBL, PSBL에서, 피싱사이트 데이터는 APWG, 오픈피시를 통하여 수집함
- 클라우드 기반의 데이터 수집 및 처리 시스템을 개발함
- IP 주소별로 구성되어있는 사이버공격 데이터를 기관별로 정리하기 위한 3단계 매핑이 필요함
- APNIC 데이터를 기준으로 인터넷 주소를 할당받은 한국 기관 정보를 수집하고, 도메인주소 정보를 이용하여 산업군별로 분류함
- 한국 기관들의 사이버위협 지수를 측정할 결과, 대학기관, 교육기관, 인터넷 망 사업자들의 사이버위협 지수가 높은 것으로 나타남

3. 사이버위협 공시의 효과 측정 실험

■ 제시된 사이버위협 지수의 효용성을 검증하기 위하여 필드실험을 실시함

- 중국어권 6개국의 1262개 기관을 대상으로 필드실험을 실시함
- 통계적으로 동일한 제어군과 실험군으로 구분하여, 실험군의 기관들에게 이메일 및 웹사이트를 통하여 위험 측정결과를 고지, 공시함
- 사이버위협 자료공시 이후, 실험군 기관들의 스팸이메일 발생량이 통계적으로 유의미한 수준으로 감소함을 확인함
- 제시된 사이버위협 평가자료를 적극적으로 확인한 기관일수록, 높은 수준의 감소량이 측정됨

IV. 사이버사고 예측모델 개발

1. 기계학습 개념 및 예측 문제 정의

- 사이버사고 예측문제는 감독기계학습(Supervised Machine Learning) 중 분류문제에 해당됨
 - 빅데이터 수집, 강력해진 계산능력 등 여러 요소들로 인하여 기계학습 기반의 예측이 정확해지고 있음
 - 예측모형을 만들기 위해 데이터유출사고 데이터를 사용하는데, 예측변수가 이산변수이므로, 분류(Classification) 알고리즘을 적용함

2. 데이터 수집

- 기 발생한 사이버사고 데이터 및 사이버위험 지수 데이터를 통합하여 학습데이터(Training Data)를 구축하였음
 - 사이버사고 중 대표적인 종류인 정보유출사고 데이터를 수집함
 - 사이버위험 지수를 바탕으로 기관별 특징변수들을 추출함. 특히 CBL상의 봇넷 정보를 이용함
 - 기관명과 스트링매칭 알고리즘을 이용하여 사이버사고 데이터와 사이버위험 데이터를 통합함

3. 예측 알고리즘 및 모델 평가

- 전통적인 기계학습 및 딥러닝 알고리즘을 통합데이터에 적용함
 - 전통적인 기계학습 알고리즘 중 KNN, Logistic Regression, Decision Tree, Random Forest를 적용함
 - 딥러닝 알고리즘 중 피드포워드 신경망을 적용함

- 최대 94%의 확률로 정보유출사고 여부를 예측할 수 있음
 - 10-fold 교차검증을 통하여 여러 알고리즘들을 평가함
 - 데이터 불균형 문제를 해결하기 위해 Precision, Recall, F1 score 기준으로 예측 정확성을 측정함

I. 서론

1. 연구배경 및 목적

빅데이터, 사물인터넷(IoT: Internet of Things), 인공지능 등으로 대표되는 정보통신기술(ICT: Information and Communication Technology)의 비약적 발전은 개인의 일상생활, 기업의 운영활동, 정부의 행정기능 등 사회전반적으로 큰 영향을 끼치고 있다. ICT의 영향력이 확대됨에 따라, 이를 악용하기 위한 사이버공격(Cyber Attack)이 급속히 증가하고 있고, 이는 막대한 사회적 비용을 발생시키고 있다. 맥아피(McAfee)사의 2007년 보고서에 따르면, 사이버범죄는 1,050억 달러 이상의 비용을 발생시켰고, 이는 불법약물 거래시장의 크기보다 큰 것이다.¹⁾ 맥아피의 최근 2018년 보고서에 따르면, 사이버범죄(Cyber Crime)로 인한 비용이 6,000억 달러 이상이 될 것이라고 예상했다.²⁾³⁾ 이에 대한 적절한 대응이 부재할 경우, 사이버위험은 디지털경제를 기반으로 하는 4차 산업혁명의 방해요소로 작용할 위험성이 있다.

이러한 사이버위험에 대응하기 위하여, 정보보안업체들은 소프트웨어 및 하드웨어 솔루션을 적극적으로 개발하고 있다. 오르비스(Orbis)사가 발행한 사이버보안 시장에 대한 보고서에 따르면, 2021년 사이버보안 시장규모는 1,300억 달러에 이를 것으로 예상되었다.⁴⁾ 정부기관들은 국가정보능력(National Intelligence)을 이용하여 불법적 사이버공격을 시행하는 해커들을 색출하는 노력을 강구하는 한편, 정보보안과 관련된

1) <https://it.slashdot.org/story/07/09/19/036203/cybercrime-now-worth-105-billion-bypasse-drug-trade>

2) <https://www.mcafee.com/enterprise/en-us/assets/reports/rp-mobile-threat-report-2018.pdf>

3) <https://www.mcafee.com/enterprise/en-us/assets/reports/rp-mobile-threat-report-2018.pdf>

4) <http://orbisresearch.com/reports/index/global-cyber-security-market-report-2017>

다양한 정책과 법률을 제안하고 있다. 예를 들어, 미국 행정부는 2013년 오바마 정부에 이어 2017년 트럼프 정부에서 사이버보안 향상을 위한 행정조치(Executive Order)들을 단행하였다.⁵⁾⁶⁾

하지만 기업들의 사이버보안 기술 개발 및 정부들의 정책적 노력에도 불구하고 사이버공격은 지속되고 있다. 이는, 해커들이 이윤추구를 목적으로 기업과 같이 조직화되고, 공격기술을 지속적으로 개발하고 있기 때문이다. 심지어 일부 사이버공격은 국가적 목적을 실현하기 위해 사이버전쟁(Cyber Warfare)의 일환으로 국가에 의해 수행된다.⁷⁾⁸⁾ 이러한 요인들로 인하여 사이버공격은 근본적으로 완전히 제거하기 어렵고, 상시적인 위험요소(Cyber Risk)로서 관리(Management)되어야 한다는 의견이 늘어나고 있다. 이에 사이버보험(Cyber Insurance)을 통한 사이버위험 관리에 대한 필요성이 대두되고 있다.

하지만 사이버보험 시장은 미국 등 일부 국가에서만 시작되었고, 한국 및 다른 국가들에는 시장 활성화가 아직 이루어지지 않은 상태이다. 공급 측면에서 보면, (1) 사이버위험 관련 데이터의 부족, (2) 사이버위험의 독특한 특성, (3) 재앙적 피해가능성 등의 이유로 보험회사들이 적극적으로 사이버위험을 인수하려고 하지 않고 있다. 한편, 수요 측면에서는 (1) 사이버보험 상품의 표준화 결여 및 (2) 사이버보험에 대한 이해 부족으로 인해 사이버보험 구매요인이 적은 상황이다. 이러한 수요 및 공급 측면의 문제점들은 공통적으로 사이버위험에 관한 데이터 부족과 표준화 미비에 기인함을 알 수 있다.

본 연구는 사이버위험 및 사이버보험의 현황에 대해서 살펴보고, 사이버보험 시장 활성화를 위해 빅데이터 기반의 사이버위험 측정 방법론을 제시하고, 제시된 사이버위험 측정결과를 기반으로 사이버사고(Cyber Incident) 예측모형(Prediction Model)을 개발하고자 한다.

5) <https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/executive-order-improving-critical-infrastructure-cybersecurity>

6) <https://www.whitehouse.gov/presidential-actions/presidential-executive-order-strengthening-cybersecurity-federal-networks-critical-infrastructure/>

7) <https://www.rand.org/topics/cyber-warfare.html>

8) <https://www.forbes.com/sites/kateoflahertyuk/2018/05/03/cyber-warfare-the-threat-from-nation-states/#4bdd430d1c78>

2. 연구범위

본 보고서의 연구범위는 다음과 같다. 첫째, 사이버위험 및 사이버보험에 대한 현황을 조사하고자 한다. 스팸이메일(Spam), 피싱공격(Phishing), 디도스(DDoS: Distributed Denial of Service) 등의 여러 사이버공격들의 현황과 그에 따른 피해규모에 대한 통계를 살펴본다. 그리고 미국 상장사들의 연간보고서(10-K Annual Report) 빅데이터를 수집해 기업들이 공시(Disclosure)하고 있는 위험요소들(Risk Factors)에 대한 텍스트를 분석하고, 그중 사이버위험의 비중이 어떻게 변화되었는지 분석한다. 또한, 사이버위험 관리를 위한 사이버보험 시장에 대한 현황을 조사한다.

둘째, 사이버공격 빅데이터를 수집/분석하고, 사이버위험의 측정체계를 제안하여, 사이버보험 데이터 표준화에 기여하고자 한다. 특별히, 여러 사이버공격 중에서 스팸 이메일 전송(Outbound Spam) 데이터를 수집하였다. 사이버공격 데이터는 대부분 IP 주소(Internet Protocol Address) 기준으로 수집이 되는데, 이를 기관별 데이터로 정리하기 위한 3단계 매핑(Mapping) 방법을 제시한다. 또한, 산업군(Industry Sector)별 랭킹을 계산하고, 한국 기관들의 사이버위험 상황을 점검한다. 사이버위험 데이터에 대한 정보공유 및 공시가 어떻게 기관들의 사이버위험 관리에 영향을 주는지 경험적으로 확인하기 위해 실시한 아시아 6개국 필드실험의 결과를 요약한다.

셋째, 사이버보험료의 정확한 측정을 위하여, 사이버위험 측정결과를 기반으로 사이버사고에 대한 예측모델을 개발한다. 사이버사고 중 대표적인 데이터유출사건 243건에 대한 자료를 수집하였고, 이를 사이버위험 측정데이터(CBL 928만건; PSBL 53만건)와 통합하였다. 특별히 CBL 데이터상에 존재하는 봇넷(Botnet) 정보를 이용하여 각 기관의 특징변수(Feature)를 추출(Extraction)하였다. 전통적 기계학습 및 딥러닝 알고리즘들을 10-fold 교차검증(Cross Validation)을 이용하여 검증하였다. 그 결과, 딥러닝(Deep Learning)의 한 방식인 전방전달 신경망(Feed Forward Neural Network)이 F1 score 기준으로 94.73%의 확률로 기업의 데이터유출사고를 예측할 수 있음을 확인하였다. 이 결과를 통하여 사이버위험 측정체계가 유의미함을 확인하였고, 나아가 빅데이터 기반의 사이버보험 데이터 표준체계 구축의 가능성을 확인하였다.

II. 사이버위험 및 사이버보험 현황

본 장에서는, 사이버위험에 대한 정의를 내리고 사이버위험과 사이버보험에 대한 현황 및 기업인식을 조사한다. 스팸이메일(Spam), 피싱공격(Phishing) 등 사이버공격(Cyber Attack) 현황과 이에 따른 피해금액에 대한 통계를 살펴본다. 또한, 미국 상장사의 연간보고서(10-K Annual Report) 텍스트 데이터를 분석하여, 기업들이 인지하고 공시하고 있는 주요 위험요소들(Risk Factors) 중 사이버위험의 비중이 어떻게 변화되었는지 살펴본다. 마지막으로, 사이버위험 관리를 위한 사이버보험 시장에 대한 현황을 조사한다.

1. 사이버위험 및 인식 현황조사

가. 사이버공격 및 사이버위험 현황

사이버위험 및 사이버보험과 관련된 여러 연구들을 종합적으로 조사한 제네바 보고서(Geneva 2016)는 사이버위험을 아래와 같이 정의하였다.

“Any risk emerging from the use of information and communication technology (ICT) that compromises the confidentiality, availability, or integrity of data or services. The impairment of operational technology (OT) eventually leads to business disruption, (critical) infrastructure breakdown, and physical damage to humans and property.”

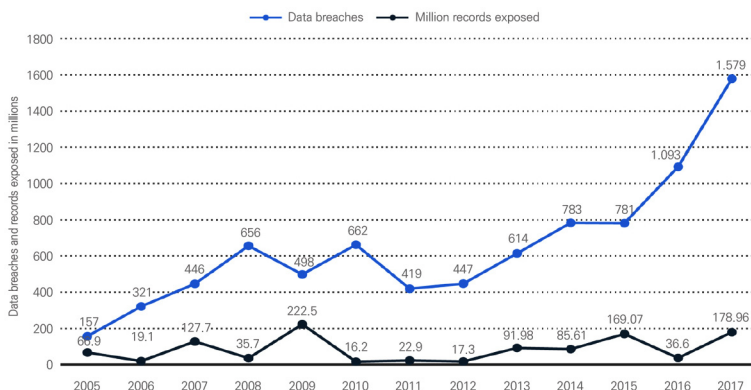
즉, 사이버위험은 정보통신시스템을 사용하는 데에 방해되는 모든 요소를 포괄하는

개념이다. 이러한 사이버위험은 정보통신기술의 필수적인 요소인 기밀유지(Confidentiality), 유효성(Availability), 무결성(Integrity)에 대한 문제를 발생시킨다. 정보통신시스템에 발생하는 결함은 경영활동 중단, 중요 인프라서비스의 중단, 심지어 인간과 재산에 대한 물리적인 손해까지 유발할 수 있다.

사이버위험은 유발요인에 따라 자연유발적(Natural) 사이버위험과 인재로 인한(Man-Made) 사이버위험으로 나눌 수 있다. 인재의 원인은 직원의 단순실수, 해커들의 불법적 사이버공격, 국가의 사이버전쟁, 사이버테러 등이 다양하게 존재한다. 인터넷기반 시스템은 여러 시스템의 상호작용을 통하여 동작하는데, 이로 인하여 사이버위험은 상호복합성(Interdependency), 재앙적 사고 가능성, 데이터 및 모델링에 대한 높은 불확실성, 변화의 위험 등의 특징이 존재한다.

사이버위험의 대표적인 종류인 정보유출사고 발생 건수는 지속적으로 증가하고 있는 추세이다. <그림 II-1>은 사이버스카우트(CyberScout)사와 신분도용자료센터(Identity Theft Resource Center)가 2017년에 공동발표한 보고서(ITRC Annual Data Breach Year-End Review)에 기재된 정보유출사고 통계이다. 2005년부터 2017년까지의 데이터를 바탕으로 작성된 이 보고서에 따르면, 정보유출사고의 발생 건수는 2005년 157건에서 2017년 1,579건으로 10배 이상 증가하였다. 흥미롭게도, 유출된 정보의 총 건수는 특별한 증가 추이를 나타내지 않았다.

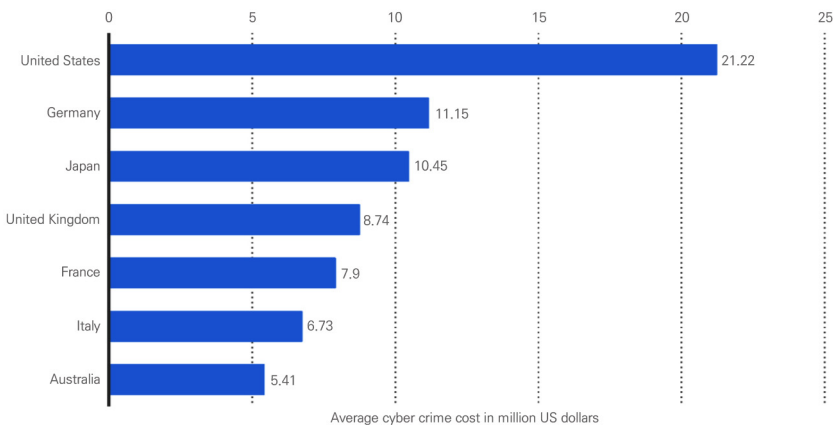
<그림 II-1> 정보유출사고 건수 및 유출량 추이(2005~2017년)



자료: Identity Theft Resource Center, CyberScout

〈그림 II-2〉는 국가별 사이버공격에 의한 피해규모에 대한 조사결과이다. 포넨몬(Ponemon Institute)과 액센추어(Accenture)사 공동으로 254개 기관을 대상으로, 사이버공격으로 인한 피해금액에 대해서 설문조사를 실시하였다. 조사결과에 따르면, 미국 기업은 평균 2,122만 달러의 사이버공격 피해를 입었고, 조사대상 국가 중 가장 큰 피해를 받은 것으로 나타났다. 다음으로는 독일(1,115만 달러), 일본(1,045만 달러), 영국(874만 달러), 프랑스(790만 달러), 이탈리아(673만 달러), 호주(541만 달러) 순으로 나타났다. 이를 통하여 정보통신기술이 발달한 국가에서 사이버공격에 의한 피해금액이 큰 것을 알 수 있다.

〈그림 II-2〉 국가별 기업평균 사이버공격 피해금액(2017년 기준)



자료: Ponemon Institute; Accenture

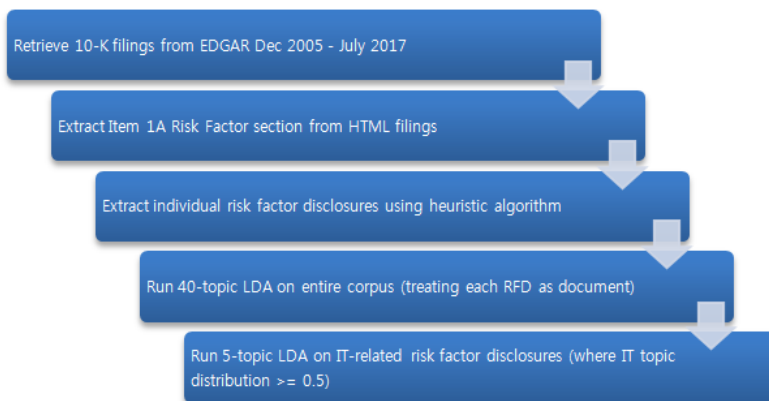
나. 기업의 사이버위험에 대한 인식 조사

이어서, 기업의 사이버위험에 대한 인식을 조사하고자 한다. 정보통신기술 발전을 선도하는 동시에, 사이버공격 또한 가장 많이 발생하는 미국 데이터를 기준으로 분석하였다. 미국 증권거래위원회(SEC: Securities and Exchanges Commission)의 기준에 따라, 미국의 모든 상장사들은 매 회계 분기 및 매 회계 연도 마다 투자자들에게 경영

에 대한 상세한 정보를 제공해야 한다. 특히 2005년 12월부터 연간회계보고서(10-K Annual Report)에 Item 1A가 추가되었는데, 이 부분에 경영과 관련된 모든 위험요소(Risk Factors)들을 기술하도록 되어있다. 사이버공격에 따른 피해가 증가함에 따라, 많은 기업들이 사이버위험에 대한 위험요소 내용을 기술하고 있다. 사이버위험에 대한 기업인식을 정량적으로 측정하기 위해, 2005년 이후에 발행된 모든 연간회계보고서를 수집하고, 위험요소에 대한 텍스트를 분석하였다. 이 부분은 필자가 공저한 Song et al.(2018) 논문에 기반한다.

〈그림 II-3〉은 미국 상장사들의 공시 데이터를 수집하고 분석하는 전체 과정을 설명한 흐름도이다. 1단계로 SEC의 공시시스템인 EDGAR⁹⁾에서 2005년 이후, 6,144개 상장사에서 발간한 26,197건의 연간보고서들을 수집하였다. 2단계로 각 보고서의 위험요소 섹션(Item 1A Risk Factors)에서 173만 건의 개별 위험요소들을 추출하였다. 각 위험요소들이 어떤 의미인지 파악하기 위해, Blei et al.(2003) 논문에서 제안된 Latent Dirichlet Allocation(LDA) 토픽모델링(Topic Modeling)을 이용하였다. LDA에 대한 상세설명은 아래에 기술한다. 추출된 전체 위험요소들 중 2.23%에 해당하는 38,622건의 위험요소들이 사이버위험과 관련이 높은 IT 관련 위험인 것을 나타냈다.

〈그림 II-3〉 미국 상장사의 공시 데이터 수집 흐름도

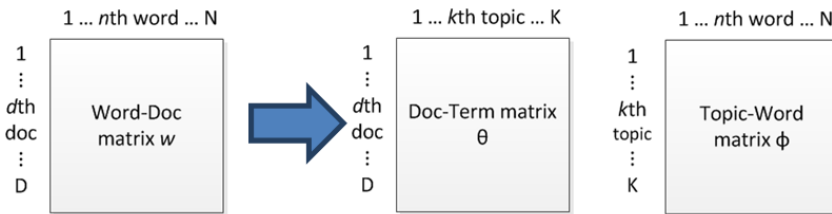


자료: Song et al.(2018)

9) <https://www.sec.gov/edgar/searchedgar/accessing-edgar-data.htm>

LDA는 텍스트 마이닝 기법 중 하나로서, 방대한 문서들 안의 공통 주제들을 자동으로 추출하는 알고리즘이다. <그림 II-4>는 LDA의 개념을 보여준다. 이 알고리즘은 전체 문서들은 K개의 주제(Topic)를 포함하고, 각 문서는 K개의 주제 중 일부를 포함한다고 가정한다. 그리고 하나의 주제는 전체 문서들에 나타난 N개 단어 중 연관단어들의 조합으로 이루어진다고 가정한다. <그림 II-4>와 같이, 행렬연산을 통해서 문서들을 표현할 수 있다고 가정한다. 이때 각 문서에서 나오는 키워드들의 순서는 고려하지 않는 Bag-of-Words 방식을 사용한다. 이 때문에 문서 내의 단어순서는 고려되지 않는 한계점이 있지만, 대용량 문서데이터를 처리하기 위한 확장성(Scalability)을 확보하기 위한 거래(Trade-Off)라고 할 수 있다. LDA는 최근 사회과학연구에 다양하게 적용되고 있다. 예를 들어 필자가 공저한 Shi et al.(2016) 논문에서는, 기업의 비즈니스 소개문서들을 LDA로 분석하여, 인수합병(M&A) 네트워크 형성과정을 설명하였다. 또한, 필자의 Lee et al.(2016) 논문에서는 소셜네트워크상에서 사용자들의 프로필을 LDA 분석하여, 사회관계망 형성 과정을 설명하였다.

<그림 II-4> Latent Dirichlet Allocation 토픽모델링 개념



자료: Song et al.(2018)

<그림 II-5>는 미국 상장사들이 인식하고 있는 위험요소 텍스트에 LDA를 적용했을 때 추출된 주제의 키워드들이다. 각 박스는 하나의 주제를 표현하며, 박스 안에는 연관 단어들이 있다. 이를 통하여 미국 상장사들이 인식하고 있는 다양한 위험요소 주제들을 확인할 수 있다. 거시경제(Interest Rate, Economic Condition), 보험/소송(Liability, Claim, Litigation), 경쟁(New Competitor), 유가 변동(Gas, Oil), 세금 관련(Income Tax), 재무 관련, 지적재산권 관련(Patent, Property, Right) 등 다양한 위험요소들이

있다. 그중 3열 첫 번째 박스에 Information System, Data Security, Service Failure와 같이 사이버위험과 관련된 단어들이 확인되었다.

〈그림 II-5〉 미국 상장사의 위험 인식 키워드



자료: Song et al.(2018); SEC Edgar 10-K Annual Reports

다음으로는, 전체 위험요소들 중에서 사이버위험 관련 요소들을 따로 추출하고, 다시 LDA 분석을 진행하여 상세한 사이버위험 키워드들을 추출하였다. 〈표 II-1〉에서 확인하는 바와 같이, 사이버공격(Cyber Risk, Attacks), 소프트웨어 제품의 결함(Software Product Error), 네트워크 서비스(Service, Network Failures), 시스템 결함(System Failure), 정보유출사고(Reputation, Breach) 등에 대한 기업들의 위험인식을 확인할 수 있다.

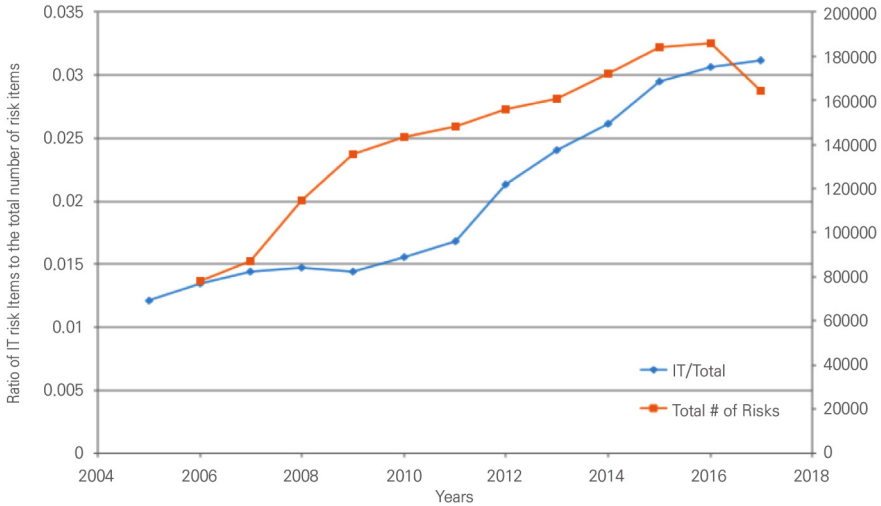
〈표 II-1〉 미국 상장사들이 인식하는 사이버위험 키워드

Topic	Top 10 keywords
1	cyber, including, risk, attacks, risks, operational, threats, from, confidential, networks
2	products, software, errors, defects, reputation, solutions, new, harm, problems, from
3	service, failures, network, failure, interruptions, from, damage, system, ability, interruption
4	failure, results, system, failures, material, company, ability, adverse, effect, its
5	reputation, breach, confidential, measures, access, unauthorized, personal, protect, liability, including

자료: Song et al.(2018); SEC Edgar 10-K Annual Reports

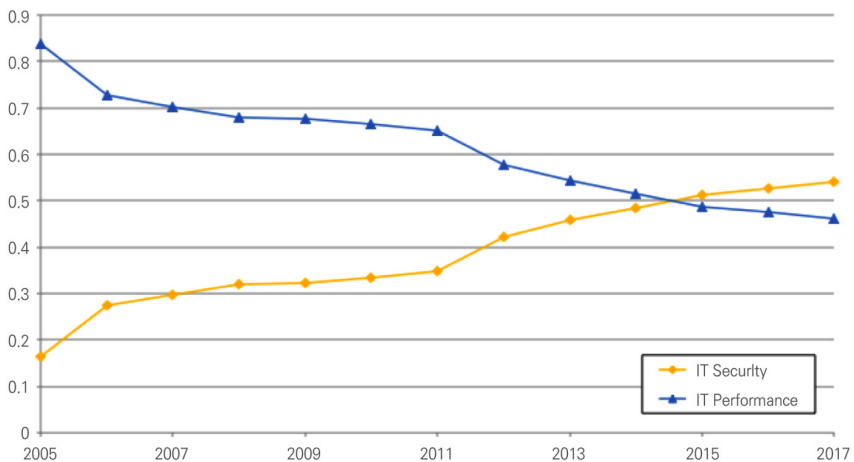
〈그림 II-6〉은 2005년부터 2017년까지 사이버위험에 대한 인식이 증가하고 있음을 보여준다. 매해 언급되었던 모든 위험요소들 중에서 IT와 관련된 위험요소들이 2005년 7만 건에서 2017년 18만 건으로 2배 이상 증가한 것으로 나타났다.

〈그림 II-6〉 미국 상장사가 인식하는 사이버위험 언급 분포의 증가



자료: Song et al.(2018); SEC Edgar 10-K Annual Reports

〈그림 II-7〉 미국 상장사가 인식하는 IT 관련 위험 분포의 증가



자료: Song et al.(2018); SEC Edgar 10-K Annual Reports

IT 관련 위험요소들은 사이버위험(IT Security; Cyber Security)과 IT 성능위험(IT Performance Risk)으로 나눌 수 있다. <그림 II-7>과 같이 IT 위험요소 중 사이버위험이 차지하는 비율이 2005년 15%에서 2017년 55%까지 크게 증가함을 볼 수 있다. 반면, IT 성능 자체에 대한 위험요소들은 크게 줄어든 것을 확인할 수 있다.

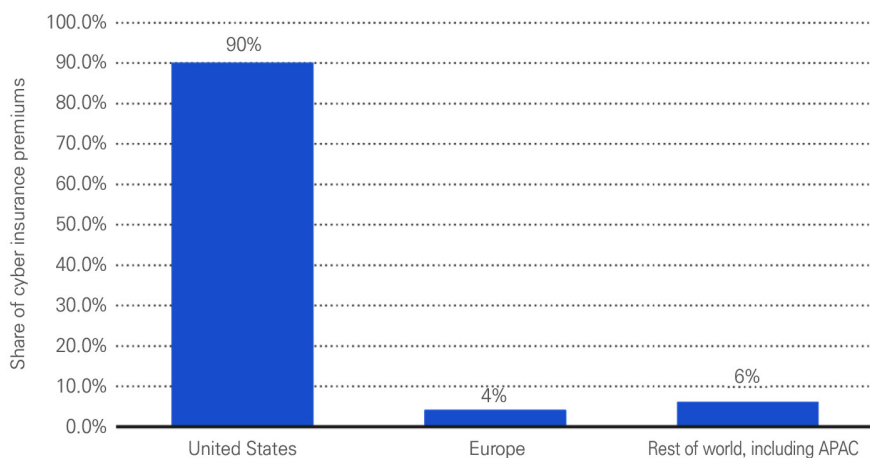
2. 사이버보험 및 인식 현황조사

다음은 사이버보험 시장현황을 조사하고, 시장 확대 가능성을 확인하기 위해 사이버보험에 대한 기업의 구매의향을 조사한다.

가. 사이버보험에 관한 현황

먼저 국가별 사이버보험 시장에 관한 상황을 살펴본다. <그림 II-8>은 2016년 MMC사가 전세계 사이버보험 보험료를 국가별로 분석한 결과이다.

<그림 II-8> 국가별 사이버보험 보험료 분포(2016년 기준)

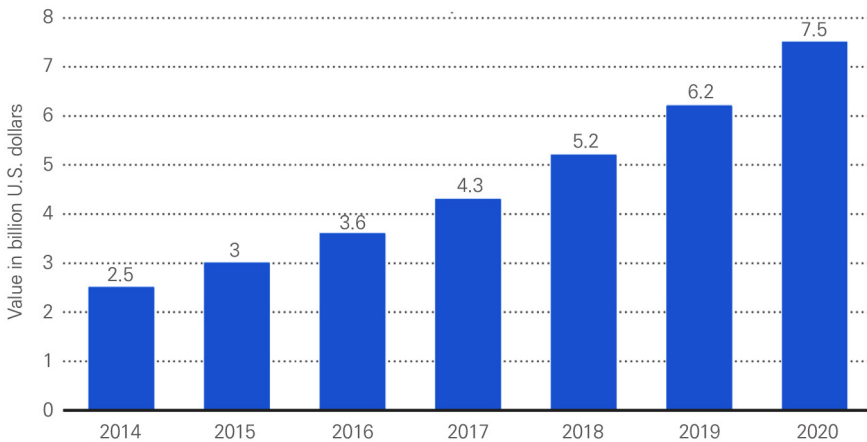


자료: MMC; Oliver Wyman

미국 시장이 전세계 보험료의 대부분(90%)을 차지하고, 유럽 시장이 4%를 차지하는 것으로 나타났다. 아시아-태평양을 포함한 기타 지역에서의 보험액은 전체의 6%에 불과한 것으로 조사되었다. 즉, 사이버공격이 가장 많이 발생하는 미국에서 사이버보험 시장규모도 가장 큰 것으로 확인되었다.

〈그림 II-9〉는 사이버보험 시장의 증가 추이를 보여준다. PwC사와 로이즈(Lloyds) 금융그룹이 2015년 5월 7일부터 2015년 6월 12일까지 PwC의 사이버보험 고객사의 CIO/CISO 임원들을 대상(설문대상: 북미 37%, 유럽 30%, 아시아 16%, 남미 14%, 중동/아프리카 3%)으로 실시한 설문조사 결과, 사이버보험 시장규모가 선형적으로 증가되고 있는 것으로 나타났다. 조사 시작기간인 2014년에는 연간 250억 달러 수준으로 추정되었고, 2018년 시장은 연간 520억 달러(2014년의 2배), 2020년 시장은 연간 750억 달러(2014년의 3배)로 추정되었다.

〈그림 II-9〉 전세계 시장의 연간 사이버보험료 추정액 추이

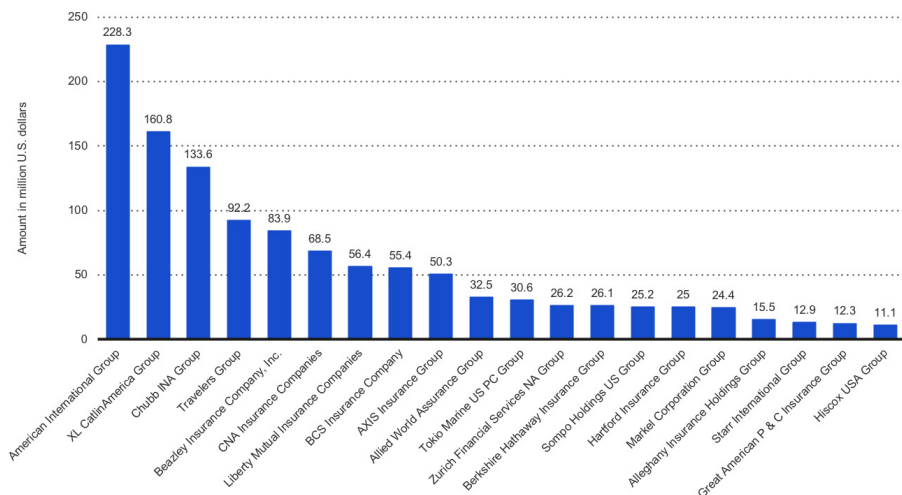


자료: PwC; Lloyds Banking Group; BI Intelligence

〈그림 II-10〉은 2016년 에이엠 베스트(A. M. Best)사가 미국 보험사들의 사이버보험료를 통계 낸 결과이다. AIG 그룹에 지불된 사이버보험료가 2.3억 달러로 가장 높은 것으로 나타났다. XL CatinAmerica 그룹과 Chubb INA 그룹 또한 1억 달러 이상의 사

이버보험료가 지불된 것으로 조사되었다. 이외에 연간보험료 5,000만 달러 이상의 주요 보험사들은 Traveler 그룹, Meazley 보험, CNA 보험, Liberty Mutual 보험, BCS 보험, AXIS 보험그룹이 있었다.

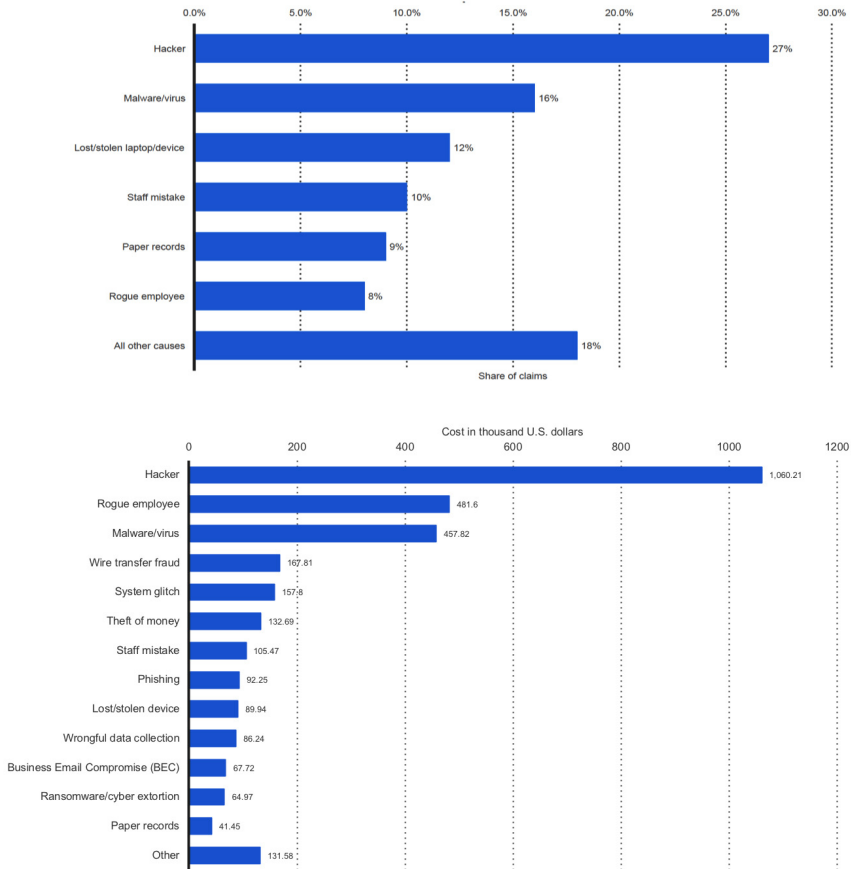
〈그림 II-10〉 사이버보험 보험료 기준 상위 보험사 리스트(2016년 기준)



자료: A. M. Best

〈그림 II-11〉은 넷딜리전스(NetDiligence)가 사이버보험에 가입된 591개 기업을 대상으로 보험 청구와 관련하여 조사한 결과이다. 조사된 사이버사건들은 2014년과 2016년 사이에 발생되었고, 전체 보험금 청구 중 대부분인 582건은 미국에서 이루어졌고, 2건은 캐나다, 4건은 영국, 2건은 호주에서 이루어졌다. 27%의 사건은 해킹에 의해 발생되었다. 사이버사고의 다른 사고원인들은 악성코드/바이러스(16%), 노트북 도난 및 분실(12%), 직원 실수(10%) 등이었다. 해킹에 의해 발생한 사건은 평균적으로 1백만 달러의 비용을 발생시킨 것으로 조사되었다. 악성코드/바이러스 및 불량직원으로 발생한 사건들은 평균 40만 달러의 비용이 발생되었다. 송금 사기, 시스템 에러, 자금 도난, 직원 실수로 발생한 사건들의 평균 피해액은 10만 달러 이상인 것으로 조사되었다.

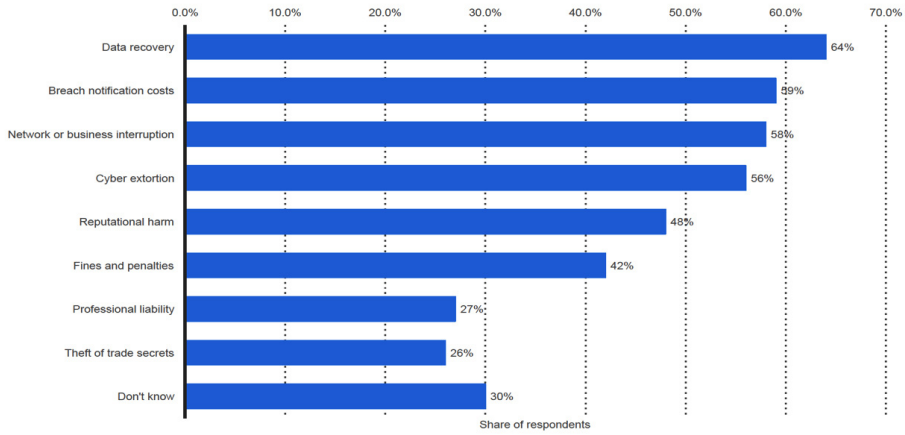
〈그림 II-11〉 사이버공격 종류별 보험 청구 통계



자료: NetDiligence

〈그림 II-12〉는 사이버보험상품의 보장 종류에 대한 통계이다. 스콧보험(Scott Insurance)사가 사이버보험에 가입된 미국 333개 기업을 대상으로 보장 종류를 조사한 결과, 64%의 보험상품이 데이터복원을 보장하는 것으로 조사되었다. 50% 이상의 상품이 데이터유출 보고 비용, 네트워크/경영 중단, 사이버 강탈 등이 보장되는 것으로 나타났다.

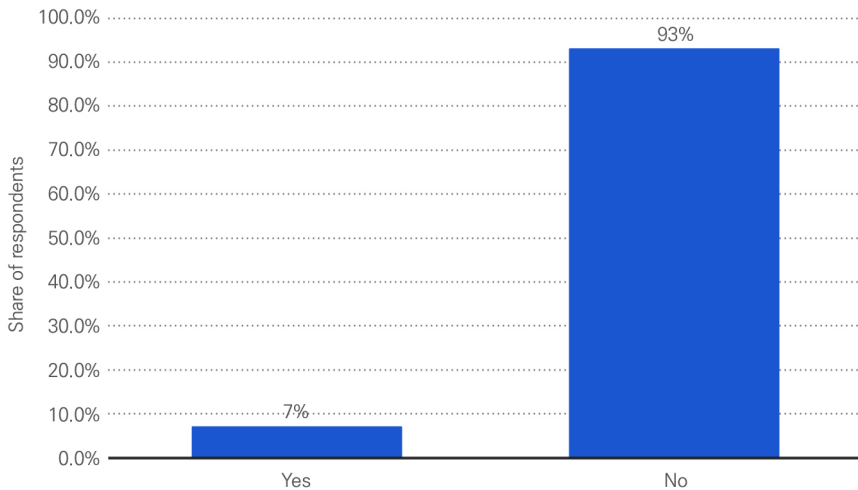
〈그림 II-12〉 사이버보험의 보장 종류



자료: Scott Insurance

이외에도 평판저하, 벌금/벌칙 비용 등이 보장 내용으로 확인되었다. 같은 조사에서, 사이버보험에 가입한 기업들 중 7%만 최근 18개월 내에 사이버보험 청구를 한 것으로 나타났다(〈그림 II-13〉 참고).

〈그림 II-13〉 사이버보험 보험 청구(2016년 기준)

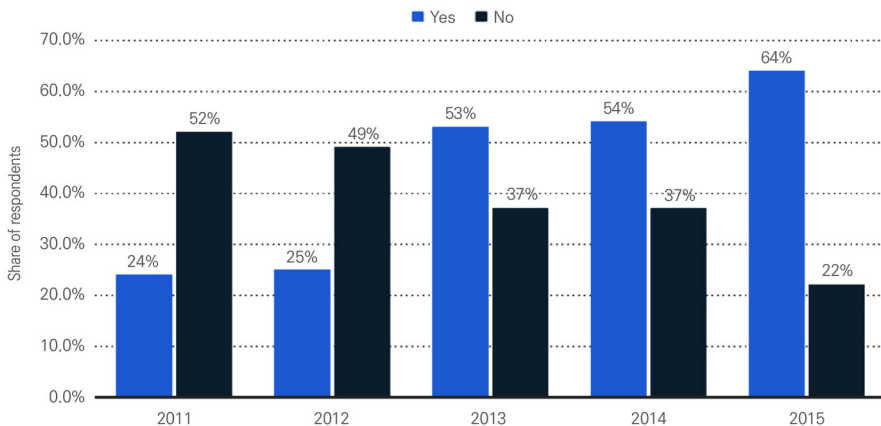


자료: Scott Insurance

나. 기업의 사이버보험에 관한 인식

다음은 사이버보험 시장 확대를 위한 방안을 모색하기 위해, 기업의 사이버보험에 관한 인식에 대해 분석한 결과이다. <그림 II-14>는 취리히(Zurich)사와 어드바이젠(Advisen)사가 공동으로 2011년부터 2015년까지 사이버보험 구매의향을 조사한 결과이다. 2015년 데이터는 448명의 위험관리자(Risk Manager), 보험구매 담당자, 위험전문가들을 대상으로 조사되었다. 구매의향 비율이 2011년 24%에서 2015년 64%로 약 3배 증가된 것으로 나타났다.

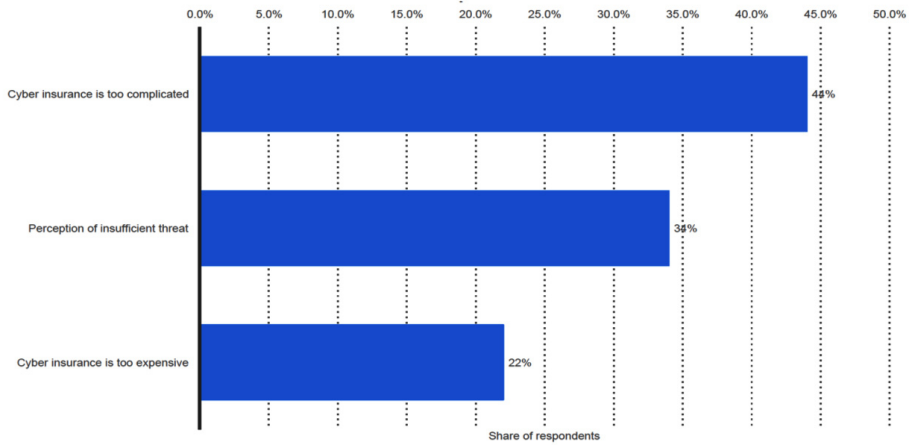
<그림 II-14> 사이버보험 구매의향 설문조사 결과(2011~2015년)



자료: Zurich, Advisen

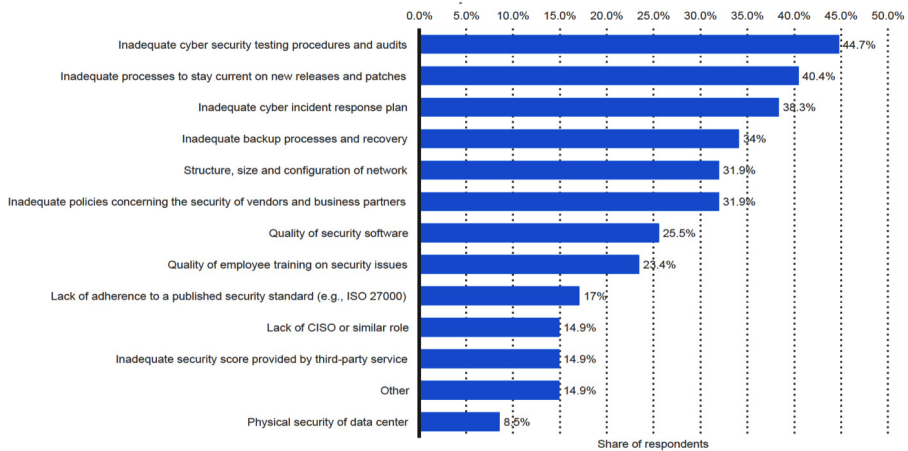
<그림 II-15>는 사이버보험 수요 측면에서 사이버보험의 비구매 이유에 대한 데이터이다. 2016년 미국 샌디에고에서 개최된 위험 및 보험 관리 학회(RIMS: Risk and Insurance Management Society Conference)에서 사이버보험을 구매하지 않는 기업의 위험관리자(Risk Manager) 102명에게 설문조사한 결과이다. 44%의 기업은 사이버보험이 너무 복잡해서 가입하지 않았다고 대답하였고, 34%의 기업은 사이버위험이 충분히 크지 않기 때문에 사이버보험을 구매하지 않은 것으로 나왔다. 22%의 기업은 사이버보험의 높은 보험료를 비구매 이유로 지적하였다.

〈그림 II-15〉 사이버보험 비구매 사유(2016년 기준)



자료: Munich Re

〈그림 II-16〉 사이버보험 가입거부 사유(2016년 기준)



자료: SANS Institute

〈그림 II-16〉은 사이버보험 공급 측면에서 사이버보험 가입신청을 거부한 이유에 대한 자료이다. 산스(SANS Institute)에서 실시한 조사에 따르면, 44.7%의 경우 사이버 보안 점검 및 심사 프로세스의 부재를 거부사유로 들었고, 40.4%의 경우 새로운 소프트웨어 업데이트 프로세스의 부재를 이유로 들었다. 사이버사고에 대한 대책 부재가

세번째로 큰 거부 이유였다. 이외에도 백업 프로세스의 부재, 네트워크 규모 및 설정, 파트너 기업들의 사이버보안 정책 부재 등이 가입신청 거부 이유로 거론되었다.

3. 소결

본 장에서는 사이버위험의 현황, 사이버위험에 대한 기업의 인식, 사이버보험 시장 현황에 대한 데이터를 분석하였다. 정보통신기술에 대한 의존도가 증가함에 따라, 불법적 이득을 취하려는 해커의 사이버공격도 증가하고 있다. 기업의 사이버위험에 대한 인식도 확대되고 있고, 이에 따라 사이버보험에 대한 구매의향도 증가되고 있다. 하지만 사이버보험에 대한 이해 부족, 사이버위험에 대한 대응부재 등의 이유로 보험시장이 아직 충분히 활성화되지 않은 상태이다. 사이버위험에 대한 상세한 데이터가 관리되는 미국이 사이버보험 시장을 선도하고 있다.

Ⅲ. 사이버위험 측정체계 개발

본 장에서는 사이버위험 지수를 체계적으로 측정하기 위한 빅데이터기반의 방법론을 제시하고자 한다. 먼저, 사이버위험과 직접적 연관이 있는 사이버공격에 대한 종류를 살펴본다. 그리고 해당 사이버공격 빅데이터를 수집 및 분석하는 시스템을 개발하여, 사이버위험의 측정체계를 제안한다. 이를 통해 궁극적으로, 측정체계를 통해 사이버보험 데이터 표준화에 기여하고자 한다. 여러 사이버공격 종류 중 스팸이메일 전송(Outbound Spam) 데이터와 피싱웹사이트(Phishing Website) 데이터를 수집하였다. 사이버공격 데이터는 대부분 IP 주소(Internet Protocol Address) 기준으로 수집이 되는데, 이를 기관별(Organization-level) 데이터로 정리하기 위한 3단계 매핑(Mapping) 방법을 제시한다. 기관별 데이터를 이용하여 산업군(Industry Sector)별 랭킹을 도출하여 한국 기관들의 사이버위험 상황을 점검한다. 마지막으로, 제안된 사이버위험 데이터에 대한 정보 공유 및 공시하는 정책이 어떻게 기관들의 사이버위험 관리에 영향을 주는지, 필드실험을 통하여 살펴본다. 필드실험 관련 부분은 필자가 공저한 Choi et al.(2018) 논문을 기반으로 한다.

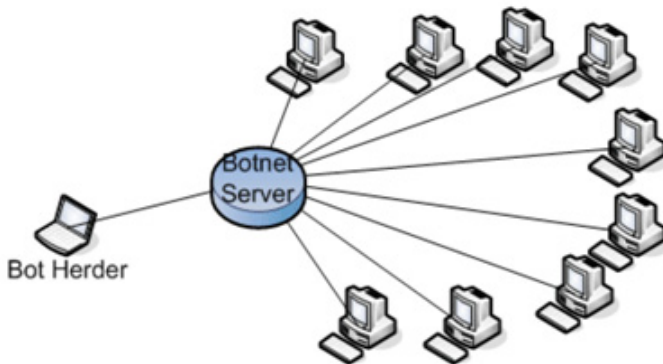
1. 사이버위험 관련 개념 및 데이터 종류

가. 사이버위험의 개념: 봇넷(Botnet)

사이버위험을 이해하기 위해서는, 그 위험의 근본요소인 봇넷(Botnet)에 대한 이해가 필요하다. 봇넷은 해커들이 악의적 목적으로 구축한 컴퓨터들의 네트워크이다. 보

통 하나의 봇넷은 수백만의 감염 컴퓨터들로 구성된다. <그림 III-1>은 봇넷의 개념도이다. 이 네트워크는 악성코드, 바이러스 등에 감염된 컴퓨터들로 구성되고, 봇넷서버(Command and Control Server)를 통하여 해커들의 의도에 따라 조정된다. 봇넷을 운영하는 해커를 봇마스터(Bot Herder: Bot Master)라고 부르는데, 봇넷의 실제 운영자를 추적하는 것은 매우 어렵다. 감염된 컴퓨터의 실제 사용자는 봇넷의 통제 여부를 모르는 경우가 대부분이다. 봇넷은 불법적 의도로 스팸이메일(Spam) 대량발송, 피싱(Phishing)공격, 디도스(DDoS: Distributed Denial of Service)공격을 수행하는 데 주로 사용된다.

<그림 III-1> 봇넷의 개념도



나. 사이버공격 데이터 종류

사이버공격은 주로 봇넷의 통제를 받는 감염된 컴퓨터의 네트워크를 통하여 이루어지는데, 그중 대표적인 사이버공격인 스팸이메일과 피싱공격에 대하여 설명하고자 한다.

스팸이메일은 수신자가 원하지 않은 이메일을 통칭한다. Anderson and Moore (2006) 논문에 따르면, 90% 이상의 스팸이메일은 봇넷에 의해 발생하는 것으로 나타났다. 원하지 않는 마케팅 이메일이 대표적인 스팸이다. 불법약품의 상품홍보를 위하여 스팸이메일이 사용되는 경우도 있고, 향후 피싱공격을 수행하기 위한 첫단계로 사용

되기도 한다. 미국의 대표적 물류업체인 타겟(Target)사에 이루어졌던 2013년 사이버 공격의 경우에서도, 직원이 실수로 열람한 스팸메일을 통하여 악성코드가 타겟 시스템에 설치된 것이 문제의 발단이었다.

피싱공격은 유사 웹사이트(Fake Website) 운영을 통해 이루어진다. 예를 들어, 실제 은행 웹사이트와 동일한 디자인을 가진 사이트에 피해자 접속을 유도하여, 아이디, 비밀번호 등 개인정보를 수집한다. 피싱사이트는 악성코드 다운로드와 같은 2차적 피해를 동반하기도 한다. 피싱사이트는 무료 웹호스팅 서비스나 해킹된 사이트를 이용하는 경우가 대부분이다.

2. 사이버위협 빅데이터 수집 및 처리 시스템

스팸이메일과 피싱사이트는 사이버위협을 측정할 수 있는 좋은 데이터로 활용될 수 있다. 앞서 언급했듯이, 스팸이메일은 봇넷에 의해 발생하는 경우가 대부분이다. 그러므로 특정 인터넷 IP 주소에서 대량 스팸이메일이 발생하는 경우, 해당 IP 주소가 운영되는 시스템의 보안이 취약함을 간접적으로 알 수 있다. 피싱사이트를 운영하기 위해서는 웹서버가 사용되는데, 해당 웹서버의 IP 주소 확인을 통해 해당 시스템의 보안 취약성을 알 수 있다. 이와 같이 스팸이메일 및 피싱사이트 데이터가 사이버위협을 알 수 있는 데이터가 될 수 있다. 또한, 해당 데이터는 피점검 시스템의 내부에 대한 적극적 점검(Active Inspection) 없이도, 데이터의 외부 수집(Externally Observable)을 할 수 있다는 큰 장점이 있다.

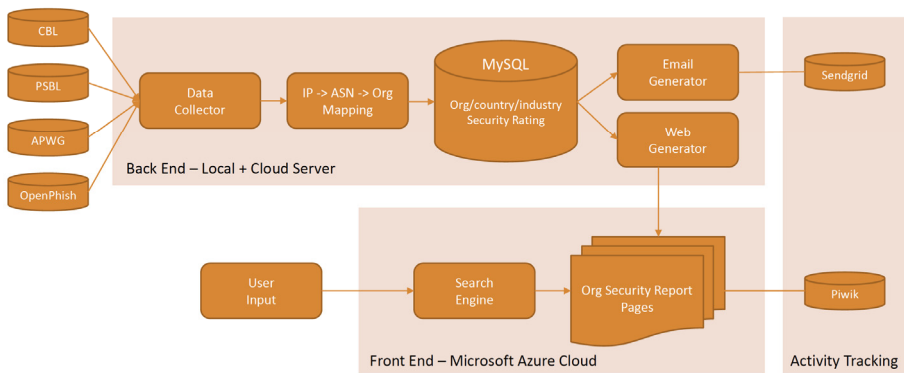
가. 시스템 구조

사이버위협 관련 빅데이터를 수집 및 분석하기 위해서는 대용량 데이터처리 시스템 개발이 요구된다. <그림 III-2>는 Choi et al.(2018) 논문에서 제시된 클라우드 기반의 사이버위협 데이터수집 시스템의 구조이다. 스팸이메일 데이터는 CBL(Composite

Blocking List)¹⁰⁾과 PSBL(Passive Spam Block List)¹¹⁾에서, 피싱사이트 데이터는 APWG(Anti-Phishing Working Group)¹²⁾와 오픈피시(OpenPhish)¹³⁾에서 수집한다. 수집된 원데이터(Raw Data)는 IP 주소 단위로 구성되어있는데, 이를 기관단위 데이터로 구성하기 위한 IP → ASN → Org의 3단계 매핑단계를 거쳐서 MySQL 데이터베이스에 정보가 입력된다. 추후 정보공시를 위하여, 기관별 사이버위협 측정결과 보고서를 이메일 및 웹페이지 형태로 생성한다.

피실험 기관들의 상세한 행위들을 추적하기 위한 시스템을 구현하였다. 먼저 이메일의 송신, 수신, 열람기록은 샌드그리드(Sendgrid) 시스템으로 추적하고, 웹페이지 방문 기록은 웹분석 시스템인 피빅(Piwik)으로 기록한다.

〈그림 III-2〉 사이버위협 데이터 수집 시스템 구조



자료: Choi et al.(2018b)

나. 스팸이메일, 피싱웹사이트 데이터

본 연구에서 사용한 스팸이메일 데이터 제공서비스는 CBL과 PSBL이다. CBL은 스팸

10) <https://www.abuseat.org/>

11) <https://psbl.org/>

12) <https://www.antiphishing.org/>

13) <https://openphish.com/>

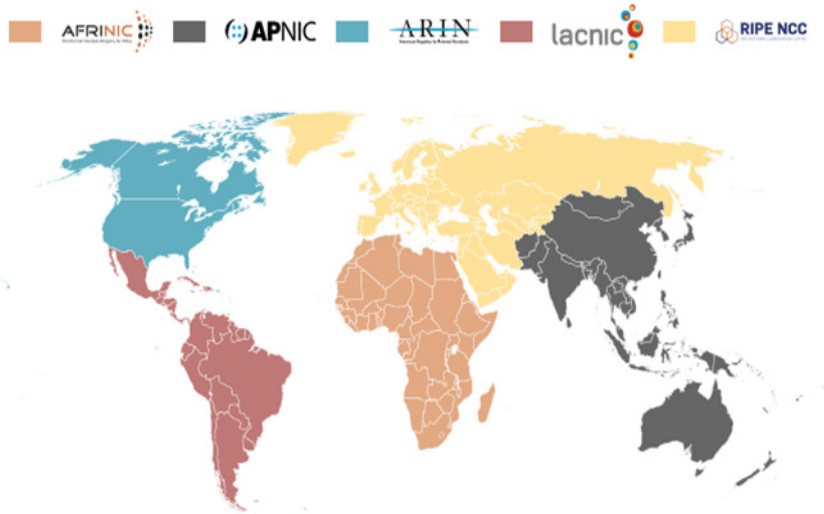
이메일이 발송된 IP 주소, 스팸 발생량, IP 주소를 운영하는 업체 정보, 봇넷 정보 등을 제공하는데, 하루에 1GB 정도의 데이터가 제공한다. 필자는 2011년부터 CBL데이터를 일(Daily) 단위로 수집하였고, 현재까지 총 3,411억 건의 스팸이메일 데이터를 수집하였다. PSBL은 CBL이 제공하는 데이터 종류에 추가하여, 스팸이메일의 내용 정보를 제공한다. 필자는 2012년부터 PSBL 데이터를 수집하였고, 현재까지 총 4억 건의 스팸이메일 데이터를 수집하였다.

다음으로, 피싱사이트 데이터 수집을 위하여 APWG와 오픈피시를 사용하였다. APWG는 현존하는 가장 큰 피싱사이트 데이터이고, 2013년부터 132만 건의 데이터를 수집하였다. 오픈피시는 2015년 7월부터 운영되었고, 76만 건의 데이터를 수집하였다. 각 데이터는 피싱사이트의 URL 주소, 보고된 시각, 사칭하는 브랜드 등의 정보가 있다.

다. 인터넷 주소의 기관 매핑 프로세스

수집된 원데이터는 IP 주소 단위로 구성되어있는데, 이를 기관단위 데이터로 구성하기 위해서는 IP → ASN → Org의 3단계 매핑단계가 거쳐야한다. 전세계의 인터넷 주소는 5개의 지역인터넷기록기관(RIR: Regional Internet Registry)에 의해 관리된다. 한국을 포함한 아시아태평양 국가들은 APNIC(Asia-Pacific Network Information Centre)가 관리하고, 북미 대륙은 ARIN(American Registry for Internet Numbers), 남미 대륙은 LACNIC(Latin America and Carribbean Network Information Centre), 유럽 대륙은 RIPE NCC(Réseaux IP Européens Network Coordination Centre), 그리고 아프리카 대륙은 AfriNIC(African Network Information Centre)에 의해 관리된다. <그림 III-3>은 5개 대륙의 RIR들을 보여준다.

<그림 III-3> 인터넷 관리 기관

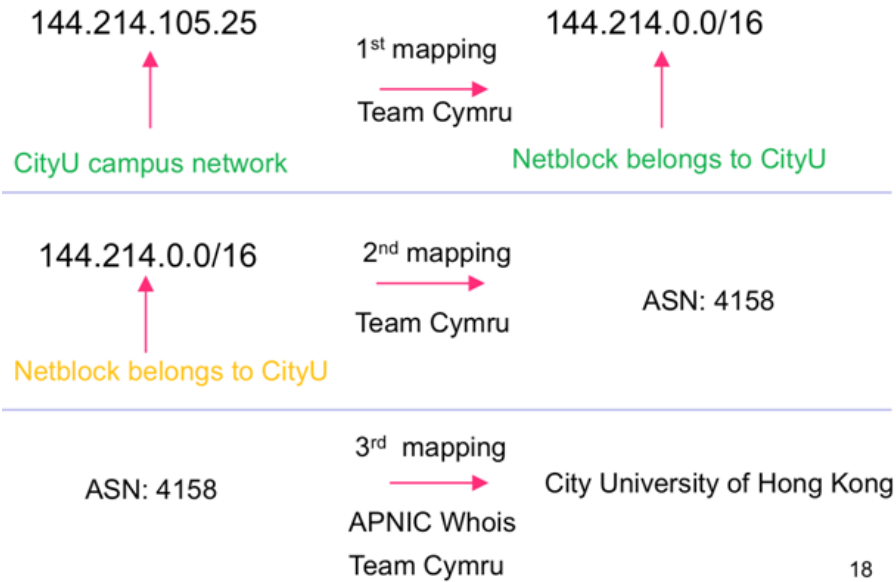


자료: <https://www.arin.net/knowledge/rirs.html>

개별 기관들이 인터넷 주소를 확보하기 위해선, 해당 국가를 관장하는 RIR에서 ASN(Autonomous System Number)을 할당받아야 한다. 경우에 따라 복수의 ASN을 할당받는 기관도 있다. 각 ASN 아래에는 IP 주소 영역대인 넷블록(Netblock)들이 할당된다.¹⁴⁾ IP 주소 기준으로 수집된 사이버위협 데이터를 기관별 데이터로 정리하기 위해서는 3단계의 매핑이 필요하다. 이를 위해 팀컴리(Team Cymru) 데이터를 이용한다. 먼저 IP 주소가 포함된 서버넷을 확보하고, 해당 서버넷이 소속된 ASN을 찾아낸다. 마지막으로 팀컴리 데이터 및 WHOIS 서비스를 통하여, 해당 ASN을 관리하는 기관 정보를 확보한다. 기업들의 인수합병, 기관 통폐합 등의 이유로 인터넷 주소 할당은 동적으로 변화한다. 즉, 특정 IP 주소와 연결된 기관이 바뀌는 경우가 있다. 이에 대응하기 위하여 인터넷 매핑 데이터를 주기적으로 업데이트해야 한다. <그림 III-4>는 IP 주소 144.214.105. 25를 해당 기관인 홍콩시립대학교에 매핑하는 과정을 보여준다.

14) 넷블록(Netblock)은 경우에 따라 서버넷(Subnet)으로 불리기도 함

〈그림 III-4〉 인터넷 주소를 기관으로 매핑하는 과정



18

자료: Choi et al.(2018b)

라. 한국기관 정보수집 및 산업 분류 방법

CBL, PSBL, APWG, 오픈피시 데이터만을 기준으로 사이버위험을 측정하는 경우에는 한 번도 해당데이터에 사이버공격이 감지되지 않은 기관(즉, 사이버위험이 매우 적은 기관)이 포함되지 않는 문제가 발생한다. 포괄적인(Comprehensive) 사이버위험 측정체계를 구축하기 위해서는 인터넷 주소를 할당받은 전체 기관 리스트를 확보해야 한다. 이를 위해, 대륙별 RIR에서 발행하는 ASN 데이터를 수집하였다. 그중 APNIC 데이터에서 인터넷 주소를 할당받은 한국 내의 기관 리스트, ASN 리스트, 해당 기관의 도메인 주소 정보를 추출할 수 있다.

한국에는 2018년 5월 기준으로 총 806개의 ASN이 할당되어있고, 총 543개의 기업 및 기관이 인터넷 주소를 할당받은 것으로 확인되었다. 복수의 ASN을 관리하는 기관들이 다수 존재한다. CJ 헬로가 18개의 ASN을 관리하고 있고, 한국인터넷진흥원(KISA)이 13개, KT는 12개의 ASN을 할당받았고, 카카오그룹, 신한은행, SK브로드밴

드가 5개 이상의 ASN을 관리한다. 이외에도 복수의 ASN을 가지고 있는 기관은 LG데이콤, 카이스트, 국민은행, LG CNS, 대법원 등이 있다. 나머지 대부분의 기관은 단수의 ASN을 관리한다.

사이버위협 지수를 상대적으로 측정하기 위해서, 기관이 소속된 산업군의 특성을 고려해야 한다. 예를 들어, KT, SK브로드밴드 등과 같은 망사업자는 일반사용자들에게 IP 주소를 할당하는 반면, 신한은행, 국민은행과 같은 금융기관들은 주로 회사 내부 시스템 운영을 위해서만 IP 주소를 할당한다.

이러한 기관별 특징을 고려하기 위해서는, 각 기관들을 산업분류에 따라 구분지를 필요성이 존재한다. 인터넷 주소를 할당받은 기관들은 크게 기업, 정부기관, 교육기관, 비영리기관 등이 있다. 먼저 기업들에 대한 산업군을 분류하기 위해서는, APNIC 데이터에서 수집한 기관의 도메인 정보를 사용하였다. 산업분류체계는 한국표준산업분류 10차 개정(KSIC 10; <그림 III-5>)을 기준으로 하였다. 나머지 상장기업이 아닌 기관은 도메인 주소의 뒷부분을 이용하여, 교육기관(edu, ac.kr), 정부기관(gov, go.kr), 비영리기관(org, or.kr), 연구기관(re.kr)으로 구분하였다.

〈그림 III-5〉 한국표준산업분류(2017년 1월 10차 개정 기준)

한국표준산업분류

- ⊕ A.농업, 임업 및 어업(01~03)
- ⊕ B.광 업(05~08)
- ⊕ C.제 조 업(10~34)
- ⊕ D.전기, 가스, 증기 및 공기 조절 공급업(35)
- ⊕ E.수도, 하수 및 폐기물 처리, 원료 재생업(36~39)
- ⊕ F.건 설 업(41~42)
- ⊕ G.도매 및 소매업(45~47)
- ⊕ H.운수 및 창고업(49~52)
- ⊕ I.숙박 및 음식점업(55~56)
- ⊕ J.정보통신업(58~63)
- ⊕ K.금융 및 보험업(64~66)
- ⊕ L.부동산업(68)
- ⊕ M.전문, 과학 및 기술 서비스업(70~73)
- ⊕ N.사업시설 관리, 사업 지원 및 임대 서비스업(74~76)
- ⊕ O.공공 행정, 국방 및 사회보장 행정(84)
- ⊕ P.교육 서비스업(85)
- ⊕ Q.보건업 및 사회복지 서비스업(86~87)
- ⊕ R.예술, 스포츠 및 여가관련 서비스업(90~91)
- ⊕ S.협회 및 단체, 수리 및 기타 개인 서비스업(94~96)
- ⊕ T.가구 내 고용활동 및 달리 분류되지 않은 자가 소비 생산활동
- ⊕ U.국제 및 외국기관(99)

자료: <http://kostat.go.kr/portal/korea/index.action>

〈그림 III-6〉을 한국 상장사들의 산업분류 분포를 보여준다. 제조업이 8,785개 업체로 가장 많은 비중을 차지하였고, 정보통신업 업체는 1,519개, 전문/과학/기술 서비스업 업체는 1,134개, 도매/소매업 업체는 924개, 그리고 금융/보험업 업체는 749개가 상장사 리스트에 존재한다.

〈그림 III-7〉은 인터넷 주소를 할당받은 한국기관에 대한 산업분류 분포이다. 산업분류정보 취합이 가능한 기관 중 가장 큰 비중을 차지한 산업은 교육기관(153)이었다. 다음으로는, 정보통신업(82), 비영리기관(75), 정부기관(67)에 50개 이상의 ASN 정보가 존재하였고, 금융/보험업(42), 전문/과학/기술 서비스업(35), 제조업(29), 연구기관(21), 도소매업(12)은 10개 이상의 ASN 정보가 확보되었다. 하지만 도메인 주소 매칭 방법으로는 280개의 ASN 정보를 얻지 못하였다. 향후, 수작업으로 매칭되지 않은 ASN을 분석하면, 좀 더 많은 산업분류정보를 취득할 수 있을 것이다. 하지만 이 방법은 대용량 데이터의 빠른 처리를 위해서는 지양해야 하는 방식일 것이다.

〈그림 III-6〉 한국 상장사들의 산업분류 분포(2018년 기준)

Rows	1	ksic10_category
8785	C. Manufacturing	
1519	J. Information and communication	
1134	M. Professional, scientific and technical activiti	
924	G. Wholesale and retail trade	
749	K. Financial and insurance activities	
378	F. Construction	
182	H. Transportation and storage	
112	N. Business facilities management and business sup	
77	D. Electricity, gas, steam and air conditioning su	
70	P. Education	
70	R. Arts, sports and recreation related services	
49	A. Agriculture, forestry and fishing	
42	E. Water supply; sewage, waste management, materia	
28	I. Accommodation and food service activities	
28	L. Real estate activities	
21		
21	B. Mining and quarrying	
7	S. Membership organizations, repair and other pers	

자료: 한국기업데이터

〈그림 III-7〉 인터넷 주소 할당 받은 한국기관의 산업분류 분포(2018년 기준)

280	
153	Education
82	J. Information and communication
75	Nonprofit
67	Government
42	K. Financial and insurance activities
35	M. Professional, scientific and technical activiti
29	C. Manufacturing
21	Research
12	G. Wholesale and retail trade
3	F. Construction
2	H. Transportation and storage
2	N. Business facilities management and business sup
2	D. Electricity, gas, steam and air conditioning su
1	B. Mining and quarrying

자료: APNIC: 한국기업데이터 통합

마. 한국기관의 사이버위험 측정 결과

다음은 2018년 5월 데이터 기준으로 한국기관의 사이버위험 지수를 측정한 결과이다. 각 산업군별로 위험지수가 높은 기관들을 나열하였다. 본 위험지수는 CBL데이터를 기준으로 하였으나, 향후 다른 사이버공격 데이터도 향후 추가 가능하다. 전체데이터는 한국 사이버위험 측정 프로젝트 사이트¹⁵⁾에서 확인가능하다.

〈그림 III-8〉은 전체 산업군을 망라한 전체 랭킹이다. 가장 사이버위험 지수가 높은 기관은 KT(이메일스팸 220만 건 이상)였고, 다음으로는 세종텔레콤(5만 건), 넥스지 보안업체(2만 8천 건), KINX 클라우드 컴퓨팅 업체(21,000건), 네이버(7,500건) 등의 정보통신업체, 성균관대학교(4,400건), 고려대학교(4,000건), 청주대학교(3,300건), 한국원자력연구원(2,500건), 대구대학교(2,100건) 등의 교육, 연구기관이 상위를 차지하였다.

15) <https://misr.sauder.ubc.ca/cyberrisk/>

〈그림 III-8〉 한국 사이버위협 랭킹 - 전체(2018년 5월 기준)

Select Year-Month reset		Select Sector reset		Select Criteria reset		Select Domain	
2018-05-00: 233		All: 233		Volume: 233		Select all	
Rank	Year-Month	Domain		Sector	Criteria	Metric	
Rankings							
1	2018-05-00	kt.com		All	Volume	2281400	
2	2018-05-00	sejongtelecom.net		All	Volume	50244	
3	2018-05-00	nexg.net		All	Volume	28008	
4	2018-05-00	kinx.net		All	Volume	21422	
5	2018-05-00	navercorp.com		All	Volume	7507	
6	2018-05-00	skku.edu		All	Volume	4401	
7	2018-05-00	korea.ac.kr		All	Volume	4054	
8	2018-05-00	cju.ac.kr		All	Volume	3349	
9	2018-05-00	kaeri.re.kr		All	Volume	2547	
10	2018-05-00	daegu.ac.kr		All	Volume	2124	

자료: <https://misr.sauder.ubc.ca/cyberrisk/>

〈그림 III-9〉는 교육기관(도메인 주소가 ac.kr이나 edu로 끝나는 기관)의 사이버위협 랭킹이다. 성균관대학교와 고려대학교가 월간 4천 건 이상의 스팸이메일을 발생시켜서 1, 2위를 차지하였다. 이외에도 청주대학교(3,300건), 대구대학교(2,100건), 연세대학교(2,000건), 강원대학교(1,900건), 우송대학교(1,900건), 서울대학교(1,500건), 전북대학교(1,200건), 광운대학교(1,200건)가 상위를 차지하였다. 146개의 교육기관들 중 53개 기관에서 한 건 이상의 스팸이메일이 발송되었다. 대학교에 할당된 IP 주소는 많은 경우 학생들의 휴대폰이나 노트북에 동적으로 할당(DHCP)되는데, 학생들의 컴퓨터에 대한 사이버위협 관리가 부족함을 알 수 있다. 학생들의 사이버보안 인식강화 방안이 필요하다고 판단된다.

〈그림 III-9〉 한국 사이버위협 랭킹 - 교육기관(2018년 5월 기준)

Select Year-Month reset		Select Sector reset		Select Criteria reset		Select Domain	
2018-05-00: 112		Education: 112		Volume: 112		Select all	
Rank	Year-Month	Domain	Sector	Criteria	Metric		
Rankings							
1	2018-05-00	skku.edu	Education	Volume	4401		
2	2018-05-00	korea.ac.kr	Education	Volume	4054		
3	2018-05-00	cju.ac.kr	Education	Volume	3349		
4	2018-05-00	daegu.ac.kr	Education	Volume	2124		
5	2018-05-00	yonsei.ac.kr	Education	Volume	2082		
6	2018-05-00	kangwon.ac.kr	Education	Volume	1998		
7	2018-05-00	wsu.ac.kr	Education	Volume	1958		
8	2018-05-00	snu.ac.kr	Education	Volume	1572		
9	2018-05-00	chonbuk.ac.kr	Education	Volume	1288		
10	2018-05-00	kw.ac.kr	Education	Volume	1281		

자료: <https://misr.sauder.ubc.ca/cyberrisk/>

〈그림 III-10〉은 20개의 정부기관(도메인 주소가 go.kr이나 gov으로 끝나는 기관)의 사이버위협 랭킹이다. 2018년 5월 기준으로 서울시(seoul.go.kr)와 광주교육청에서만 사이버위협 요소가 발견되었다. 이외에 한국투자공사, 한국수출입은행, 한국장학재단, 울산광역시교육청, 정보통신산업진흥원, 서울특별시교육청, 중앙선거관리위원회, 한국교통안전공단 등 18개의 정부기관에서는 CBL 데이터상으로 사이버위협이 감지되지 않았다. 2018년 전체 데이터를 기준으로 했을때, 서울시, 서울시교육청, 광주교육청, 인천교육청 등에서 사이버위협 요소가 발견되었다. 대학기관에 이어, 교육기관에서의 사이버보안 관리가 부족함을 확인하였다. 이외의 정부기관은 보안 관리가 대체적으로 양호함이 확인되었다.

〈그림 III-10〉 한국 사이버위험 랭킹 - 정부기관(2018년 5월 기준)

Select Year-Month reset Select Sector reset Select Criteria reset Select Domain					
2018-05-00: 11		Government: 11	Volume: 11	Select all	
Rank	Year-Month	Domain	Sector	Criteria	Metric
Rankings					
1	2018-05-00	seoul.go.kr	Government	Volume	1703
2	2018-05-00	korea.kr	Government	Volume	130
3	2018-05-00	kic.go.kr	Government	Volume	0
3	2018-05-00	koreaexim.go.kr	Government	Volume	0
3	2018-05-00	kosaf.go.kr	Government	Volume	0
3	2018-05-00	use.go.kr	Government	Volume	0
3	2018-05-00	nipa.kr	Government	Volume	0
3	2018-05-00	sen.go.kr	Government	Volume	0
3	2018-05-00	nec.go.kr	Government	Volume	0
3	2018-05-00	ts2020.kr	Government	Volume	0

자료: <https://misr.sauder.ubc.ca/cyberrisk/>

〈그림 III-11〉은 비영리기관(도메인 주소가 or.kr이나 org로 끝나는 기관)의 사이버 위험 지수 측정결과이다. 경기테크노파크와 한국콘텐츠진흥원에서 스팸이메일이 각 600건 발생되었다. 2018년 전체 데이터 기준으로는 수자원공사, 고대의료원, 한국정보화지능원, 서울대분당병원에서 스팸이메일 발송 문제가 감지되었다. 반면, 보험개발원, 한국사학진흥재단, 동국대학교의료원, 한국농어촌공사, 한국교통안전공단 등에서는 위험이 감지되지 않았다.

〈그림 III-11〉 한국 사이버위험 랭킹 - 비영리기관(2018년 5월 기준)

Select Year-Month reset	Select Sector reset	Select Criteria reset	Select Domain		
2018-05-00: 37	Nonprofit: 37	Volume: 37	Select all		
Rank	Year-Month	Domain	Sector	Criteria	Metric
Rankings					
1	2018-05-00	gtp.or.kr	Nonprofit	Volume	697
2	2018-05-00	kocca.kr	Nonprofit	Volume	662
3	2018-05-00	kidi.or.kr	Nonprofit	Volume	0
3	2018-05-00	sahak.or.kr	Nonprofit	Volume	0
3	2018-05-00	dumc.or.kr	Nonprofit	Volume	0
3	2018-05-00	ekr.or.kr	Nonprofit	Volume	0
3	2018-05-00	khug.or.kr	Nonprofit	Volume	0
3	2018-05-00	cef.or.kr	Nonprofit	Volume	0
3	2018-05-00	fsec.or.kr	Nonprofit	Volume	0

자료: <https://misr.sauder.ubc.ca/cyberrisk/>

〈그림 III-12〉는 상장사들 중 16개의 제조업체에 대한 사이버위험 지수 결과이다. 한화, 오투기, 현대중공업, 농심 등에서 스팸이메일이 다수 발생하였다. 하지만 신영, 남양유업, 포스코, 대우조선해양, 대원제약, 보령제약에서는 문제가 발견되지 않았다.

〈그림 III-12〉 한국 사이버위험 랭킹 - 제조업(2018년 5월 기준)

Select Year-Month reset	Select Sector reset	Select Criteria reset	Select Domain		
2018-05-00: 16	C. Manufacturing: 16	Volume: 16	Select all		
Rank	Year-Month	Domain	Sector	Criteria	Metric
Rankings					
1	2018-05-00	hanwha.co.kr	C. Manufacturing	Volume	22
2	2018-05-00	ottogi.co.kr	C. Manufacturing	Volume	8
3	2018-05-00	hhi.co.kr	C. Manufacturing	Volume	3
4	2018-05-00	shinyoung.com	C. Manufacturing	Volume	0
4	2018-05-00	namyangi.com	C. Manufacturing	Volume	0
4	2018-05-00	posco.com	C. Manufacturing	Volume	0
4	2018-05-00	dsme.co.kr	C. Manufacturing	Volume	0
4	2018-05-00	daewonpharm.com	C. Manufacturing	Volume	0
4	2018-05-00	boryung.co.kr	C. Manufacturing	Volume	0

자료: <https://misr.sauder.ubc.ca/cyberrisk/>

〈그림 III-13〉은 상장사들 중 38개의 정보통신업체에 대한 2018년 5월의 사이버위험 지수 결과이다. KT, 세종텔레콤, 한솔넥스지, KINX, 네이버, 가비아, 이니텍, LG홈쇼핑 등의 업체에서 스팸이메일 대량발송 문제가 감지되었다. 특히, 통신사업자들이 관리하는 IP 주소에서 월간 수만에서 수백만 건의 스팸이메일이 발송되었다. 통신업체들은 사업운영을 위한 시스템뿐만 아니라, 개인이용자들에게 IP 주소를 할당된다. 본 결과를 통해 개인사용자들의 보안관리가 비교적 부실함을 확인할 수 있다. 통신사업자가 개인사용자들의 사이버위험 문제를 선제적으로 발견하고 이에 대응방안을 적극적으로 제시함으로써, 전체적인 사이버위험 요소들을 줄일 수 있을 것으로 기대된다.

필자가 공저한 Choi et al.(2016a) 논문에서는 이러한 점을 이용하여 통신사업자가 사이버보안과 관련하여 방어자(Defender) 및 보험사업자(Insurer)의 역할을 동시에 수행할 것을 제안하였다. 즉, 개인사용자들에게 망접속서비스 이외에도 보안관리 서비스와 사이버보험 서비스까지 제공할 것을 제안한다. 이를 Managed Security Service with Warranty(MSSW)라고 명명하였다. 이러한 사이버위험 관리 및 보험 서비스를 제공함으로써, 인터넷의 전체적인 사이버보안이 강화되는 것으로 수학적으로 증명하였다.

〈그림 III-13〉 한국 사이버위험 랭킹 - 정보통신업(2018년 5월 기준)

Select Year-Month reset	Select Sector reset	Select Criteria reset	Select Domain		
2018-05-00: 27	J. Information and communication	Volume: 27	Select all		
Rank	Year-Month	Domain	Sector	Criteria	Metric
Rankings					
1	2018-05-00	kt.com	J. Information and communication	Volume	2281400
2	2018-05-00	sejongtelecom.net	J. Information and communication	Volume	50244
3	2018-05-00	nexg.net	J. Information and communication	Volume	28008
4	2018-05-00	kinx.net	J. Information and communication	Volume	21422
5	2018-05-00	navercorp.com	J. Information and communication	Volume	7507
6	2018-05-00	gabia.com	J. Information and communication	Volume	524
7	2018-05-00	initech.com	J. Information and communication	Volume	64
8	2018-05-00	gsshop.com	J. Information and communication	Volume	28
9	2018-05-00	douzone.com	J. Information and communication	Volume	0
9	2018-05-00	infrawaretech.com	J. Information and communication	Volume	0

자료: <https://misr.sauder.ubc.ca/cyberrisk/>

〈그림 III-14〉는 상장사들 중 14개의 금융업체에 대한 2018년 5월의 사이버위험 지수 결과이다. 다행히도, 해당 데이터에서는 모든 금융업체에서는 스팸이메일 문제가 발생하지 않았다. 금융시스템은 비교적 철저하게 사이버보안이 관리되는 것으로 보인다. 하지만 2018년도 전체 데이터 기준으로는, 간헐적으로 현대해상과 산업은행에서 사이버위험 요소가 발견되었다.

〈그림 III-14〉 한국 사이버위험 랭킹 - 금융/보험업(2018년 5월 기준)

Select Year-Month reset	Select Sector reset	Select Criteria reset	Select Domain		
2018-05-00: 6	K. Financial and insurance activ	Volume: 6	Select all		
Rank	Year-Month	Domain	Sector	Criteria	Metric
Rankings					
1	2018-05-00	hi.co.kr	K. Financial and insurance activities	Volume	0
1	2018-05-00	jprovest.com	K. Financial and insurance activities	Volume	0
1	2018-05-00	kbfg.com	K. Financial and insurance activities	Volume	0
1	2018-05-00	kjbank.com	K. Financial and insurance activities	Volume	0
1	2018-05-00	ktb.co.kr	K. Financial and insurance activities	Volume	0
1	2018-05-00	miraeeasset.com	K. Financial and insurance activities	Volume	0

자료: <https://misr.sauder.ubc.ca/cyberrisk/>

3. 사이버위험 공시의 효과 측정 실험

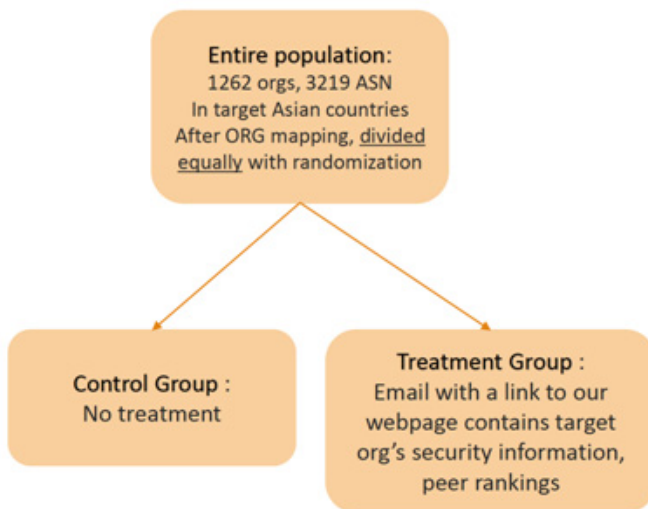
제시된 사이버위험 측정체계의 효용성을 검증할 필요성이 있다. 이에, 사이버위험 측정에 대한 자료공유 및 공시가 어떻게 기관들의 사이버보안에 영향을 미치는지 경험적으로 측정하고자 한다. 이를 위하여 He et al.(2016) 논문과 같이, 대규모의 현장실험(Randomized Field Experiment)을 실시하였다. 한국과 사회문화적으로 유사도가 높은 중국어권 국가들을 대상으로 실험을 실시하였다. 해당 부분은 필자가 공저한 Choi et al.(2018b) 논문을 바탕으로 한다.

먼저 실험 설계에 대해서 논하고, 각 실험군에 대하여 사이버위험 측정결과를 공시하는 이메일, 웹사이트 채널들에 대하여 소개한다. 마지막으로 데이터 분석을 통하여 실험결과를 도출한다.

가. 실험 설계

한국 기관에 대한 사이버위협을 측정한 방법과 동일한 과정을 거쳐서, 중국, 홍콩, 싱가포르, 말레이시아, 대만, 마카오 등 6개국에 소속된 1,262개의 기관(3,219개 ASN)에 대한 사이버위협 지수를 측정하였다. 통계적 임의화(Randomization) 과정을 거쳐서, 631개 기관은 제어군(Control Group)으로, 나머지 631개 기관은 실험군(Treatment Group)으로 선정하였다. 제어군에 소속된 기관들에게는 아무런 조치를 취하지 않은 반면, 실험군에 소속된 기관들에게는 이메일과 홈페이지를 통하여 사이버위협 측정결과를 고지하였다. 실험시작 이전 제어군과 실험군은 통계적으로 동일한 조건을 만족하는 것(Internal Validity)을 확인하였다. 그러므로 실험 이후에 발생한 두 군의 차이점은 실험에 의하여 발생된 것임을 인과적으로 밝힐 수 있다. <그림 III-15>는 실험 설계에 대한 개념도이다. <표 III-1>은 피실험 1,262기관에 대한 통계이다.

<그림 III-15> 실험 설계



자료: Choi et al.(2018b)

〈표 Ⅲ-1〉 아시아 6개국 기관 통계

(단위: 개)

Countries	Number of Organizations	Control Group	Treatment Group
China	309	631	631
Hong Kong	309		
Singapore	264		
Malaysia	171		
Taiwan	138		
Macau	4		
Total	1,262	1,262	

자료: Choi et al.(2018b)

나. 사이버위험 자료 공시 채널

사이버위험 측정결과에 대한 자료를 고지하기 위하여, 이메일과 웹페이지 등 두 가지의 자료공시 채널을 구축하였다. 중국어권 기관들을 대상으로 하는 실험이므로, 중국어와 영어로 작성된 이메일을 전송하였다. 실험 이메일은 2개월 주기로 2017년 7월, 9월, 11월에 전송되었다. 이메일의 송신, 수신, 열람 등의 상세한 정보를 수집하기 위하여 샌드그리드(Sendgrid) 서비스를 사용하였다.

각 실험 이메일에는, 해당 기관의 사이버위험 순위, 상세 사이버위험 지수를 기재하였고, 웹페이지로의 링크를 첨부하였다. 〈그림 Ⅲ-16〉을 통하여 이메일 샘플을 확인할 수 있다.

사이버위험 자료고지 사이트에는 〈그림 Ⅲ-17〉과 같이 기업 및 산업군을 검색할 수 있는 검색시스템이 구현되었다. 그리고 실험군의 각 기관별로 산업/국가 랭킹, 월간, 일간 위험지수 등의 상세정보가 제공된다. 〈그림 Ⅲ-18〉은 알리바바(Alibaba)사의 2017년 10월 사이버위험 지수 보고서이다.

〈그림 III-16〉 사이버위험 지수 공유 이메일 샘플

Security Advisory | 資訊安全建議 | 信息安全建议

This advisory indicates the level of potential spam emails and phishing websites using IP addresses owned by Henan Telcom Union Technology Co Ltd, compared to other organizations in Pan-Asia. This information may be useful in determining information security improvements.

這份報告提供由Henan Telcom Union Technology Co Ltd所擁有的IP地址中有關垃圾郵件和釣魚網站指數，以及貴機構在資訊安全方面與泛亞洲地區其他機構的對比情況。這些訊息或有助於貴機構了解自身在資訊安全方面的改善情況。

這份報告提供由Henan Telcom Union Technology Co Ltd所擁有的IP地址中有關垃圾郵件和釣魚網站指數，以及貴機構在信息安全方面與泛亞洲地區其他機構的對比情況。這些信息或有助於貴機構了解自身在信息安全方面的改善情況。

[Composite Borda](#) ranking for Henan Telcom Union Technology Co Ltd in May 2017:

Henan Telcom Union Technology Co Ltd in May 2017的[綜合波達指數](#)排名：

Henan Telcom Union Technology Co Ltd in May 2017的[綜合波達指數](#)排名：

Borda Rank	Among	HSIC Code	Description
109	1262	631100	Data processing, hosting and related activities

[Composite Borda score](#) is based on four different data sources. Higher [Borda score](#) and ranking indicates a higher level of potential vulnerabilities. For graphics and more information about spam volume and phishing sites originating from your organization, please visit our [Organizational Analysis](#) page. Note that the information provided on this security advisory is publicly searchable on the [cybeRatings](#) website.

[綜合波達指數](#)是由四個不同的數據源經綜合計算得出，較高的[波達指數](#)排名表明貴機構可能有嚴重的安全漏洞。如需獲取以上表格的配套圖表，或了解更多有關由貴機構發出的垃圾郵件或者釣魚網站的資訊，請訪問我們的[機構分析](#)頁面。請注意，此報告中所提供的資訊對所有訪問[cybeRatings](#)網站的用戶開放。

[綜合波達指數](#)是由四個不同的數據源經綜合計算得出，較高的[波達指數](#)排名表明貴機構可能有嚴重的安全漏洞。如需獲取以上表格的配套圖表，或了解更多有關由貴機構發出的垃圾郵件或者釣魚網站的信息，請訪問我們的[機構分析](#)頁面。請注意，此報告中所提供的信息對所有訪問[cybeRatings](#)網站的用戶開放。

Data source details | 數據源詳情 | 数据源详情

The table below shows the data source breakdown of three months of phishing and spam volume rankings based on information collected from March 2017 to May 2017:

這份表格分別列出貴公司在四個數據源中最近三個月由March 2017到May 2017的釣魚網站和垃圾郵件的容量和排名：

這份表格分別列出貴公司在四個數據源中最近三個月由March 2017到May 2017的钓鱼网站和垃圾邮件的容量和排名：

Source	Rank	Volume
CBL	108	736
PSBL	107	48
APWG	N/A due to zero volume	0
OpenPhish	58	2

We hope you find this advisory useful in evaluating and perhaps improving your organization's security. Should you have any questions, please reply to this email or contact us at advisory@cityu.edu.hk. The Dossier ID assigned to your company is 1089. Please include this Dossier ID in all correspondence sent to us.

我們希望這份報告對貴機構在進行資訊安全評估和提升時有所幫助。若閣下有任何查詢，請回復本郵件，或直接發送郵件至 advisory@cityu.edu.hk。貴公司的Dossier ID是1089。向我們發送郵件時，請一併附上這個編號。

我们希望这份报告对贵机构在进行信息安全评估和提升时有所帮助。若閣下有任何查詢，請回復本郵件，或直接發送郵件至 advisory@cityu.edu.hk。貴公司的Dossier ID是1089。向我們發送郵件時，請一併附上這個編號。

자료: Choi et al.(2018b)

〈그림 III-17〉 사이버위험 자료공유 웹사이트 - 메인페이지

cybeRatings About Glossary FAQ Contact Us English 繁體中文 簡體中文

cybeRatings

Search:

2017-10

Top 10 ranks
This ranking is based on the Borda score. It does not contain organizations in the control group.

Name	ASN	DOSSIER ID	HSIC	Rank	Organization name	Country	CBL	PSBL	AWPG	Openphish
Tencent cloud computing (Beijing) Co Ltd	58835	2840	631100	1	China Internet Network Information Center	CN	19928885	29817	364	14
				2	Chunghwa Telecom Co Ltd	TW	9959369	2125	3	3
				3	Alibaba (China) Technology Co Ltd	CN	2854793	747	2	6
				4	TIME	MY	1209566	1627	2	2
				5	China Mobile Communications Corporation	CN	5249049	11301	0	1
				6	China Telecom (Group)	CN	959195	7846	1	0
				7	StarHub	SG	4962891	3280	0	0
				8	KbTelecom	TW	1573615	529	0	2
				9	MAGIX Software GmbH	SG	1120842	4850	0	0
				10	YTL Communications Sdn Bhd	MY	573176	5023	0	0

previous 1 next

자료: <https://cyberatings.is.cityu.edu.hk/>

〈그림 III-18〉 사이버위험 자료공유 웹사이트 - 개별페이지 샘플(Alibaba)



자료: <https://cyberatings.is.cityu.edu.hk/>

다. 데이터 분석 및 실험결과 분석

〈표 III-2〉는 본 실험에 사용된 주요변수 및 제어변수들에 대한 통계이다. 본 실험의 주요변수는 CBL, PSBL, APWG, 오픈피시 데이터 기준으로 계산된 월간 스팸이메일 발생 건수와 월간 피싱사이트 발생 건수이다. 제어변수로는, 기관에 할당된 총 IP 주소 갯수, 소셜미디어 사용 여부, 산업분류코드 등이 사용되었다.

〈표 III-2〉 실험 데이터 통계

Variable	Variable description	Mean	S.D.	Max	Min
CV	CBL Volume	151661.8	2269080	1.00e8	0
PV	PSBL Volume	147.9001	2698.253	157765	0
AV	APWG Volume	0.2372	8.1761	456	0
OV	OpenPhish Volume	0.3249	3.1254	105	0
Number of IP addresses	Total number of IP addresses owned by each company	610223.4	7273093	2.33e8	0
If has social media account	If the company has at least one social media account	0.7035	0.4569	1	0
HSIC	Hong Kong Standard Industrial Classification Code	-	-	960299	50000
If has opened treatment emails	If an organization has opened a treatment email on or before this month	0.2062	0.4048	1	0
If has visited treatment website	If an organization has visited our website on or before this month	0.07080	0.2566	1	0

자료: Choi et al.(2018b)

CBL 데이터의 월 평균은 15만 건으로 주요변수 중 가장 규모가 큰 것으로 나타났다. PSBL의 월 평균값은 147건이었다. 이에 비해 평균 피싱웹사이트 건수는 0.23에서 0.32로 규모가 매우 작은 것을 알 수 있다. 한 기관은 평균적으로 61만 개의 IP 주소를 할당받았고, 70%의 기관이 소셜미디어를 사용하고 있다. 실험군에 소속된 20%의 기관에서 사이버위험 측정결과 고지 이메일을 열람하였고, 7%의 기관에서 웹사이트를 방문한 것으로 나타났다.

본 실험에서 경험적으로 측정하고자 하는 것은, 실험 전후로 제어군과 실험군의 위험지수가 어떻게 변화하는지 여부이다. 이를 위해 차이 내 차이(DiD: Difference-in-Difference)라는 통계방법론을 사용하였다. 아래는 실험결과 측정을 위한 방정식이다.

$$y_{it} = \alpha_0 + \alpha_1 * emailtreat_{it} + \theta_i + \sigma_t + \epsilon_{it},$$

y_{it} 는 기관 i 가 t 월에 발생시킨 스팸이메일/피싱사이트 발생 건수의 로그값이다.¹⁶⁾ 실험변수인 $emailtreat_{it}$ 는 0(제어군) 혹은 1(실험군)의 값을 갖는다. α_1 은 본 실험을 통해 경험적으로 측정하고자 하는 파라미터이다. σ 는 시간 고정효과, θ 는 기관 고정효과 제어를 위한 변수들이다.

〈표 III-3〉은 모집단(1,130기관) 기준으로, 이메일을 수신한 기관들에 대한 실험효과를 측정한 결과이다. CBL 기준으로 실험군의 스팸이메일 발생량이 감소한 것으로 나타났다. 하지만 다른 지수들에 대해서는 통계적으로 의미 있는 변화가 발견되지 않았다.

〈표 III-3〉 실험효과 측정(이메일 수신 기관 기준)

구분	CBL(log)	PSBL(log)	APWG(log)	OP(log)
-	(1)	(2)	(3)	(4)
Email_treat	-0.201*	-0.0237	0.0464	0.00577
-	(0.115)	(0.0659)	(0.0291)	(0.0353)
Observations	13,560	13,560	13,560	13,560
Number of orgs	1,130	1,130	1,130	1,130

주: 통계적 유의성: * p<0.1, ** p<0.05, *** p<0.01

자료: Choi et al.(2018b)

〈표 III-4〉는 실험 이전에 사이버위험 지수가 양의 값을 갖는 기관에 한정하여, 실험효과를 측정한 것이다. CBL과 PSBL 기준으로 스팸이메일 발생량이 통계적으로 유의미한 수준으로 감소되었다. 하지만 여전히 피싱사이트 지수는 통계적으로 의미 있는 변화를 감지하지 못하였다.

16) 사이버위험 지수 모두 비뮴어진 분포(Skewed Distribution)를 따르는데, 이 지수들을 정규 분포화하여 선형회귀분석하기 위하여 위험지수들을 로그변환함

〈표 III-4〉 실험효과 측정(실험 전 사이버위험 감지된 기관 기준)

구분	CBL(log)	PSBL(log)	APWG(log)	OP(log)
-	(1)	(2)	(3)	(4)
Email_treat	-0.4591*	-0.4381**	0.4931	-0.2563
-	(0.2428)	(0.2119)	(0.3875)	(0.3944)
Observations	5,472	3,144	744	1,104
Number of orgs	456	262	62	92

주: 통계적 유의성: * p<0.1, ** p<0.05, *** p<0.01

자료: Choi et al.(2018b)

〈표 III-5〉는 이메일 수신 이후 이메일을 열람한 기관들을 기준으로 실험효과를 측정
한 결과이다.

〈표 III-5〉 실험효과 측정(이메일 열람한 기관 기준)

구분	CBL(log)	PSBL(log)	APWG(log)	OP(log)
-	(1)	(2)	(3)	(4)
Open a treatment email	-1.440*	-2.016**	1.182	-1.451
-	(0.811)	(0.832)	(1.209)	(1.964)
Observations	5,472	3,144	744	1,104
Number of orgs	456	262	62	92

주: 통계적 유의성: * p<0.1, ** p<0.05, *** p<0.01

자료: Choi et al.(2018b)

〈표 III-6〉은 이메일 열람 이후 실험 웹사이트에 방문한 기관들을 기준으로 실험효
과를 측정한 결과이다.

〈표 III-6〉 실험효과 측정(실험 웹사이트 방문한 기관 기준)

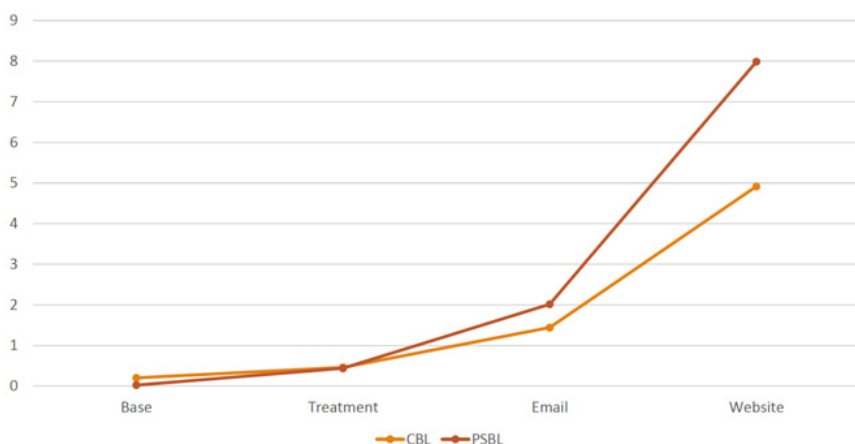
구분	CBL(log)	PSBL(log)	APWG(log)	OP(log)
-	(1)	(2)	(3)	(4)
Visit treatment website	-4.915*	-7.989**	5.403	-3.824
-	(2.836)	(3.940)	(6.840)	(5.446)
Observations	5,472	3,144	744	1,104
Number of orgs	456	262	62	92

주: 통계적 유의성: * p<0.1, ** p<0.05, *** p<0.01

자료: Choi et al.(2018b)

〈그림 III-19〉는 기관의 반응별로 실험효과 크기를 비교한 결과이다. 실험군의 기관들이 적극적으로 사이버위험 보고서에 반응함에 따라 실험효과 또한 증가함이 확인되었다. 즉, 실험군 중 이메일을 열람한 그룹과 공시 사이트에 방문한 그룹에 대한 실험효과가 높은 것으로 나타났다. 이는, 사이버위험 지수의 측정 및 고지하는 것이 기관들의 사이버보안 수준을 높이는 유인으로 작용하는 것을 보여준다.

〈그림 III-19〉 기관 반응별 실험효과 비교



자료: Choi et al.(2018b)

4. 소결

본 장에서는 빅데이터 기반의 사이버위험 측정 방법론을 제시하였다. 사이버공격에 대한 개념을 정리하였고, 여러 가지 사이버공격 종류에 대하여 설명하였다. 그리고 스팸이메일과 피싱사이트 데이터를 수집하고, 기관별 사이버위험 지수를 계산하는 방법을 제시하였다. 해당 방법론을 이용하여 한국 기관들의 사이버위험 지수를 계산하고, 산업별 사이버위험 평가 결과를 도출하였다. 한국의 경우에는 교육기관 및 개인사용자들의 사이버위험 수준이 높은 것으로 나타났다. 제시된 사이버위험 평가 시스템의 유용성을 검증하기 위해, 중국계 기관들을 대상으로 위험평가 결과를 공유하고 공시하

는 현장실험을 수행하였다. 이를 통해 사이버위험 결과공유가 사이버위험 요소를 감소시키는 것으로 확인되었다. 본 장에서 제시된 빅데이터 기반의 사이버위험 측정 방법론은 향후 사이버보험 체계구축을 위한 데이터 플랫폼으로 사용될 수 있을 것으로 기대된다.

IV. 사이버사고 예측모델 개발

사이버보험 시장 활성화를 위해서는 표준데이터 개발이 필요하다. 이를 위하여 이전 장에서는 빅데이터 기반의 사이버위험 측정체계를 제안하였다. 본 장에서는 제안된 사이버위험 지수를 이용하여 사이버사고(Cyber Incident)를 예측하는 모델을 개발하고자 한다. 이는 향후 정확한 보험금 산출에 기여할 것으로 기대한다.

최근 빅데이터, 인공지능(Artificial Intelligence), 기계학습(Machine Learning)의 급속한 발전으로 미래에 대한 정확한 예측이 가능해졌다. 이러한 발전을 바탕으로, 사이버사고를 예측하는 모델을 제시하고자 한다. 사이버사고에 대한 체계적 데이터가 수집되어있는 미국 자료를 이용하여, 예측모델을 개발하였다. 딥러닝 모델 이용 시 최대 94%의 확률로 기관의 사이버사고 여부를 예측할 수 있음을 확인하였다.

1. 기계학습 개념 및 예측 문제 정의

가. 기계학습의 개념

사이버사고 예측모델 개발을 위하여 기계학습 방법론을 사용하고자 하는데, 독자들의 이해를 돕기 위해서 개념을 먼저 설명하고자 한다. 기계학습에 대한 다양한 정의들이 존재하는데, 공통적으로 “학습데이터(Training Data)로부터 통계적 모델(Model)을 학습하여 새로운 데이터에 대한 결과를 예측(Predict)하는 기술”이라고 정의한다. 기계학습은 예측모델(Predictive Modeling), 데이터마이닝(Data Mining), 예측분석(Predictive Analytics)이라고 불리기도 한다.

기계학습은 다양한 예측 문제에 사용되고 있다. 예를 들어, 수신된 이메일이 스팸인지 여부를 예측하는 스팸필터링(Spam Filtering) 기술, 신용카드 거래의 진위 여부를 예측하는 위조방지(Fraud Detection) 기술, 인터넷쇼핑 사용자가 어떤 광고를 클릭할지 예측하는 타겟광고(Targeted Advertising) 기술 등에 사용되고 있다.

기계학습은 학습데이터 내의 정답 존재 여부에 따라 감독학습(Supervised Learning)과 비감독학습(Unsupervised Learning)으로 구분지을 수 있다.

감독학습에서는 데이터의 속성을 나타내는 특징(Feature)과 정답 레이블(Label)이 필요하다. 예를 들어, 소셜네트워크서비스(SNS) 사진 내의 얼굴인식 문제를 푸는 데에 있어서 SNS 사용자가 직접 입력한 얼굴 위치 및 이름 데이터를 사용하는 경우, 이는 감독학습 문제라고 할 수 있다. 영화 추천시스템(Recommendor System)에서 사용자의 이전 평점을 이용하는 것도 감독학습 문제의 사례이다. 감독학습 모델은, 주어진 레이블 데이터의 속성에 따라 이산값(Discrete Value)의 레이블이 주어지는 분류문제(Classification)와 연속값(Continuous Value)의 레이블로 주어지는 회귀문제(Regression)로 나뉜다. 스팸필터링 문제나 얼굴인식 문제는 분류문제의 사례이다. 반면, 주가예측 문제는 회귀문제라고 할 수 있다.

감독학습 문제를 해결하기 위한 대표적 알고리즘은 K-Nearest Neighbor(KNN), Support Vector Machine(SVM), Decision Tree, Random Forest, Logistic Regression, Linear Regression 등이 있다.

비감독학습에서는 정답 레이블 데이터가 주어지지 않고, 특징들만을 이용하여 유사한 데이터를 클러스터링(Clustering)하거나 전체 데이터의 특성을 요약하는 문제를 해결한다. 비감독학습 알고리즘은 크게 클러스터링과 차원감소(Dimensionality Reduction)로 나뉜다.

비감독학습 문제를 해결하는 대표적 알고리즘은 K-means Clustering, Principal Component Analysis(PCA), Multidimensional Scaling(MDS) 등이 있다.

나. 사이버사고 예측문제 정의

사이버위험 예측문제는 사이버위험 지수를 사용하여 주어진 기관이 사이버사고를 당할지 여부를 예측하는 문제이다. 본 문제의 경우 사이버위험 지수가 특징(Feature) 변수이고, 사고 여부가 이산값의 레이블(Label)이다. 그러므로 본 문제는 감독학습 중 분류문제로 치환할 수 있다.

2. 데이터 수집

가. 사이버사고 데이터 수집

사이버사고 중에서 가장 대표적인 것이 정보유출(Data Breach) 사고이다. 미국 등 여러 국가들은 법률에 의해 정보유출사고를 당한 기관들이 사고정보를 공시하도록 되어있다. 이 때문에 미국기관의 정보유출사고에 데이터는 비교적 잘 정리되어있다. 관련 데이터는 Privacy Clearing House(PCH),¹⁷⁾ VERIS 커뮤니티 데이터베이스(VCDB)¹⁸⁾에서 수집하였다. PCH 데이터는 미국 보건복지부(U.S. Department of Health and Human Services), Dataloss DB, 외부미디어 등 다양한 데이터소스에서 제공된 8,068건의 정보유출에 대한 상세한 정보를 제공한다. 버라이즌(Verizon)사가 관리하는 VCDB는 여러 데이터소스와 자발적 자료공유로 수집된 8,451건의 정보유출에 관한 정보를 제공한다. 각 사고기록에는 사고날짜, 사고기관, 유출된 기록 건수 등의 정보가 존재한다.

17) <https://www.privacyrights.org/data-breaches>

18) <https://github.com/vz-risk/VCDB>

나. 사이버위험 특징추출

기계학습 알고리즘을 적용하기 위해서는 사이버위험 지수 데이터를 이용하여 각 기관들에 대한 특징변수를 추출(Feature Extraction)해야 한다. 이번 연구에서는 CBL과 PSBL 데이터를 기준으로 특징추출을 수행하였다. 향후 다른 사이버위험 데이터 확보 시, 추가적으로 특징변수로 사용할 수 있다. 먼저 두 데이터상에서 각 기관들의 스팸이메일 발생총량(Volume), 발생 건수(Count), 스팸발생 IP 개수(Host) 등 6개의 특징을 추출하였다. CBL에서는 스팸이메일을 발생시킨 봇넷이 포함되는데, 현재까지 400여 개의 봇넷 정보가 수집되었다. 각 기관에서 발생된 스팸발생량을 400여 개의 봇넷별로 분류한 데이터를 특징으로 추출하였다.

다. 사이버사고와 사이버위험 데이터 통합

다음으로는 정보유출 데이터와 사이버위험 지수 데이터를 통합하는 과정이 필요하다. 하지만 두 가지의 데이터는 통합 ID를 사용하지 않으므로, 이중 데이터를 하나로 통합하는 것은 매우 어려운 과정이다. 이번 연구에서는 양 데이터에서 나온 기관의 이름에 스트링매칭(String Matching) 알고리즘을 적용하여 데이터를 매칭하였다. 그 결과 2,102개의 기관에 대한 데이터를 통합할 수 있었다. 통합데이터에는 243건의 정보유출사고가 포함되었다.

이는 전체 정보유출사고 건수에 비해 적은 비율이다. 많은 매칭이 이루어지지 않은 이유는 정확한 피해기관의 이름이 기재되지 않은 경우, 그리고 한 기관이 다양한 이름으로 기재되는 경우가 많기 때문이다. 수작업으로 데이터를 통합하는 방법을 통하여 더 많은 통합데이터를 수집할 수 있을 것이다. 하지만 이는 확장성(Scalability)이 있어서, 사이버보험 시장에 직접 사용되기는 어려울 것이다. 향후 사이버위험, 사이버사고 등의 데이터 수집 및 통합을 용이하게 하기 위해서는 각 데이터에서 통합 ID를 사용하게 하는 방안이 필요하다.

3. 예측 알고리즘 및 모델 평가

이미 언급된 것처럼, 사이버사고 예측문제는 감독 기계학습 문제 중 분류문제로 생각할 수 있는데, 이미 개발된 전통적 기계학습 알고리즘들을 적용하고, 또한 최근 새롭게 개발된 딥러닝(Deep Learning) 알고리즘을 예측문제에 적용하였다.

가. 전통적 분류 알고리즘

분류 알고리즘 중 대표적인 K-Nearest Neighbor(KNN),¹⁹⁾ Logistic Regression(Logit),²⁰⁾ Decision Tree(DT),²¹⁾ Random Forest(RF)²²⁾ 등 4가지 알고리즘을 사용하였다. Python 내의 오픈소스 기계학습 라이브러리인 Scikit-Learn을 사용하였다.²³⁾

Logit은 독립변수의 선형결합을 이용하여 사건의 발생 가능성을 예측하는 데 사용되는 통계 기법인데, 사건의 종류가 2가지인 경우에 주로 사용된다. 독립변수가 수치 자료인 경우에 많이 사용된다.

KNN 알고리즘은 각 데이터를 다차원 특징 공간(Feature Space)상에 배열하였을 때, 지역적으로 근접한 데이터들이 같은 성질을 가진다고 가정한다. 데이터 간의 거리를 측정하는 방법으로는 유클리드 거리(Euclidean Distance) 혹은 코사인거리(Cosine Distance)를 사용한다. 이는 인스턴스 기반 학습이고 가장 간단한 기계학습 알고리즘에 속한다. 특징공간이 초고차원일 때, 데이터 간의 거리가 기하급수적으로 멀어진다는 단점을 가지고 있다.

Decision Tree 학습법은 어떤 항목에 대한 관측값과 목표값을 연결시켜주는 예측 모델로서, 결정 나무를 사용한다. 의사결정 분석에서 DT는 시각적이고 명시적인 방법으로 의사 결정 과정과 결정된 의사를 보여주는 장점이 있다. 반면, 훈련데이터가 제대

19) https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

20) https://en.wikipedia.org/wiki/Logistic_regression

21) https://en.wikipedia.org/wiki/Decision_tree

22) https://en.wikipedia.org/wiki/Random_forest

23) http://scikit-learn.org/stable/supervised_learning.html

로 일반화하지 못할 경우 너무 복잡한 결정 나무를 만든다는 한계점을 지니고 있다. 그리고 약간의 훈련데이터 차이에 따라 나무의 모양이 크게 달라질 수 있다.

이러한 문제점을 해결하기 위한 방법이 Random Forest이다. 훈련 과정에서 구성한 다수의 Decision Tree로부터 분류치를 종합하는 앙상블(Ensemble) 학습방법이다.

마지막으로, 벤치마킹을 위해서 임의예측(Random) 알고리즘을 사용하였다. 즉, 학습데이터에서 사고발생 확률을 계산한 후, 개별 기관에 대한 개별적 특징을 고려하지 않고, 확률에 따라 사고가 발생할 것으로 예측하는 방법이다.

나. 딥러닝 알고리즘

전통적인 기계학습 방법은 특징추출의 단계를 거치는데, 이때 해당분야에 대한 전문가의 직관이 필요하다. 예를 들어, 사이버보안 문제가 발생하는 데에 영향을 끼치리라 직관적으로 예상되는 변수들을 추출하게 된다. 하지만 이러한 방법은, 특징변수가 전문가의 직관에 제한되는 한계점을 가진다. 이를 극복하기 위한 방법이 인공신경망(Artificial Neural Network)을 이용하는 딥러닝(Deep Learning) 기술이다. 딥러닝 기술에서는 특징추출 과정 없이 원데이터를 직접 인공신경망에 적용하여 원데이터 간의 복잡한 상호작용을 복구할 수 있다. 다만, Decision Tree와 같이 의사결정 과정을 투명하게 보여주기 어려운 특징(Interpretability Issue)을 가지고 있다. 이에 대한 연구를 지속적으로 진행하고 있다.

필자가 공저한 Shin et al.(2016) 논문 사례에서 볼 수 있듯이, 딥러닝 기술은 사진과 동영상과 같은 비정형 데이터를 분석하는 데 널리 사용되고 있다. 일부 사진분석 문제에 있어서는 인간의 능력보다 뛰어남이 밝혀졌다.²⁴⁾ 본 연구에서는 딥러닝 알고리즘 중에 하나인 전방전달 신경망(Feed-Forward Neural Network)을 이용하였다.²⁵⁾

24) <https://www.forbes.com/sites/michaelthomsen/2015/02/19/microsofts-deep-learning-project-outperforms-humans-in-image-recognition/>

25) https://en.wikipedia.org/wiki/Feedforward_neural_network

다. 예측모델 평가 방법

머신러닝으로 개발된 예측모델을 평가할 때 교차검증(Cross Validation) 방법이 사용된다. 즉, 수집된 데이터를 학습데이터(Training Data)와 평가데이터(Testing Data)로 나누고, 학습데이터만을 이용하여 예측모델을 학습하고, 학습된 모델을 평가데이터에 적용하여 예측 정확성을 평가하는 방법이다. 본 연구에서는 90%의 데이터를 학습데이터로 10%의 데이터를 평가데이터로 사용하는 10-Fold 교차검증을 사용하였다.

전체 기관들의 데이터를 기준으로 보았을 때, 사이버사고는 비교적 자주 발생하지 않는 사고이다. 그러므로 레이블 데이터에 불균형 문제(Class Imbalance Problem)가 발생한다. 이러한 경우에 공정한 평가를 위해서는 Accuracy를 사용하지 않고, Precision, Recall, F1 score를 사용한다.²⁶⁾ 아래는 4가지의 예측 정확도 측정 방법이다.

이때, TP는 실제 발생사건을 정확히 예측하는 경우(True Positive)의 건수, FP는 실제 발생하지 않은 사건을 발생한 것으로 잘못 예측하는 경우(False Positive)의 건수, TN은 실제 발생하지 않은 사건을 미발생한 것으로 정확히 예측한 경우(True Negative)의 건수, FN은 실제 발생사건에 대하여 예측하지 못한 경우(False Negative)의 건수를 의미한다.

라. 예측모델 성능비교

〈표 IV-1〉은 앞서 설명한 6가지 분류 알고리즘에 대한 예측성능을 비교한 결과이다. 딥러닝 방식이 가장 높은 94.73%(F1 기준) 예측률을 기록하였다. 전통적인 기계학습 알고리즘들 중에서는 Decision Tree(91.63%)와 Random Forest(91.92%)가 높은 예측률을 보였다. 반면, KNN과 Logistic Regression은 90% 이하의 예측률을 기록하였다. 벤치마크로 사용한 임의예측(Random) 방식은 59%의 예측률을 보였다.

이 결과를 통해, 딥러닝 방식의 특징 자동추출이 전통적 기계학습 대비 3%의 성능향상을 이룬 것이 확인되었다. 임의예측을 포함한 모든 알고리즘의 90% 이상의

26) https://en.wikipedia.org/wiki/F1_score

Precision 값을 가지지만, Recall 값에서 큰 차이를 보였다. 임의예측은 Recall 값이 45%인 반면, 딥러닝은 Recall 96%를 기록하였다.

〈표 IV-1〉 예측모델 성능비교

(단위: %)

구분	Precision	Recall	F1 score
Random	91.93	45.97	59.46
KNN	93.66	85.31	88.86
Logistic Regression	95.47	67.3	76.94
Decision Tree	92.88	90.52	91.63
Random Forest	92.94	91	91.92
Deep Learning	96.35	96.21	94.73

4. 소결

본 장에서는 사이버사고의 대표적인 종류인 데이터유출사고 데이터를 이용하여, 사이버사고 예측에 대한 가능성을 타진하였다. 예측모델 개발을 위하여, 본 연구에서 제시된 사이버위험 지수들을 이용하여 기관들의 특징변수들을 추출하였다. 그리고 다양한 기계학습 알고리즘들을 사이버사고/사이버위험 통합데이터에 적용하였다. 특징변수를 연구자가 직접 설계하는 전통적인 기계학습 방법에서는 90% 정도(F1 기준)의 예측률이 달성되었다. 이어 원데이터에서 직접 변수들의 상호작용을 추출하는 딥러닝 알고리즘을 적용하였을때, F1 기준으로 94.73%의 예측률이 달성되었다. 본 결과를 통해, 제시된 사이버위험 지수 데이터가 사이버사고를 예측하는 데 유용하다는 것을 입증한 것이다. 향후 기계학습 기반의 사고예측 모델은 사이버보험 시장에서 보험금 계산에 유용하게 사용될 것으로 기대된다.

V. 결론

정보통신기술의 발전은 우리 사회에 여러 긍정적인 효과를 가져다 주고 있다. 하지만 동시에 기술에 대한 의존도가 증가함에 따라, 이를 악용하는 악의적 사이버공격 또한 늘어나고 있고, 이에 따른 사회적 손실이 크게 발생하고 있다. 이러한 사이버위험을 관리하기 위한 사이버보험에 대한 수요가 늘어나고 있다. 하지만 사이버위험 및 사이버보험에 대한 데이터 공유 및 표준화가 이루어지지 않은 한계 때문에, 사이버보험 시장이 아직 활성화되지 않고 있는 상황이다.

본 보고서에서는 사이버공격, 사이버위험, 사이버보험과 관련된 현황을 조사하고, 사이버공격 빅데이터 기반의 사이버위험 측정체계를 제안하며, 기계학습을 이용하여 사이버사고에 대한 예측모델을 개발하였다. 이를 통하여, 사이버보험 시장의 활성화에 기여하고자 한다.

하지만 본 보고서에서 제안된 방법론들을 적용하기 위해서는 기본 원데이터 수집이 반드시 필요하다. 하지만 한국기관과 관련된 사이버위험 데이터는 아직 크게 부족한 상황이다. 본 연구에서는 인터넷 주소 기반의 사이버공격 데이터에서 한국 기관 데이터를 추출할 수 있었다. 반면, 사이버사고 및 사이버보험에 대한 한국 데이터는 전무한 상황이다. 한국 사이버보안 시장이 본격적으로 활성화되기 위해서는 사이버위험과 관련된 데이터를 공유하고 표준화하는 정책이 필요하다고 판단한다. 더불어 사이버보험과 관련된 데이터 공유를 촉진할 수 있는 인센티브 제도를 개발함으로써, 사회 전체적으로 사이버보안을 강화할 수 있을 것이다.

참고문헌

- Anderson, R. and Moore, T.(2006), “The Economics of Information Security”, *Science*, 314(5799)
- Blei, D., Ng A., and Jordan, M.(2003), “Latent Dirichlet Allocation”, *Journal of Machine Learning Research*, 3
- Choi, J. H., Choi, Y., Lee, G. M., and Whinston, A. B.(2018a), “To Disconnect or Not: A Cybersecurity Game”, SSRN Working Paper
- Choi, Y., He, S., Zhuang, Y., Lee, G. M., Leung, A. C. M., and Whinston, A. B.(2018b), “Information Disclosure and Security Policy Design: A Large-Scale Randomization Experiment in Pan-Asia”, Workshop on Economics of Information Security
- Geneva Association(2016), *Ten Key Questions on Cyber Risk and Cyber Risk Insurance*
- He, S., Lee, G. M., Han S., and Whinston, A. B.(2016), “How Would Information Disclosure Influence Organizations’ Outbound Spam Volume? Evidence from a Field Experiment”, *Journal of Cybersecurity*, 2(1)
- Lee, G. M., Qiu, L., and Whinston, A. B.(2016), “A Friend Like Me: Modeling Network Formation in a Location-Based Social Network”, *Journal of Management Information Systems*, 33(4)
- Shi, Z., Lee, G. M., and Whinston, A. B.(2016), “Toward a Better Measure of Business Proximity: Topic Modeling for Industry Intelligence”, *MIS Quarterly*, 40(4)
- Shin, D., He, S., Lee, G. M., and Whinston, A. B.(2016), “Analyzing Content and Customer Engagement in Social Media with Deep Learning”, INFORMS Conference on Information Systems and Technology
- Song, V., Cavusoglu, H., Lee, G. M., and Ma, M. L. Z.(2018), “IT Risk Factor Disclosure and Stock Price Crash Risk”, UBC Cybersecurity Summit

보험연구원(KIRI) 발간물 안내

※ 2017년부터 기존의 연구보고서, 정책보고서, 경영보고서, 조사보고서가 연구보고서로 통합되었습니다.

■ 연구보고서

- 2017-1 보험산업 미래 / 김석영·윤성훈·이선주 2017.2
- 2017-2 자동차보험 과실상계제도 개선방안 / 전용식·채원영 2017.2
- 2017-3 상호협정 관련 입법정책 연구 / 정호열 2017.2
- 2017-4 저소득층 노후소득 보장을 위한 공사연계연금 연구 / 정원석·강성호·마지혜 2017.3
- 2017-5 자영업자를 위한 사적소득보상체계 개선방안 / 류건식·강성호·김동겸 2017.3
- 2017-6 우리나라 사회안전망 개선을 위한 현안 과제 / 이태열·최장훈·김유미 2017.4
- 2017-7 일본의 보험회사 도산처리제도 및 사례 / 정봉은 2017.5
- 2017-8 보험회사 업무위탁 관련 제도 개선방안 / 이승준·정인영 2017.5
- 2017-9 부채시가평가제도와 생명보험회사의 자본관리 / 조영현·이혜은 2017.8
- 2017-10 효율적 의료비 지출을 통한 국민건강보험의 보장성 강화 방안 / 김대환 2017.8
- 2017-11 인슈어테크 혁명: 현황 점검 및 과제 고찰 / 박소정·박지윤 2017.8
- 2017-12 생산물 배상책임보험 역할 제고 방안 / 이기형·이규성 2017.9
- 2017-13 보험금청구권과 소멸시효 / 권영준 2017.9
- 2017-14 2017년 보험소비자 설문조사 / 동향분석실 2017.10
- 2017-15 2018년도 보험산업 전망과 과제 / 동향분석실 2017.11
- 2017-16 퇴직연금 환경변화와 연금세제 개편 방향 / 강성호·류건식·김동겸 2017.12
- 2017-17 자동차보험 한방진료 현황과 개선방안 / 송윤아·이소양 2017.12
- 2017-18 베이비부머 세대의 노후소득 / 최장훈·이태열·김미화 2017.12
- 2017-19 연금세제 효과연구 / 정원석·이선주 2017.12
- 2017-20 주요국의 지진보험 운영 현황 및 시사점 / 최창희·한성원 2017.12
- 2017-21 사적연금의 장기연금수령 유도방안 / 김세중·김유미 2017.12
- 2017-22 누적전망이론을 이용한 생명보험과 연금의 유보가격 측정 연구 / 지홍민 2017.12
- 2018-1 보증연장 서비스 규제 방안 / 백영화·박정희 2018.1
- 2018-2 건강생활서비스 공·사 협력 방안 / 조용운·오승연·김동겸 2018.2
- 2018-3 퇴직연금 가입자교육 개선 방안 / 류건식·강성호·이상우 2018.2

- 2018-4 IFRS 9과 보험회사의 ALM 및 자산배분 / 조영현·이혜은 2018.2
 2018-5 보험상품 변천과 개발 방향 / 김석영·김세영·이선주 2018.2
 2018-6 계리적 관점에서 본 실손의료보험 개선 방안 / 조재린·정성희 2018.3
 2018-7 국내 보험회사의 금융겸업 현황과 시사점 / 전용식·이혜은 2018.3
 2018-8 장애인의 위험보장 강화 방안 / 오승연·김석영·이선주 2018.4
 2018-9 주요국 공·사 건강보험 연계 체계 분석 / 정성희·이태열·김유미 2018.4
 2018-10 정신질환 위험보장 강화 방안 / 이정택·임태준·김동겸 2018.4
 2018-11 기초서류 준수의무 위반 시 과징금 부과기준 개선방안 / 황현아·백영화·권오경 2018.9
 2018-12 2018년 보험소비자 설문조사 / 동향분석실 2018.9
 2018-13 상속법의 관점에서 본 생명보험 / 최준규 2018.9
 2018-14 호주 퇴직연금제도 현황과 시사점 / 이경희 2018.9

■ 연구보고서(구)

- 2008-1 보험회사의 리스크 중심 경영전략에 관한 연구 / 최영목·장동식·김동겸 2008.1
 2008-2 한국 보험시장과 공정거래법 / 정호열 2008.6
 2008-3 확정급여형 퇴직연금의 자산운용 / 류건식·이경희·김동겸 2008.3
 2009-1 보험설계사의 특성분석과 고능률화 방안 / 안철경·권오경 2009.1
 2009-2 자동차사고의 사회적 비용 최소화 방안 / 기승도 2009.2
 2009-3 우리나라 가계부채 문제의 진단과 평가 / 유경원·이혜은 2009.3
 2009-4 사적연금의 노후소득보장 기능제고 방안 / 류건식·이창우·김동겸 2009.3
 2009-5 일반화선형모형(GLM)을 이용한 자동차보험 요율상대도 산출 방법 연구 / 기승도·김대환 2009.8
 2009-6 주행거리에 연동한 자동차보험제도 연구 / 기승도·김대환·김혜란 2010.1
 2010-1 우리나라 가계 금융자산 축적 부진의 원인과 시사점 / 유경원·이혜은 2010.4
 2010-2 생명보험 상품별 해지율 추정 및 예측 모형 / 황진태·이경희 2010.5
 2010-3 보험회사 자산관리서비스 사업모형 검토 / 진 익·김동겸 2010.7

■ 정책보고서(구)

- 2008-2 환경오염리스크관리를 위한 보험제도 활용방안 / 이기형 2008.3
 2008-3 금융상품의 정의 및 분류에 관한 연구 / 유지호·최 원 2008.3
 2008-4 2009년도 보험산업 전망과 과제 / 이진면·이태열·신종협·황진태·유진아·

- 김세환·이정환·박정희·김세중·최이섭 2008.11
- 2009-1 현 금융위기 진단과 위기극복을 위한 정책제언 / 진 익·이민환·유경원·최영목·최형선·최 원·이경아·이혜은 2009.2
- 2009-2 퇴직연금의 급여 지급 방식 다양화 방안 / 이경희 2009.3
- 2009-3 보험분쟁의 재판외적 해결 활성화 방안 / 오영수·김경환·이종욱 2009.3
- 2009-4 2010년도 보험산업 전망과 과제 / 이진면·황진태·변혜원·이경희·이정환·박정희·김세중·최이섭 2009.12
- 2009-5 금융상품판매전문회사의 도입이 보험회사에 미치는 영향 / 안철경·변혜원·권오경 2010.1
- 2010-1 보험사기 영향요인과 방지방안 / 송윤아 2010.3
- 2010-2 2011년도 보험산업 전망과 과제 / 이진면·김대환·이경희·이정환·최 원·김세중·최이섭 2010.12
- 2011-1 금융소비자 보호 체계 개선방안 / 오영수·안철경·변혜원·최영목·최형선·김경환·이상우·박정희·김미화 2010.4
- 2011-2 일반공제사업 규제의 합리화 방안 / 오영수·김경환·박정희 2011.7
- 2011-3 퇴직연금 적립금의 연금전환 유도방안 / 이경희 2011.5
- 2011-4 저출산·고령화와 금융의 역할 / 윤성훈·류건식·오영수·조용운·진 익·유진아·변혜원 2011.7
- 2011-5 소비자 보호를 위한 보험유통채널 개선방안 / 안철경·이경희 2011.11
- 2011-6 2012년도 보험산업 전망과 과제 / 윤성훈·황진태·이정환·최 원·김세중·오병국 2011.12
- 2012-1 인적사고 보험금의 지급방식 다양화 방안 / 조재린·이기형·정인영 2012.8
- 2012-2 보험산업 진입 및 퇴출에 관한 연구 / 이기형·변혜원·정인영 2012.10
- 2012-3 금융위기 이후 보험규제 변화 및 시사점 / 임준환·유진아·이경아 2012.11
- 2012-4 소비자중심의 변액연금보험 개선방안 연구: 공시 및 상품설계 개선을 중심으로 / 이기형·임준환·김해식·이경희·조영현·정인영 2012.12
- 2013-1 생명보험의 자살면책기간이 자살에 미치는 영향 / 이창우·윤상호 2013.1
- 2013-2 퇴직연금 지배구조체계 개선방안 / 류건식·김대환·이상우 2013.1
- 2013-3 2013년도 보험산업 전망과 과제 / 윤성훈·전용식·이정환·최 원·김세중·채원영 2013.2
- 2013-4 사회안전망 체계 개편과 보험산업 역할 / 진 익·오병국·이성은 2013.3
- 2013-5 보험지주회사 감독체계 개선방안 연구 / 이승준·김해식·조재린 2013.5
- 2013-6 2014년도 보험산업 전망과 과제 / 윤성훈·전용식·최 원·김세중·채원영 2013.12
- 2014-1 보험시장 경쟁정책 투명성 제고방안 / 이승준·강민규·이해량 2014.3

- 2014-2 국내 보험회사 지급여력규제 평가 및 개선방안 / 조재린·김해식·김석영 2014.3
- 2014-3 공·사 사회안전망의 효율적인 역할 제고 방안 / 이태열·강성호·김유미 2014.4
- 2014-4 2015년도 보험산업 전망과 과제 / 윤성훈·김석영·김진익·최 원·채원영·이아름·이해랑 2014.11
- 2014-5 의료보장체계 합리화를 위한 공·사건강보험 협력방안 / 조용운·김경환·김미화 2014.12
- 2015-1 보험회사 재무건전성 규제 - IFRS와 RBC 연계방안 / 김해식·조재린·이경아 2015.2
- 2015-2 2016년도 보험산업 전망과 과제 / 윤성훈·김석영·김진익·최 원·채원영·이아름·이해랑 2015.11
- 2016-1 정년연장의 노후소득 개선 효과와 개인연금의 정책방향 / 강성호·정봉은·김유미 2016.2
- 2016-2 국민건강보험 보장률 인상 정책 평가: DSGE 접근법 / 임태준·이정택·김혜란 2016.11
- 2016-3 2017년도 보험산업 전망과 과제 / 동향분석실 2016.12

■ 경영보고서(구)

- 2009-1 기업휴지보험 활성화 방안 연구 / 이기형·한상용 2009.3
- 2009-2 자산관리서비스 활성화 방안 / 진 익 2009.3
- 2009-3 탄소시장 및 녹색보험 활성화 방안 / 진 익·유시용·이경아 2009.3
- 2009-4 생명보험회사의 지속가능성장에 관한 연구 / 최영목·최 원 2009.6
- 2010-1 독립판매채널의 성장과 생명보험회사의 대응 / 안철경·권오경 2010.2
- 2010-2 보험회사의 윤리경영 운영실태 및 개선방안 / 오영수·김경환 2010.2
- 2010-3 보험회사의 퇴직연금사업 운영전략 / 류건식·이창우·이상우 2010.3
- 2010-4(1) 보험환경변화에 따른 보험산업 성장방안 / 산업연구실·정책연구실·동향분석실 2010.6
- 2010-4(2) 종합금융서비스를 활용한 보험산업 성장방안 / 금융제도실·재무연구실 2010.6
- 2010-5 변액보험 보증리스크관리연구 / 권용재·장동식·서성민 2010.4
- 2010-6 RBC 내부모형 도입 방안 / 김해식·최영목·김소연·장동식·서성민 2010.10
- 2010-7 금융보증보험 가격결정모형 / 최영수 2010.7

- 2011-1 보험회사의 비대면채널 활용방안 / 안철경·변혜원·서성민 2011.1
- 2011-2 보증보험의 특성과 리스크 평가 / 최영목·김소연·김동겸 2011.2
- 2011-3 충성도를 고려한 자동차보험 마케팅전략 연구 / 기승도·황진태 2011.3
- 2011-4 보험회사의 상조서비스 기여방안 / 황진태·기승도·권오경 2011.5
- 2011-5 사기성클레임에 대한 최적조사방안 / 송윤아·정인영 2011.6
- 2011-6 민영의료보험의 보험리스크관리 방안 / 조용운·황진태·김미화 2011.8
- 2011-7 보험회사의 개인형 퇴직연금 운영방안 / 류건식·김대환·이상우 2011.9
- 2011-8 퇴직연금시장의 환경변화에 따른 확정기여형 퇴직연금 운영방안 / 김대환·류건식·이상우 2011.10
- 2012-1 국내 생명보험회사의 기업공개 평가와 시사점 / 조영현·전용식·이혜은 2012.7
- 2012-2 보험산업 비전 2020 : @sure 4.0 / 진 익·김동겸·김혜란 2012.7
- 2012-3 현금흐름방식 보험료 산출의 시행과 과제 / 김해식·김석영·김세영·이혜은 2012.9
- 2012-4 보험회사의 장수리스크 발생원인과 관리 방안 / 김대환·류건식·김동겸 2012.9
- 2012-5 은퇴가구의 경제형태 분석 / 유경원 2012.9
- 2012-6 보험회사의 날씨리스크 인수 활성화 방안: 지수형 날씨보험을 중심으로 / 조재린·황진태·권용재·채원영 2012.10
- 2013-1 자동차보험시장의 가격경쟁이 손해율에 미치는 영향과 시사점 / 전용식·채원영 2013.3
- 2013-2 중국 자동차보험 시장점유율 확대방안 연구 / 기승도·조용운·이소양 2013.5
- 2016-1 뉴 노멀 시대의 보험회사 경영전략 / 임준환·정봉은·황인창·이혜은·김혜란·정승연 2016.4
- 2016-2 금융보증보험 잠재 시장 연구: 지방자치단체 자금조달 시장을 중심으로 / 최창희·황인창·이경아 2016.5
- 2016-3 퇴직연금시장 환경변화와 보험회사 대응방안 / 류건식·강성호·김동겸 2016.5

■ 조사보고서(구)

- 2008-1 보험회사 글로벌화를 위한 해외보험시장 조사 / 양성문·김진익·지재원·박정희·김세중 2008.2

- 2008-2 노인장기요양보험 제도 도입에 대응한 장기간병보험 운영 방안 / 오영수 2008.3
- 2008-3 2008년 보험소비자 설문조사 / 안철경·기승도·이상우 2008.4
- 2008-4 주요국의 보험상품 판매권유 규제 / 이상우 2008.3
- 2009-1 2009년 보험소비자 설문조사 / 안철경·이상우·권오경 2009.3
- 2009-2 Solvency II의 리스크 평가모형 및 측정 방법 연구 / 장동식 2009.3
- 2009-3 이슬람 보험시장 진출방안 / 이진면·이정환·최이섭·정중영·최태영 2009.3
- 2009-4 미국 생명보험 정산거래의 현황과 시사점 / 김해식 2009.3
- 2009-5 헤지펀드 운용전략 활용방안 / 진 익·김상수·김종훈·변귀영·유시용 2009.3
- 2009-6 복합금융 그룹의 리스크와 감독 / 이민환·전선애·최 원 2009.4
- 2009-7 보험산업 글로벌화를 위한 정책적 지원방안 / 서대교·오영수·김영진 2009.4
- 2009-8 구조화금융 관점에서 본 금융위기 분석 및 시사점 / 임준환·이민환·윤건용·최 원 2009.7
- 2009-9 보험리스크 측정 및 평가 방법에 관한 연구 / 조용운·김세환·김세중 2009.7
- 2009-10 생명보험계약의 효력상실·해약분석 / 류건식·장동식 2009.8
- 2010-1 과거 금융위기 사례분석을 통한 최근 글로벌 금융위기 전망 / 신종협·최형선·최 원 2010.3
- 2010-2 금융산업의 영업행위 규제 개선방안 / 서대교·김미화 2010.3
- 2010-3 주요국의 민영건강보험의 운영체제와 시사점 / 이창우·이상우 2010.4
- 2010-4 2010년 보험소비자 설문조사 / 변혜원·박정희 2010.4
- 2010-5 산재보험의 운영체제에 대한 연구 / 송운아 2010.5
- 2010-6 보험산업 내 공정거래규제 조화방안 / 이승준·이종욱 2010.5
- 2010-7 보험종류별 진료수가 차등적용 개선방안 / 조용운·서대교·김미화 2010.4
- 2010-8 보험회사의 금리위험 대응전략 / 진 익·김해식·유진아·김동겸 2011.1
- 2010-9 퇴직연금 규제체계 및 정책방향 / 류건식·이창우·이상우 2010.7
- 2011-1 생명보험설계사 활동실태 및 만족도 분석 / 안철경·황진태·서성민 2011.6
- 2011-2 2011년 보험소비자 설문조사 / 김대환·최 원 2011.5
- 2011-3 보험회사 녹색금융 참여방안 / 진 익·김해식·김혜란 2011.7
- 2011-4 의료시장 변화에 따른 민영실손의료보험의 대응 / 이창우·이기형 2011.8
- 2011-5 아세안 주요국의 보험시장 규제제도 연구 / 조용운·변혜원·이승준·김경환·오병국 2011.11
- 2012-1 2012년 보험소비자 설문조사 / 황진태·전용식·윤상호·기승도·이상우·최 원 2012.6

- 2012-2 일본의 퇴직연금제도 운영체계 특징과 시사점 / 이상우·오병국 2012.12
- 2012-3 솔벤시 II의 보고 및 공시 체계와 시사점 / 장동식·김경환 2012.12
- 2013-1 2013년 보험소비자 설문조사 / 전용식·황진태·변혜원·정원석·박선영·이상우·최 원 2013.8
- 2013-2 건강보험 진료비 전망 및 활용방안 / 조용운·황진태·조재린 2013.9
- 2013-3 소비자 신뢰 제고와 보험상품 정보공시 개선방안 / 김해식·변혜원·황진태 2013.12
- 2013-4 보험회사의 사회적 책임 이행에 관한 연구 / 변혜원·조영현 2013.12
- 2014-1 주택연금 연계 간병보험제도 도입 방안 / 박선영·권오경 2014.3
- 2014-2 소득수준을 고려한 개인연금 세제 효율화방안: 보험료 납입단계의 세제방식 중심으로 / 정원석·강성호·이상우 2014.4
- 2014-3 보험규제에 관한 주요국의 법제연구: 모집채널, 행위 규제 등을 중심으로 / 한기정·최준규 2014.4
- 2014-4 보험산업 환경변화와 판매채널 전략 연구 / 황진태·박선영·권오경 2014.4
- 2014-5 저시경제 환경변화의 보험산업 파급효과 분석 / 전성주·전용식 2014.5
- 2014-6 국내경제의 일본식 장기부진 가능성 검토 / 전용식·윤성훈·채원영 2014.5
- 2014-7 건강생활관리서비스 사업모형 연구 / 조용운·오승연·김미화 2014.7
- 2014-8 보험개인정보 보호법제 개선방안 / 김경환·강민규·이해랑 2014.8
- 2014-9 2014년 보험소비자 설문조사 / 전용식·변혜원·정원석·박선영·오승연·이상우·최 원 2014.8
- 2014-10 보험회사 수익구조 진단 및 개선방안 / 김석영·김세중·김혜란 2014.11
- 2014-11 국내 보험회사의 해외사업 평가와 제언 / 전용식·조영현·채원영 2014.12
- 2015-1 보험민원 해결 프로세스 선진화 방안 / 박선영·권오경 2015.1
- 2015-2 재무건전성 규제 강화와 생명보험회사의 자본관리 / 조영현·조재린·김혜란 2015.2
- 2015-3 국내 배상책임보험 시장 성장 저해 요인 분석 - 대인사고 손해배상액 산정 기준을 중심으로 - / 최창희·정인영 2015.3
- 2015-4 보험산업 신뢰도 제고 방안 / 이태열·황진태·이선주 2015.3
- 2015-5 2015년 보험소비자 설문조사 / 동향분석실 2015.8
- 2015-6 인구 및 가구구조 변화가 보험 수요에 미치는 영향 / 오승연·김유미 2015.8
- 2016-1 경영환경 변화와 주요 해외 보험회사의 대응 전략 / 전용식·조영현 2016.2
- 2016-2 시스템리스크를 고려한 복합금융그룹 감독방안 / 이승준·민세진 2016.3
- 2016-3 저성장 시대 보험회사의 비용관리 / 김해식·김세중·김현경 2016.4
- 2016-4 자동차보험 해외사업 경영성과 분석과 시사점 / 전용식·송윤아·채원영

- 2016.4
- 2016-5 금융·보험세제연구: 집합투자기구, 보험 그리고 연금세제를 중심으로 / 정원석·임 준·김유미 2016.5
- 2016-6 가용자본 산출 방식에 따른국내 보험회사 지급여력 비교 / 조재린·황인창·이경아 2016.5
- 2016-7 해외 사례를 통해 본 중·소형 보험회사의 생존전략 / 이태열·김해식·김현경 2016.5
- 2016-8 생명보험회사의 연금상품 다양화 방안: 종신소득 보장기능을 중심으로 / 김세중·김혜란 2016.6
- 2016-9 2016년 보험소비자 설문조사 / 동향분석실 2016.8
- 2016-10 자율주행자동차 보험제도 연구 / 이기형·김혜란 2016.9

■ 조사자료집

- 2014-1 보험시장 자유화에 따른 보험산업 환경변화 / 최 원·김세중 2014.6
- 2014-2 주요국 내부자본적정성 평가 및 관리 제도 연구 - Own Risk and Solvency Assessment - / 장동식·이정환 2014.8
- 2015-1 고령층 대상 보험시장 현황과 해외사례 / 강성호·정원석·김동겸 2015.1
- 2015-2 경증치매자 보호를 위한 보험사의 치매실태 도입방안 / 정봉은·이선주 2015.2
- 2015-3 소비자 금융이해력 강화 방안: 보험 및 연금 / 변혜원·이해랑 2015.4
- 2015-4 글로벌 금융위기 이후 세계경제의 구조적 변화 / 박대근·박춘원·이향용 2015.5
- 2015-5 노후소득보장을 위한 주택연금 활성화 방안 / 전성주·박선영·김유미 2015.5
- 2015-6 고령화에 대응한 생애자산관리 서비스 활성화 방안 / 정원석·김미화 2015.5
- 2015-7 일반 손해보험 요율제도 개선방안 연구 / 김석영·김혜란 2015.12
- 2018-1 변액연금 최저보증 및 사업비 부과 현황 조사 / 김세환 2018.2

■ 연차보고서

- 제 1 호 2008년 연차보고서 / 보험연구원 2009.4
- 제 2 호 2009년 연차보고서 / 보험연구원 2010.3
- 제 3 호 2010년 연차보고서 / 보험연구원 2011.3

- 제 4 호 2011년 연차보고서 / 보험연구원 2012.3
- 제 5 호 2012년 연차보고서 / 보험연구원 2013.3
- 제 6 호 2013년 연차보고서 / 보험연구원 2013.12
- 제 7 호 2014년 연차보고서 / 보험연구원 2014.12
- 제 8 호 2015년 연차보고서 / 보험연구원 2015.12
- 제 9 호 2016년 연차보고서 / 보험연구원 2017.1
- 제 10 호 2017년 연차보고서 / 보험연구원 2018.1

■ 영문발간물

- 제 7 호 Korean Insurance Industry 2008 / KIRI, 2008.9
- 제 8 호 Korean Insurance Industry 2009 / KIRI, 2009.9
- 제 9 호 Korean Insurance Industry 2010 / KIRI, 2010.8
- 제10호 Korean Insurance Industry 2011 / KIRI, 2011.10
- 제11호 Korean Insurance Industry 2012 / KIRI, 2012.11
- 제12호 Korean Insurance Industry 2013 / KIRI, 2013.12
- 제13호 Korean Insurance Industry 2014 / KIRI, 2014.8
- 제14호 Korean Insurance Industry 2015 / KIRI, 2015.8
- 제15호 Korean Insurance Industry 2016 / KIRI, 2016.8
- 제16호 Korean Insurance Industry 2017 / KIRI, 2017.8
- 제 7 호 Korean Insurance Industry Trend 2Q FY2013 / KIRI, 2014.2
- 제 8 호 Korean Insurance Industry Trend 3Q FY2013 / KIRI, 2014.5
- 제 9 호 Korean Insurance Industry Trend 1Q FY2014 / KIRI, 2014.8
- 제10호 Korean Insurance Industry Trend 2Q FY2014 / KIRI, 2014.10
- 제11호 Korean Insurance Industry Trend 3Q FY2014 / KIRI, 2015.2
- 제12호 Korean Insurance Industry Trend 4Q FY2014 / KIRI, 2015.4
- 제13호 Korean Insurance Industry Trend 1Q FY2015 / KIRI, 2015.8
- 제14호 Korean Insurance Industry Trend 2Q FY2015 / KIRI, 2015.11
- 제15호 Korean Insurance Industry Trend 3Q FY2015 / KIRI, 2016.2
- 제16호 Korean Insurance Industry Trend 4Q FY2015/ KIRI, 2016.6
- 제17호 Korean Insurance Industry Trend 1Q FY2016/ KIRI, 2016.9
- 제18호 Korean Insurance Industry Trend 2Q FY2016/ KIRI, 2016.12
- 제19호 Korean Insurance Industry Trend 3Q FY2016/ KIRI, 2017.2
- 제20호 Korean Insurance Industry Trend 4Q FY2016/ KIRI, 2017.5

제21호 Korean Insurance Industry Trend 1Q FY2017/ KIRI, 2017.9

제22호 Korean Insurance Industry Trend 2Q FY2017/ KIRI, 2017.11

■ CEO Report

- 2008-1 자동차보험 물적담보 손해를 관리 방안 / 기승도 2008.6
- 2008-2 보험산업 소액지급결제시스템 참여 관련 주요 이슈 / 이태열 2008.6
- 2008-3 FY2008 수입보험료 전망 / 동향분석실 2008.8
- 2008-4 퇴직급여보장법 개정안의 영향과 보험회사 대응과제 / 류건식·서성민 2008.12
- 2009-1 FY2009 보험산업 수정전망과 대응과제 / 동향분석실 2009.2
- 2009-2 퇴직연금 예금보험요율 적용의 타당성 검토 / 류건식·김동겸 2009.3
- 2009-3 퇴직연금 사업자 관련규제의 적정성 검토 / 류건식·이상우 2009.6
- 2009-4 퇴직연금 가입 및 인식실태 조사 / 류건식·이상우 2009.10
- 2010-1 복수사용자 퇴직연금제도의 도입 및 보험회사의 대응과제 / 김대환·이상우·김혜란 2010.4
- 2010-2 FY2010 수입보험료 전망 / 동향분석실 2010.6
- 2010-3 보험소비자 보호의 경영전략적 접근 / 오영수 2010.7
- 2010-4 장기손해보험 보험사기 방지를 위한 보험금 지급심사제도 개선 / 김대환·이기형 2010.9
- 2010-5 퇴직금 중간정산의 문제점과 개선과제 / 류건식·이상우 2010.9
- 2010-6 우리나라 신용카드시장의 특징 및 개선논의 / 최형선 2010.11
- 2011-1 G20 정상회의의 금융규제 논의 내용 및 보험산업에 대한 시사점 / 김동겸 2011.2
- 2011-2 영국의 공동계정 운영체제 / 최형선·김동겸 2011.3
- 2011-3 FY2011 수입보험료 전망 / 동향분석실 2011.7
- 2011-4 근퇴법 개정에 따른 퇴직연금 운영방안과 과제 / 김대환·류건식 2011.8
- 2012-1 FY2012 수입보험료 전망 / 동향분석실 2012.8
- 2012-2 건강생활서비스법 제정(안)에 대한 검토 / 조용운·이상우 2012.11
- 2012-3 보험연구원 명사초청 보험발전 간담회 토론 내용 / 윤성훈·전용식·전성주·채원영 2012.12
- 2012-4 새정부의 보험산업 정책(I): 정책공약집을 중심으로 / 이기형·정인영 2012.12
- 2013-1 새정부의 보험산업 정책(II): 국민건강보험 본인부담경감제 정책에 대한

- 평가 / 김대환·이상우 2013.1
- 2013-2 새정부의 보험산업 정책(III): 제18대 대통령직인수위원회 제안 국정과제를 중심으로 / 이승준 2013.3
- 2013-3 FY2013 수입보험료 수정 전망 / 동향분석실 2013.7
- 2013-4 유럽 복합금융그룹의 보험사업 매각 원인과 시사점 / 전용식·윤성훈 2013.7
- 2014-1 2014년 수입보험료 수정 전망 / 동향분석실 2014.6
- 2014-2 인구구조 변화가 보험계약규모에 미치는 영향 분석 / 김석영·김세중 2014.6
- 2014-3 『보험 혁신 및 건전화 방안』의 주요 내용과 시사점 / 이태열·조재린·황진태·송윤아 2014.7
- 2014-4 아베노믹스 평가와 시사점 / 임준환·황인창·이혜은 2014.10
- 2015-1 연말정산 논란을 통해 본 소득세제 개선 방향 / 강성호·류건식·정원석 2015.2
- 2015-2 2015년 수입보험료 수정 전망 / 동향분석실 2015.6
- 2015-3 보험산업 경쟁력 제고 방안 및 이의 영향 / 김석영 2015.10
- 2016-1 금융규제 운영규정 제정 의미와 시사점 / 김석영 2016.1
- 2016-3 2016년 수입보험료 수정 전망 / 동향분석실 2016.7
- 2016-4 EU Solvency II 경과조치의 의미와 시사점 / 황인창·조재린 2016.7
- 2016-5 비급여 진료비 관련 최근 논의 동향과 시사점 / 정성희·이태열 2016.9
- 2017-1 보험부채 시가평가와 보험산업의 과제 / 김해식 2017.2
- 2017-2 2017년 수입보험료 수정 전망 / 동향분석실 2017.7
- 2017-3 1인 1 퇴직연금시대의 보험회사 IRP 전략 / 류건식·이태열 2017.7
- 2018-1 2018년 수입보험료 수정 전망 / 동향분석실 2018.7
- 2018-2 북한 보험산업의 이해와 대응 / 안철경·정인영 2018.7

■ Insurance Business Report

- 26호 퇴직연금 중심의 근로자 노후소득보장 과제 / 류건식·김동겸 2008.2
- 27호 보험부채의 리스크마진 측정 및 적용 사례 / 이경희 2008.6
- 28호 일본 금융상품판매법의 주요내용과 보험산업에 대한 영향 / 이기형 2008.6
- 29호 보험회사의 노인장기요양 사업 진출 방안 / 오영수 2008.6
- 30호 교차모집제도의 활용의향 분석 / 안철경·권오경 2008.7
- 31호 퇴직연금 국제회계기준의 도입영향과 대응과제 / 류건식·김동겸 2008.7
- 32호 보험회사의 헤지펀드 활용방안 / 진 익 2008.7
- 33호 연금보험의 확대와 보험회사의 대응과제 / 이경희·서성민 2008.9

■ 간행물 _____

○ 보험동향 / 연 4회

○ 보험금융연구 / 연 4회

※ 2008년 이전 발간물은 보험연구원 홈페이지(<http://www.kiri.or.kr>)에서 확인하시기 바랍니다.

『 도서 회원 가입 안내 』

회원 및 제공자료

	법인회원	특별회원	개인회원
연회비	₩ 300,000원	₩ 150,000원	₩ 150,000원
제공자료	- 연구보고서 - 기타보고서 - 연속간행물 · 보험금융연구 · 보험동향 · KIRI 포커스 모음집 · KIRI 이슈 모음집 · KOREA INSURANCE INDUSTRY	- 연구보고서 - 기타보고서 - 연속간행물 · 보험금융연구 · 보험동향 · KIRI 포커스 모음집 · KIRI 이슈 모음집 · KOREA INSURANCE INDUSTRY	- 연구보고서 - 기타보고서 - 연속간행물 · 보험금융연구 · 보험동향 · KIRI 포커스 모음집 · KIRI 이슈 모음집 · KOREA INSURANCE INDUSTRY
	- 영문연차보고서	-	-

※ 특별회원 가입대상 : 도서관 및 독서진흥법에 의하여 설립된 공공도서관 및 대학도서관

가입문의

보험연구원 도서회원 담당

전화 : (02) 3775 - 9080 팩스 : (02) 3775 - 9102

회비납입방법

- 무통장입금 : 국민은행 (400401 - 01 - 125198)

예금주 : 보험연구원

가입절차

보험연구원 홈페이지(www.kiri.or.kr)에 접속 후 도서회원가입신청서를 작성·등록 후 회비입금을 하시면 확인 후 1년간 회원자격이 주어집니다.

자료구입처

서울 : 보험연구원 자료실 (02-3775-9115 / cbyun@kiri.or.kr)

저 자 약 력

이 진 무(Gene Moo Lee)

University of Texas at Austin 컴퓨터과학 박사

University of British Columbia 경영정보시스템학과 교수

(E-mail: gene.lee@sauder.ubc.ca)

연구보고서 2018-15

빅데이터 기반의 사이버위험 측정 방법 및 사이버사고 예측모형 연구

발행일 2018년 9월

발행인 한 기 정

발행처 **보 험 연 구 원**

서울특별시 영등포구 국제금융로 6길 38

화재보험협회빌딩

대표전화 : (02) 3775-9000

조판및
인 쇄 고려씨엔피

ISBN 979-11-85691-92-3 94320

979-11-85691-50-3 (세트)

정가 10,000원