Journal of The Korea Institute of Information Security & Cryptology
VOL.30, NO.5, Sep. 2020
ISSN 1598-3986(Print)
ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2020.30.5.835

AI를 통한 BEC (Business Email Compromise) 공격의 효과적인 대응방안 연구

이 도 경,[1†] 장 건 수,[2] 이 경 호[3‡]
[1,3]고려대학교(대학원생,교수), [2]두산중공업(과장)

A Study on the Effective Countermeasure of Business Email Compromise (BEC) Attack by AI

Dokyung Lee,[1†] Gunsoo Jang,[2] Kyung-ho Lee[3‡]
[1,3]Korea University(Graduate student, Professor), [2]DHI&C(Manager)

요 약

이메일을 통해서 거래처나 경영진을 사칭하여 금전이나 민감한 정보를 탈취하는 BEC(Business Email Compromise) 공격이 빈번하게 발생하고 있다. 이러한 공격 형태는 최근 발생한 무역사기 중 가장 큰 비중을 차지하며 FBI에서 추정 2019년 피해 금액만 약 17억 달러에 이른다. 하지만 이에 비해서 기업들의 대응 실태를 살펴보면 전통적인 SPAM 차단시스템에 의존하고 있어 보다 치밀해져만 가는 BEC 공격에는 사실상 무방비 상태이며, 임직원에 관련 사고를 안내하고 주의를 당부하는 변화관리 수준의 대응에 머물고 있다. 이에 BEC 사고 유형과 방법들을 분석하고, AI를 통한 기업에서의 효과적인 BEC 공격 대응방안을 제안하고자 한다.

ABSTRACT

BEC (Business Email Compromise) attacks are frequently occurring by impersonating accounts or management through e-mail and stealing money or sensitive information. This type of attack accounts for the largest portion of the recent trade fraud, and the FBI estimates that the estimated amount of damage in 2019 is about $17 billion. However, if you look at the response status of the companies compared to this, it relies on the traditional SPAM blocking system, so it is virtually defenseless against the BEC attacks that social engineering predominates. To this end, we will analyze the types and methods of BEC accidents and propose ways to effectively counter BEC attacks by companies through AI(Artificial Intelligence).

Keywords: Business email compromise, BEC, SCAM, Social engineering, email attack, Machine Learning

I. 서 론

최근 기업들의 보안 수준은 과거와 비교하여 상당히 높아졌다. 대부분 국내 대기업들은 방화벽을 설치하여 외부의 공격을 차단하고 접근제어기술을 통해서 서버로의 악의적인 접근을 차단한다. 그리고 실시간으로 공격 트래픽을 분석하는 등 기본적인 관제 시스템을 운영하고 있다. 이러한 변화로 과거 비교적 손쉽게 공격이 가능하던 서버 중심의 시스템을 직접 공격하기 더욱 어려워졌다.

그 결과 최근의 기업 사이버 공격은 서버를 공격하기보다는 비교적 손쉽게 공격이 가능한 사용자를 노리는 공격이 급증하였으며 공격의 목적 또한 시스템 중단이나 대규모 정보 탈취의 목적보다는 개인을

Received(07. 01. 2020), Modified(09. 22. 2020), Accepted(10. 07. 2020)
† 주저자, leedkarmy@korea.ac.kr
‡ 교신저자, kevinlee@korea.ac.kr(Corresponding author)

속여 금전적 이득을 취하려는 형태의 공격이 주를 이룬다.

이런 기업의 사용자를 공격하기 위해서 가장 많이 사용되는 공격 수단으로 이메일이 주로 사용되고 있으며 이메일의 경우 상대적으로 위장하기 쉽고 한 번에 다수를 공격할 수 있는 특성으로 인해 간단한 공격으로도 기업에 큰 피해를 주고 있다. 초기 이메일 공격은 대량의 광고 메시지를 전달하는 SPAM이나 ID/PW 탈취 목적의 피싱 공격 위주에서 앞서 이야기한 금전 취득을 목적으로 하는 BEC[1] 비즈니스 이메일 침해 공격으로 변화하고 있다.

2019년 FBI의 발표에 따르면 미국에서 발생한 BEC 공격의 피해액은 2019년 17.7억 달러로, 작년 12억 달러, 2017년 6억 달러로 해마다 그 피해가 급격하게 증가하고 있으며 사이버 범죄 중 피해액 규모에서 압도적인 1위를 차지하였다. 공격으로 인해 피해를 본 기업만 23,000개가 넘을 정도로 심각한 상태에며 매해 신고 접수도 100% 이상으로 해마다 증가하고 있다 [1].

BEC 공격 피해 사례를 살펴보면, 2014년 LG화학의 240억 피해, 2018년 한국에너지기술연구소의 1억 원 송금 피해, 2020년 미래에셋 홍콩법인 60억 피해와 더불어 해외 구글, 페이스북까지도 BEC 송금 피해의 희생양이 되었다. 주로 큰 금전적 이득을 취할 수 있는 대기업이 공격의 주된 대상이 되었으며 많은 금전 거래가 발생하기 때문에 상대적으로 공격의 표적이 되기 쉬우며 또 큰 규모의 거래가 이루어지기 때문에 금전적으로 한 번에 수백억의 큰 이익을 취할 수 있다는 점에서 공격의 표적이 되고 있다.

표적의 대상인 기업들은 매년 사이버보안을 위해서 상당한 투자를 하고 있으며 외부 침해 공격에 대응하기 위한 시스템 방어체계와 체계적인 프로세스를 갖추고 있다. 하지만, 더욱 교활하고 치밀해져 가는 BEC 공격은 현 SPAM이나 악성코드 탐지 위주의 방어체계를 쉽게 통과하여 무력화시키며 기업의 피해를 주고 있으며 대부분 기업이 효과적으로 대응하지 못하고 있다.

본 논문은 AI(Artificial Intelligence)를 통한 BEC 공격의 효과적인 탐지방안을 제시를 목적으로 하며, 기업의 비즈니스 이메일 및 BEC 공격의 실질 데이터를 기반으로 공격 형태와 특징을 분석 공격행위의 특징을 패턴화하고 Machine Learning 기술을 통해 최적의 알고리즘을 탐색 및 적용하여 탐지 시스템을 구현함으로써 BEC 공격에 대해 AI 기반의 효과적인 대응방법을 제안한다.

1) Business Email Compromise 비즈니스 이메일을 통한 금전 취득 목적의 침해 공격

II. 연구 배경 및 선행연구

2.1 연구 배경

미 연방 수사국(FBI)의 내부 범죄 대응 센터(IC3)가 2019년 발표한 Cyber Crime Report에 따르면 2015년부터 2019년까지 발생한 피해액은 전체 102억 달러, 46만 7361건이 신고 되었다 [1].

이 중 BEC 공격 비중은 17.7억 달러로 압도적으로 가장 높았으며 참고로 최근 기업에서 이슈화되는 Ransomware 공격의 피해액이 0.89억 달러로 피해 규모 면에서 압도적인 차이를 보인다.

글로벌 보안 전문기업 Symantec, CISCO, Trend micro 그리고 국내 기관 KISA, 금융보안원 등 수많은 보안 관련 기관에서 이메일을 통한 BEC 공격 위협을 가장 심각한 사이버보안 이슈 및 공격 Trend로 다루고 있다 [2][3][4].

매년 큰 피해가 발생하며 많은 보안 기관에서 BEC 공격에 대한 사이버위협을 경고하고 있지만 해마다 피해액은 급증하고 있으며 기존의 SPAM 위주의 기업 방어체계로는 대응이 어려운 상황이다. 또한, 현재 대응방안으로 제시되는 방안들이 대부분 임직원 가이드 수준의 변화관리에 그치며 기술적 측면에서는 SPF[2] 적용을 통한 도메인 검증 방식을 제안하고 있으나 실질적인 차단 효과가 크지 않아 기업에 적용 가능한 효과적인 차단 방안 제시가 시급한 상황이다.

2.2 선행연구

이메일 공격 위협에 관한 연구는 1990년대 들어 SPAM 메일에 대한 탐지를 중심으로 연구가 시작되었으며 2000년대 들어서는 피싱 공격 중심의 사회공학 기법이 유행하면서 사회공학 피해 사례나 공격 기법에 관한 연구들이 함께 활발히 이루어졌다 [5][6][7]. 연구들을 살펴보면, URL 빈도나 본문의 키워드, 평판을 이용한 블랙리스트나 패턴을 중심으로 스

2) Sender Policy Framework 발신자 도메인 검증

팸메일을 필터링하는 기술적 보호 방안과 사전 승인을 통한 발송만을 허용하는 옵트인(Opt-in), 수신 거부 의사에 따른 거부 옵션을 적용한 옵트아웃(Opt-out), 실명인증, 특별법 제정 등 규제를 통한 해결 방법들이 주로 연구되었다.

그리고 최근 들어서는 스팸 필터링의 정확도를 높이고 효과적으로 운영하기 위해서 AI에 관한 연구들이 시도되고 있다. 선행 연구된 'DNN을 이용한 텍스트 기반 스팸 메일 필터링' 논문에서는 AI 기술 중 하나인 DNN(Deep Neural Network)를 이용한 텍스트 기반 스팸 메일의 필터링을 위한 방법을 제안하고 메일 본문 학습을 통한 필터링 방법을 구현하였다.

하지만 위 연구들은 대부분은 일반적인 사용자 관점에서의 상업용 광고 메일에 대한 대응이나 URL, 사이트 속이는 형태의 악성코드에 대한 기술적 조치에 관한 연구들로 광고성 메일이나 악성코드를 필터링하는 데 초점이 맞춰져 있어 경영진이나 거래처를 주로 사칭하는 BEC 공격에는 효과적이지가 못하다.

비즈니스 이메일 공격에 관한 연구는 2006년 Nigerian SCAM 또는 419 SCAM으로 불리는 비즈니스 무역 사기가 유행하면서 연구가 시작되었고 사고 사례에 대한 분석 중심의 연구가 이루어지고 있다 [8][9][10].

Naver 및 Google의 학술자료를 참고하여 Email과 관련된 보안 연구들을 정량적으로 분석해 본 결과, 1990년 후반에 들어 이메일이 중요한 커뮤니케이션의 수단이 되고 급속도로 이용이 늘면서 보안 이슈들도 함께 부상하며 연구가 활발해졌고 2007년에는 관련 연구가 130건에 이를 정도로 연구의 정점을 이루었다. 이후의 연구들은 좀 더 분야가 세분되며 사회공학 기법이나 SCAM 등으로 전문화되며 꾸준하게 관련 연구들이 이루어지고 있다 [27][28].

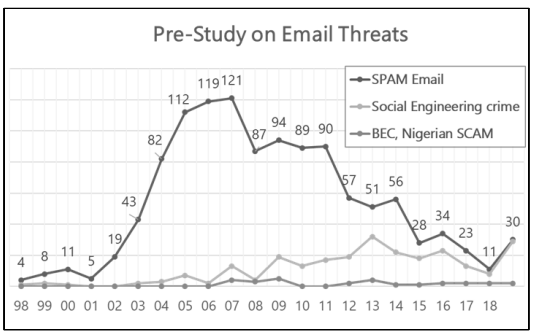

Fig. 1. Pre-Study on Email Threats [27][28]

본 연구에 핵심이 되는 인공지능과 관련한 연구를 살펴보면 보안 연구의 대부분은 Malware analysis에 초점 맞춰져 있으며 이메일과 관련해서는 Phishing Filtering 관점의 악성 URL이나 본문의 Contents 학습을 통한 탐지방안이 연구되었다 [11].

SPAM이나 Malware 중심으로 연구들이 이루어지고 있으며 BEC 공격에 대해서는 BEC 공격의 큰 상위 범주에 해당하는 사회공학적 기법이라는 측면에서 주로 연구가 되고 있고, 기업을 특정하여 공격 대상으로 하는 BEC 공격에 대한 구체적인 대응방안에 관한 연구는 찾기 힘든 실정이다.

2017년 발표된 Detection of business email compromise 논문에서는 이메일 수신 과정에서 개인 이메일, 핸드폰 번호 등의 부가 정보 확인 과정의 인증 프로세스를 추가하여 예방하는 방안을 제안하였으나 최근 공격자 대부분은 발신자의 2차 정보 등을 인지하고 있어 통제를 우회할 수 있는 한계점과 한번 정상 메일로 인증받은 발신자 주소로 메일 헤더 주소를 변경하여 공격하는 경우 대응이 쉽지 않은 한계점을 가지고 있다 [12].

Business email compromise and executive impersonation 논문에서는 BEC 공격 위협과 위험성을 주제로 인식 관점에서의 연구가 주로 이루어졌으며 실무 관점에서 구체적으로 기업이 다루어야 할 대응방안에 대한 부분은 언급되지 않고 있다 [13][14].

2.3 연구 차별성

본 논문은 BEC 이메일 공격 데이터를 기준으로 공격 형태와 특징을 분석하여 공격행위의 특징을 패턴화하고 Machine Learning 기술을 통해 최적의 알고리즘을 탐색 및 적용함으로써 AI 기반의 BEC 침해 공격에 대한 실질적인 탐지방안을 제시한다는데 기존 연구와 차별성을 갖는다.

2.3.1 기존 Machine Learning

Machine Learning은 Spam과 달리 Phishing과 분야에는 거의 사용되지 않고 있는데 일부 연구에서 적용한 대표적인 논문 Classification of Phishing Email Using Random Forest Machine Learning

Technique [26] 연구를 살펴보면, Random Forest를 통해 URL이나 Link의 특정 IP 주소나 문자 형태, 메일 도메인의 특수문자 포함 여부, HTML 본문, 자바스크립트 존재 여부, 본문에 자주 사용되는 키워드 그룹을 특성으로 지정하여 blacklists와 heuristics를 조합하는 방식으로 피싱 메일을 탐지하였다. 학습에는 2천 개의 데이터를 사용하였고 97%의 높은 탐지율을 보였다.

하지만, BEC 공격에서는 대부분 AWS나 Azure와 같은 정상 서버에서 동작하는 일반 메일과 동일한 형태의 공격이 대부분으로 위 연구에 사용된 기법의 URL이나 본문의 특성을 이용한 탐지방법으로는 BEC 공격을 효과적으로 구분해 낼 수가 없는 한계점을 확인할 수 있었다.

기존 보안 제품의 경우도, 머신러닝 기술이 작년부터 활발히 적용되고 있으나 Black List 방식으로 URL, 본문 키워드, 악성코드 Signature, 신고된 IP, 신고된 메일 주소를 기반으로 학습하여 탐지하고 있으며 앞선 연구와 같이 정상적인 메일 서버에서 Header를 변조하여 들어오는 공격에는 효과적으로 방어하지 못했다.

본 연구에서는 글로벌 Top Vendor 제품의 머신러닝 기반의 Email 및 ATP 방어체계를 구현하고 있는 기업을 대상으로 임직원 10명에게 금전 요청, 자료 요청 12건의 공격을 시도하였고 공격 메일은 100% 지인에게 전달되었으며 80% 이상 공격이 성공하여 원하는 목적을 달성할 수 있었다.

헤더 변조 프로그램을 제작하여 일반 메일 형태로 공격을 시도하였고 이러한 경우, 기존 Email 및 ATP에서 이상 여부를 탐지하지 못하였다. Black List에 포함되지 않은 정상적인 서버 IP와 정상적인 메일 주소, 내용을 바탕으로 일상 메일로 공격이 이루어졌기 때문에 방어체계를 우회할 수 있었고 또한 메일이 통과하여 당사자에게 수신된 경우 사전 내부 정보를 수집한 내용을 바탕으로 메일 내용을 구성하고 관계자로 발신자를 조작하였기 때문에 쉽게 공격

Table 1. BEC Attack Test

Devision	Attack Try	Mail pass	Accident Damage
Money request	5	5 (100%)	4 (80%)
Data Request	7	7 (100%)	6 (85%)

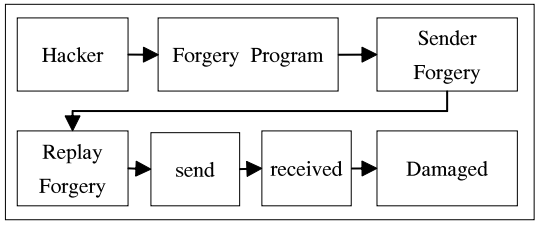

Fig. 2. Attack Flow

에 성공할 수 있었으며 Header 변조를 통해 만들어낸 사칭 주소도 공격받는 처지에서는 조작되었음을 구분하지 못했다.

공격은 BEC 공격 사례를 참고하여 시나리오를 구성하였고 Header 변조 프로그램 제작 및 외부 서버를 구축하여 실 공격 환경과 같은 조건으로 공격을 진행하였다.

From Name은 사칭할 표시 이름, From Email은 사칭 이메일로 정상 메일과 동일하게 설정, Reply To는 공격 성공 후 회신받을 메일 주소를 입력하였다.

아래 메일은 공격에 사용된 실제 공격 이메일이다.

내부 진행되는 상황을 이용해서 메일을 작성하여 공격하였고, 발신자 주소와 표시이름까지 정상메일과 동일하기 때문에 표면상 구분할 수 없으며 더욱이 발신자 사진까지도 동일하게 표시되기 때문에 메일 상에서 차이점을 찾을 수가 없었다.

테스트 결과 사회공학 기법인 BEC 공격에 대해서는 특정 URL, 도메인, IP 등의 특성 위주의 Data를 기반으로 하는 기존 탐지방법으로는 효과적으로 공격을 탐지할 수 없었으며 공격 메일이 수신된 경우 임직원의 관점에서 메일의 Interface에서 정상메일과 BEC 공격 메일의 차이를 확인할 수 없었다.

Fig. 3. BEC Attack Program

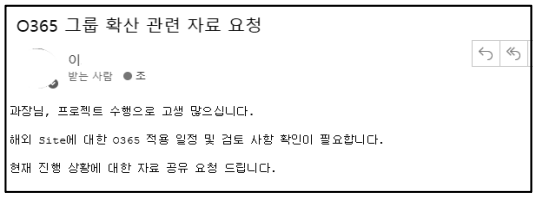

Fig. 4. BEC Attack Mail Example

해당 공격을 살펴보려면, 메일의 속성에 Header를 하나하나 분석해야 공격임을 확인할 수 있어서 일반 임직원이 수신 메일 상에서 공격을 구분해 내기는 사실상 불가능에 가까웠다.

2.3.2 제안 모델

본 연구에서는 기존의 악성 URL, IP, Domain의 특성 학습을 통한 Black List 방식의 탐지 방식이 아닌 Mail의 상관성에 기반을 두어 Header의 특징을 구분하여 학습 모형을 설계하였다.

① 메일의 송수신 관계 분석

Sender 중심으로 메일의 수발신 관계를 학습할 수 있도록 설계하여 변형된 공격자의 침입을 식별할 수 있게 하였다.

A와 B의 메일 송수신 관계에서 만들어진 특성을 X, Y, Z라고 했을 때 X'와 같은 변형된 특성이 확인되었을 때 공격으로 탐지하게 된다.

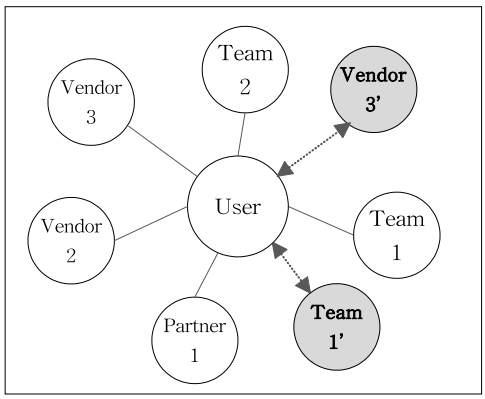

Fig. 5 User Impersonation

② Mail Path Verification

메일 송수신 과정에는 Sender와 Receiver의 IP 뿐만 아니라 메일 서버들을 거치며 Path들을 형성

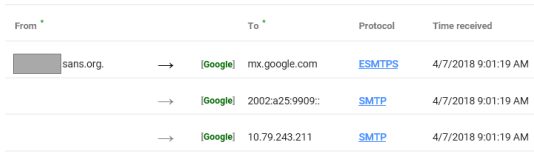

Fig. 6. Mail Path Analysis

하게 된다. 메일이 발신되고 수신되는 과정에서 일련의 Path를 이어서 학습할 수 있게 함으로써 Path를 기반으로 하는 관계를 만들어 낼 수 있다.

A와 B가 메일을 송수신하는 과정의 Path가 X라고 했을 때, 만약 공격자가 Header를 변조하여 발송하더라도 발신 서버가 다르므로 Path가 변하게 된다.

A와 B의 정상 메일을 기준으로 Mail의 Path를 그렸을 때, 외부에서 공격용 서버를 구축하여 A로 속여 발송하였을 시 Path 상에 경로 및 전체 Path 상의 Hop Count도 변하였다. 예로 A 회사의 메일과 Gmail과의 관계에서 발생하는 경로상의 Hop이 3이라고 했을 때 AWS에 메일 서버를 구축하여 발송한 경우 경로상의 Hop은 7로 변하게 된다.

③ User/Domain Impersonation

BEC 공격에 상당 부분은 유사한 메일 주소를 생성하여 사칭하는 공격이다. 메일의 주소에 숫자나 영문자 O를 숫자 0으로 변경하는 등 본 주소와 유사한 계정을 생성하여 공격을 감행한다. 이러한 공격을 차단하기 위해서 Sender의 주소를 학습하여 유사한 String에 관한 판단이 가능하도록 하였다. String의 차이에 관한 확인을 위해 Levenshtein Distance Algorithm을 사용하였고 기존에 정상적으로 학습된 Sender의 주소 값이 1025223라고 할 때 1025227의 주소가 들어오게 된 경우 발신 주소가 매우 유사함을 판단할 수 있게 된다.

④ SPF의 발신자 검증

Sender Policy Framework 정보에 대한 학습이다. 국내 KISA나 미국의 FBI에서 피싱공격에 대응하기 위해 권장하는 기술이며 해당 기술은 발신자의 서버 정보를 사전에 DNS에 등록하여 발신지의 이상유무를 확인할 수 있게 된다.

하지만, SPF 기능은 기업 내 의무사항이 아니므로 많은 기업들이 사용하지 않고 있으며 또한 해커가 Google이나 Naver와 같은 정상 메일 서버를 이용

하거나 또는 정상적인 메일 도메인 서비스를 통해서 공격하는 경우에는 구분할 수 없다.

그럼에도 불구하고 해당 필드를 학습하는 이유는 발신자의 도메인이 SPF가 적용되었는지 아닌지 그리고 적용되었다면 정상인지 아닌지 구분하여 도메인의 상태 값으로 학습할 수 있어서 연관분석을 통한 관계도를 그리는 데 이점이 있기 때문이다.

⑤ Reply to 변조 여부 학습

BEC 공격의 주요 유형 중 하나로 Reply to 변조 공격이 있다. Header 변조를 통해 발신자, 표시 이름을 변경하여 수신자가 공격을 의심할 수 없게 만들고 나서 필요한 정보를 획득하기 위해 메일의 회신 버튼을 클릭 시, 사전에 설정해 놓은 Reply to의 해커의 메일 주소로 메일이 발신되도록 하는 방식이다.

위 공격을 구분해 내기 위해서, Sender와 Reply to가 같은지 학습할 수 있도록 설계하였으며 Reply to 변경이 정상적인 메일에 대해서도 지도학습을 통해서 사전에 학습하여 판단할 수 있게 하였다.

일부 리서치나 마케팅을 목적으로 의도적으로 Reply to 주소를 수정하여 사용하는 경우가 일부 있으며 이러한 경우라도 전체 관계를 학습하기 때문에 몇 차례라도 관계가 맺어지면 사전 학습 정보를 통해 정상 메일임을 구분해 낼 수 있었다.

위 크게 5가지의 view에서 머신러닝을 활용한 Header의 Feature를 구분하여 학습하였고 그 결과, 기존 시스템에서 탐지하기 어려웠던 BEC 공격에 대한 효과적인 탐지가 가능하였다.

III. AI기반 탐지 모델 제시

3.1 탐지 모델 설계

BEC 공격에 대한 실시간 탐지 체계를 구현하기 위해 그림 7.과 같이 시스템을 설계하였으며 지속적인 패턴 학습을 통한 알고리즘 최적화로 지능화된 공격에 대응, 효과적인 탐지방안을 제시한다.

BEC 공격 탐지 모델은 크게 아래 6단계의 프로세스를 통해 모델링이 이루어진다. (①Data extract ②Data Collection ③Data Pre processing ④Data Analysis ⑤Model Selection ⑥ Evaluation & Application)

BEC 공격을 분석하기 위해서, 수신된 Email Header 데이터 추출하고 알고리즘을 최적화하기 위

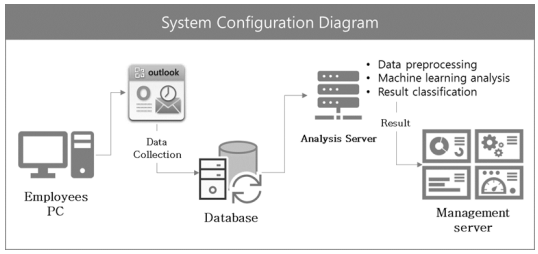

Fig. 7. AI-based Business Email Compromise (BEC) Detection System Diagram

한 Data 분석 및 분류 작업을 통해서 공격 탐지를 위한 주요 Feature를 재정의한다.

최종적으로 학습에 최적화된 데이터 모델을 위한 Feature engineering 작업이 반복되며 ML 학습에 적합한 학습 데이터 형태로 Preprocessing을 진행하게 된다. Preprocessing 완료 후 Model Selection 과정을 통해 최적의 알고리즘을 탐색하게 되고 성능 평가 과정을 통해서 모델을 확정, 탐지 시스템을 구현하였다. 분석 시스템 사양은 아래와 같다.

Table 2. Analysis System

	Sub Group	Content
Platform	OS	Windows 10 10.0.18362
	DBMS	Microsoft Office 2016 Accessdata
Program	Data extract	Visual Studio 2019 C#
	Mail Header Analysis	Google G Suite Toolbox, MS email analyzer
	Preprocessing	Visual Studio 2019 C#
Analysis	File Extraction	Microsoft Office 2016 Excel
	Framework	Weka 3.84
UI	Interface	Microsoft Power BI
Server	Analysis Server	Xeon 3.56 Ghz *2, quadro p2000 * 2, 64 GB (16*4)

3.2 데이터 분석

모든 이메일은 메시지의 상단에는 헤더라고 하는 텍스트 블록이 포함되어 있다. 이 헤더에는 보낸 사람의 정보, 받는 사람의 정보, 보낸 날짜, 전송에 사용한 서버, 메시지 수발신 시 이동한 경로 정보 등

메시지에 관련된 상세 정보를 포함한다. 이런 특성을 고려, 해커의 BEC 공격 메일과 정상 메일 간의 Header 정보의 차이와 공격 메일의 특성들을 학습시켜 공격을 구분할 수 있다. 이런 헤더는 매우 길지만 일관된 구조와 특성을 보인다. 표준 RFC 822에서 기술적 정의와 규약을 다루고 있으며 표준을 통해, 해더의 추출 및 Feature 단위로 특성을 구분할 수 있다 [15].

3.2.1 Data 추출

Exchange 메일 사서함의 수발신 메일들에 대한 Header 정보를 추출하는 과정이다.

추출 과정은 크게 Header를 의미 있는 Feature 단위로 Parsing 하는 단계와 두 번째 Parsing 한 데이터에서 필요한 정보 외에는 삭제하는 Cleaning 단계, 그리고 마지막으로 Database에 저장하는 3단계로 구성하였다.

Email Header에 추출은 Outlook Add-in 모듈로 구현하였으며, Outlook에 프로그램 추가 설치 후 Extractor 버튼 클릭 시 이벤트를 받아 Thread로 헤더 추출 Job을 실행하게 된다.

메일의 Header 정보를 수집하게 되고, Header를 미리 정의한 48개의 field에 따라서 Parsing을 진행한다. Parsing을 완료하게 되면, 해당 Data는 Excel 파일로 추출하게 된 후 다시 Access database에 저장하여 전체 Data Collection 과정을 완료하게 된다.

3.2.2 Data 전처리

앞서 Data Collection에서 처리한 Data Set은 바로 분석할 수 없는 Raw 한 형태의 자료이다. 머신러닝 학습을 위해서 해당 데이터는 전처리 작업을 통해서 Digitalization 해야 한다.

Header에 대한 Digital Number로 변환하는 작업은 String Distance Algorithm 중 하나인 Levenshtein Distance Algorithm을 사용하였으며 [29] 전체 전처리 작업의 구조는 다음과 같다.

1단계 전처리 Data 대상을 선택하는 단계이다. 전처리 대상 Database 파일을 선택하게 되면, Text 정보를 Digital Number로 변환할 Thread를 실행하게 된다.

2단계에서는 Digital Number로 변환하는 작업을 수행하며, Levenshtein Distance Algorithm을 사용하여 Text를 Number로 변환한다 [29].

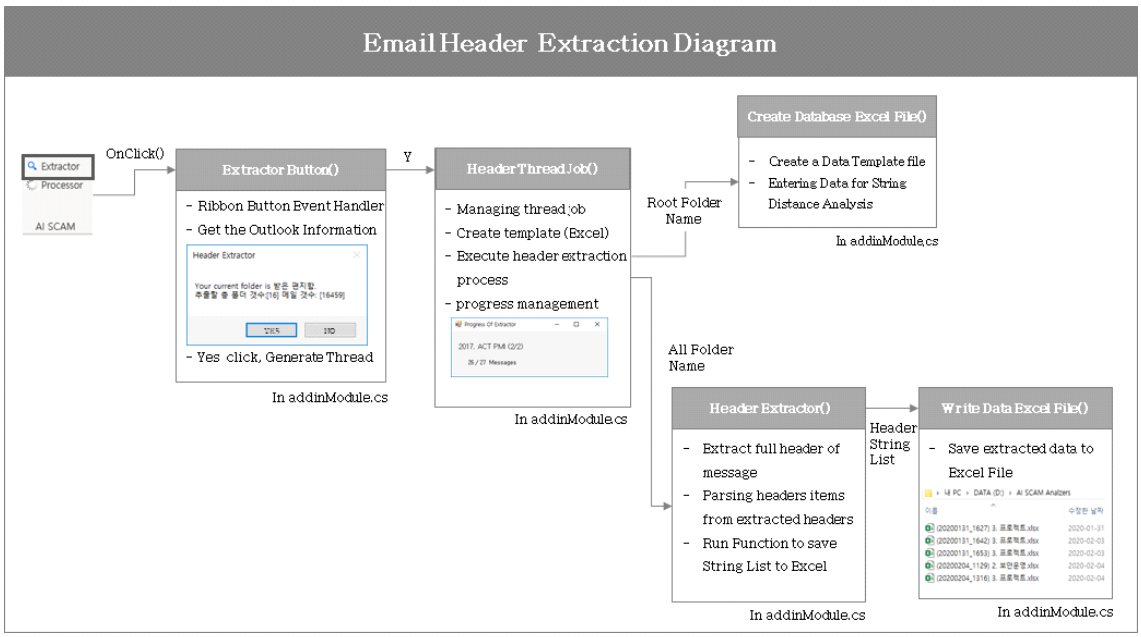

Fig. 8. Email Header Extraction Diagram

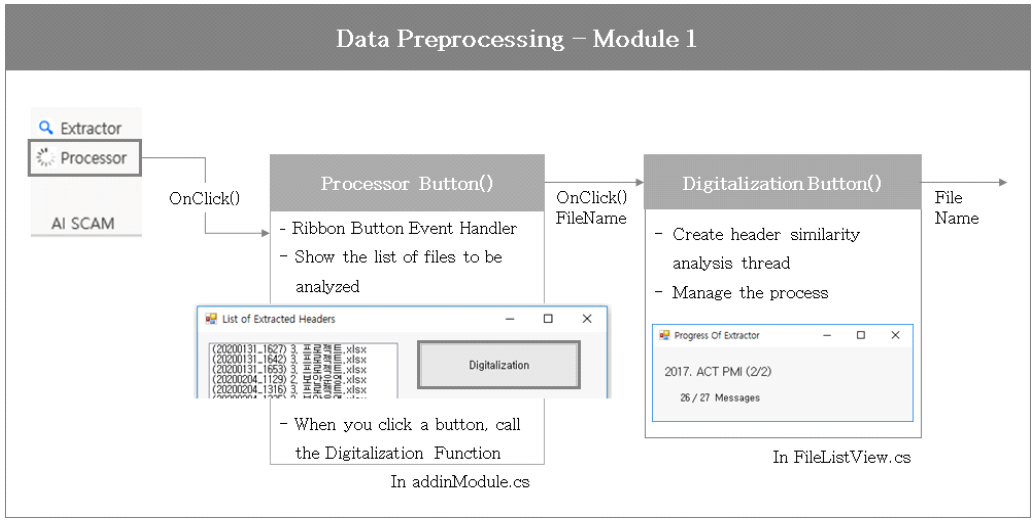

Fig. 9. Data Preprocessing - Phase 1

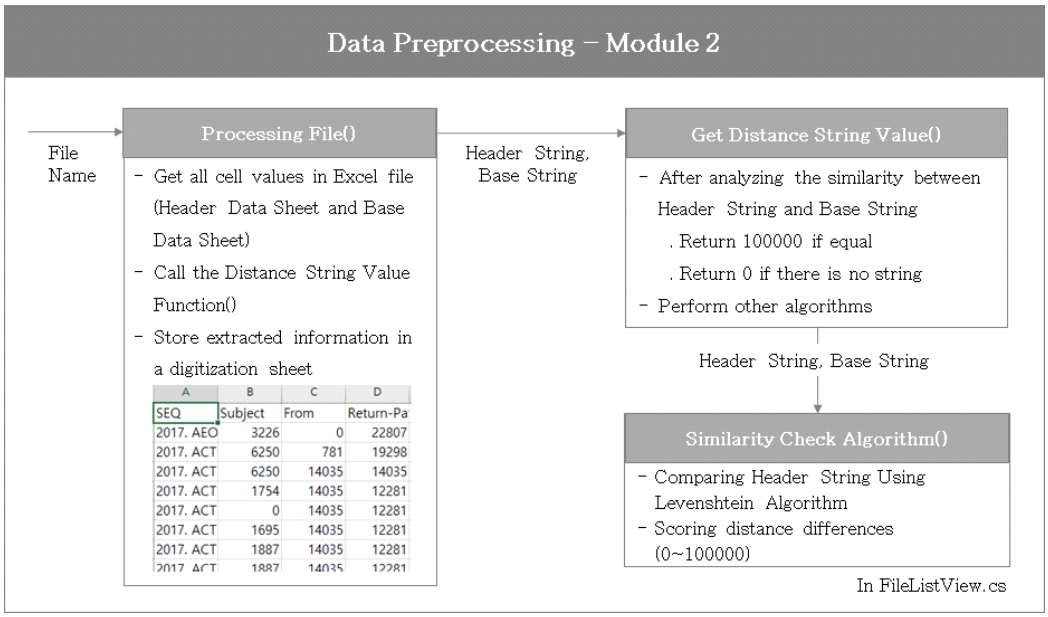

Fig. 10. Data Preprocessing - Phase 2

3.2.3 Data Feature Engineering

본 연구에서는 Business driven Features 관점과 Feature Representation, Error Analysis를 주로 사용하여 모델링을 진행하였다. 최종 48개의 Feature로 Data Set을 재정리하였으며 주요 특성은 다음과 같다.

1) Received 특성을 통해 메일 발송의 경로를 분류하여 정상메일과 사칭메일의 경로상의 Path 차이를 구분

2) Received Path의 Hop을 계산하여 정상메일의 패턴을 학습할 수 있도록 정의. 정상적으로 발신자 B의 메일 Path의 Hop이 8인데 7로 수신된 경우 공격의 가능임을 확인

3) Reply to와 Sender와의 정보가 다른 경우 사칭

임을 의심
4) Domain IP 확인을 통한 비정상 발신자 확인
5) Sender 특성을 통한 유사 메일 비교
6) SPF 특성을 통한 비정상 메일 확인
7) Encoding 특성을 통한 발신자 설정 값 비교

Error Analysis 과정을 통해 재정리된 최적화된 48개의 Feature는 다양한 ML 알고리즘을 적용하여 효과적인 알고리즘을 선택하게 된다.

3.3 Machine Learning 학습

앞서 전처리된 Header 정보는 Exploratory Data Analysis 과정을 반복하여 최적화된 Feature Set을 정의하며 분석 Model을 선정하게 된다.

알고리즘 탐색에는 AI 전문 기업에서 분류한 ML Algorithms Cheat Sheet 2가지를 참고하여 모델링에 사용될 알고리즘을 선정하였고 기업에서 실질적으로 적용되어 사용 가능하도록 속도 측면이나 탐지 예측 모델에 적합한 6가지 알고리즘 (J48, Naive Bayes, REPTree, kNN, SVM, OneR) 선택하였다 [16][17].

IV. 실험 결과 및 고찰

Data set은 국내 제조업 분야 대기업 A의 협조를 받아 비식별화된 Data Preprocessing 과정을 거친 실험 형태의 학습데이터로 추출하여 사용하였다. 전체 Data Set은 108,699개로 Training Data - 6만5천개, Test Data - 4만3천개로 60%:40%로 비율을 조정하여 실험을 진행하였으며 실험데이터는 아래와 같은 형태로 이루어져 있다.

알고리즘 평가는 크게 Accuracy, Precision,

Table 3. Confusion Matrix [19]

		Predict	
		Positive	Negative
Actual	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

PK_Value	Subject	From	Return-Path	Reply-To	Auth~	SPF	Path Hop	Path -1	Path-2
188235	20833	188235	161765	280702	43127	251594	2	157360	215190
164706	41667	164706	184615	245614	46440	262500	3	162437	199367
152941	83333	152941	210526	228070	64257	176796	4	167513	162437
152941	41667	152941	158730	210526	50725	259494	4	172589	203463
129412	91954	129412	119403	192982	37464	237500	2	152284	195918
117647	83333	117647	174603	175439	57554	265432	4	157360	197479
105882	20833	105882	175000	157895	41995	248731	3	157360	228395
141176	20833	141176	55556	154930	48110	227027	3	197970	214286
152941	62500	152941	123077	144444	35623	243902	3	203046	217252
141176	20833	141176	155556	140351	40816	267380	2	147208	185455
117647	62500	117647	121951	136752	38647	254144	3	162437	205047
152941	20833	152941	138462	133333	35623	243902	3	177665	213376
117647	41667	117647	106667	122807	42667	218750	3	157360	221184
129412	62500	129412	115942	120000	42553	234568	3	192893	205882
141176	41667	141176	87500	114583	53957	376190	12	747525	464539
141176	62500	141176	140845	109890	43243	264706	3	162437	185950
141176	83333	141176	132353	107143	44199	248485	3	177665	200837
141176	20833	141176	130435	105882	41667	246988	3	162437	203226

Fig. 11. Experimental Data Set

Recall과 학습 및 탐지속도 4가지를 기준으로 전체 평균값을 사용하여 평가하였다. [19]

1) Accuracy(정확도)
True, False 모두 올바르게 예측한 경우를 측정한 값의 지표이다. 표현식은 아래와 같다.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

2) Precision(정밀도)
A로 분류한 것 중 진짜 A가 얼마나 되는지 확인할 수 있는 지표이다. 표현식은 아래와 같다.

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

3) Recall(재현율)
Recall은 실제 A를 얼마나 A로 잘 예측했는지 확인할 수 있는 지표이다. 표현식은 아래와 같다.

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

4) Time (학습 및 처리까지 소요시간)

4.1 실험 결과

본 연구에서 제안한 머신러닝을 통한 탐지 방식의 모델링 결과, Accuracy가 평균 90% 이상이며 Precision 및 Recall에 대해서도 0.9 이상의 우수한 결과를 확인할 수 있었다. 모델링을 통해, 메일 발신

자를 기준으로 Header를 학습을 시켜 발신 정보가 기존과 다른 경우 탐지가 가능하도록 구현되었으며, 일반적인 Phishing이나 악성코드 탐지의 머신러닝은 각 Data Set 개별 단위로 독립적으로 특성을 학습하는 것이 일반적이나, 본 연구에서 제안한 방식은 메일의 Sender 단위로 관계를 학습할 수 있도록 기준을 제시하고 학습을 진행하였다. 따라서 BEC 공격자인 해커가 유사한 메일 주소나 표시이름 변경 등을 통한 공격을 진행하더라도 사전에 Sender 단위로 학습한 데이터에 따라서 기존 Sender와 해커가 변조한 Sender와의 Gap을 분석하여 탐지율을 획기적으로 높일 수 있었다.

세부적으로 살펴보면, 본 연구에서 BEC 공격을 가장 효과적으로 탐지할 수 있는 모델은 J48 알고리즘이었다. Accuracy 98.5%로 전반적으로 성능이 높았으며 알고리즘 분석 시간도 평균 이상으로 빠르게 처리할 수 있었다. kNN, SVM의 경우 탐지율은 높은 수준이었으나 알고리즘 분석 및 탐지를 위해 걸린 전체 시간이 402초와 2,755초로 매우 처리 속도가 느려 실시간으로 메일을 분석해야 하는 업무환경의 특성을 고려 BEC 공격 탐지에는 적합하지 않은 모델이었다.

Table 4. Analysis of experimental results

	Accuracy	Precision	Recall	Time
J48	98.5%	0.989	0.989	16초
Naive Bayes	67.1%	0.931	0.671	3초
kNN	98.6%	0.987	0.986	402초
SVM	93.6%	0.941	0.937	2,755초
OneR	95.7%	0.955	0.957	3초
평균	90.7%	0.960	0.908	635.8초

4.2 시스템 구현 및 검증

시스템 구현은 정확도 및 처리 속도 등 전반적으로 가장 우수한 성능을 보였던 J48 머신러닝 알고리즘을 통해서 구현하였다.

구현은 프로그램 작업을 통해 Mobile App으로 개발하여 업무에 적용하였고 J48 알고리즘의 빠른 학습

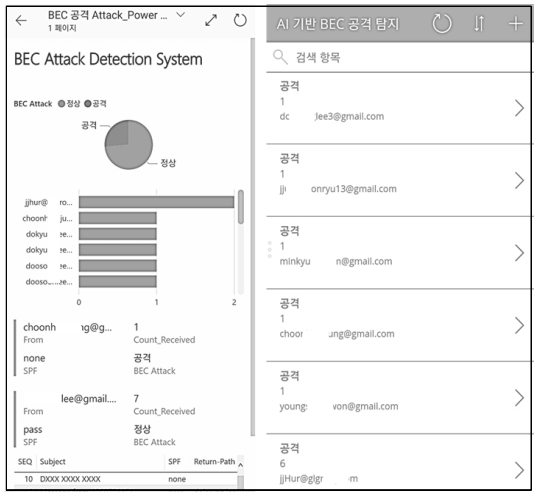

Fig. 12. AI-based BEC attack detection system screen

능력으로 실제 시스템 구현 시, 실시간 BEC 공격 탐지가 가능하며 새로운 공격 패턴에 대해서 탐지 및 자동으로 학습하는 구조로 기존 Rule base나 시나리오 기반의 탐지 방식과 비교하여 월등한 관리적 이점을 확인할 수 있었다.

V. 결 론

본 연구에서는 BEC 침해 공격을 탐지하기 위해 이메일 Header 정보에 대한 머신러닝 학습을 토대로 이상 여부를 탐지할 수 있는 모델을 구축하고 그 효과성을 검증하였다. 제안한 모델의 성능을 측정하기 위해 Accuracy, Precision 및 Recall을 통해 Score 분석을 하였으며 실험 결과 기존 키워드, 평판 중심의 SPAM 차단 시스템이나 악성코드 탐지 중심의 보안 시스템을 우회하는 BEC 공격에 대해서 의미 있는 효과성을 확인할 수 있었다.

본 연구를 토대로, 보다 폭넓고 다양한 Data를 기반으로 모델을 발전시켜 간다면 점점 치밀해지는 BEC 공격에 의한 기업 피해를 보다 최소화하고 효과적으로 대응할 수 있을 것이라고 기대한다.

본 연구는 과학기술정보통신부 및 정보통신기기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음

References

[1] FBI, "FBI IC3 2019 Internet Crime Report," https://www.fbi.gov/news/stories/2019-internet-crime-report-released-021120, Jan. 2020

[2] FSEC, "2020 Cybersecurity Issue Forecast Report," http://www.fsec.or.kr/fsec/index.do, Mar. 2020

[3] KISA, "2019 KISA Cyber Threat Trend Report," KISA,https://www.krcert.or.kr/data/reportView.do?bulletin_writing_sequence=35239, Feb. 2020

[4] Trendmicro, "Business Email Compromise," https://www.trendmicro.com/vinfo/us/security/definition/business-email-compromise-(bec), Mar. 2020

[5] R Sikorski, R Peters, "A privacy primer for the Web: spam, bread crumbs, and cookies," The Journal of the American Medical Association, ISSN 0098-7484, E-ISSN 1538-3598, Apr. 1998

[6] Hinde Stephen, "Spam: the evolution of a nuisance," pp. 474-478, ISSN 0167-4048, Sept. 2003

[7] W.G. Hoover, C.G. Hoover, "SPAM-based recipes for continuum simulations," pp. 78- 85, ISSN 1521-9615 , Mar/Apr. 2001

[8] DuBoff, Leonard D, King, Christy O, Educators Beware, "Avoiding the Scams," pp. 11-13, ISSN 8756-3894, E-ISSN 1559-7075, Mar/Apr. 2009

[9] Deborah Schaffer, "THE LANGUAGE OF SCAM SPAMS: LINGUISTIC FEATURES OF NIGERIAN FRAUD E-MAILS," pp. 157-179, ISSN 0014-164X , Apr. 2012

[10] Peter Ribic, "The Nigerian email scam novel," pp. 424-436, ISSN 1744-9855, Jan. 2019

[11] Blanzieri, Enrico and Anton Bryl, "A survey of learning-based techniques of email spam filtering," Artificial Intelligence Review 29.1, pp. 63-92, July. 2008

[12] JAKOBSSON and Bjorn Markus, "Detection of business email compromise," U.S. Patent Application No 15/414,489, Aug. 2017.

[13] ZWEIGHAFT and David, "Business email compromise and executive impersonation: are financial institutions exposed?," Journal of Investment Compliance, May. 2017.

[14] Remorin, Lord, Ryan Flores, and Bakuei Matsukawa, "Tracking Trends in Business Email Compromise (BEC) Schemes," Trend Micro 18.1, 2018

[15] Wikipedia, "email header Standard," https://tools.ietf.org/html/rfc822, Feb. 2020

[16] BecomingHuman.AI, "AI algorithms," https://becominghuman.ai/, Feb. 2020

[17] Microsoft, "ML Algorithms Sheet," https://docs.microsoft.com/, Feb. 2020

[18] Hyun-Jun Kirn, Jason J. Jung and 0eun-Sik Jo. "Spam-Mail Filtering System Using Weighted Bayesian Classifier," Journal of KIISE: software and usage products 31.8, pp. 1092-1100, Aug. 2004

[19] Wikipedia, "Confusion Matrix," https://en.wikipedia.org/wiki/Confusion_matrix, Mar. 2020

[20] Sango Lee, "Spam-Filtering by Identifying Automatically Generated Email Accounts," Journal of the Society of Information Sciences," Software and Applications 32.5, pp. 378-384, May. 2005

[21] Jindal, Nitin, and Bing Liu. "Review spam detection," Proceedings of the 16th international conference on World Wide Web, pp. 1189-1190, May. 2007.

[22] Markines, Benjamin, Ciro Cattuto,

and Filippo Menczer. "Social spam detection," Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web, pp. 41-48, Apr. 2009.
〔23〕 Barreno, Marco, et al. "The security of machine learning," Machine Learning 81.2, pp. 121-148, Apr. 2010
〔24〕 Buczak, Anna L., and Erhan Guven. "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Communications surveys & tutorials 18.2, pp. 1153-1176, Oct. 2015
〔25〕 Papernot, Nicolas, et al. "Towards the science of security and privacy in machine learning," :1611.03814, Nov. 2016
〔26〕 Akinyelu, Andronicus A and Aderemi O. Adewumi, "Classification of phishing email using random forest machine learning technique," Journal of Applied Mathematics 2014, Apr. 2014
〔27〕 Academic Information, "BEC, SCAM, social engineering," https://academic.naver.com/, Feb, 2020
〔28〕 Scholar, "BEC, SCAM, social engineering," https://scholar.google.co.kr/,Feb, 2020
〔29〕 Blog, "levenshtein algorithm," https://www.cuelogic.com/blog/the-levenshtein-algorithm, Jan. 2020

〈저자소개〉

이 도 경 (Dokyung Lee) 정회원
2007년 8월: 아주대 정보통신학과 졸업
2017년 9월~현재: 고려대 정보보호대학원 석사과정
2007년 12월~2011년: ㈜두산 정보통신 Data Center
2011년~2014년: ㈜두산중공업 보안감사팀
2014년~현재: ㈜두산 지주부문 CDO Security Audit팀
〈관심분야〉 정보보호정책, 머신러닝, 디지털포렌식, Operation Technology 보안

장 건 수 (Gunsoo Jang) 정회원
2009년 2월: 명지대 컴퓨터공학과 졸업
2008년 12월~2017년: ㈜두산 정보통신 Data Center
2017~현재: ㈜두산중공업 Security Audit팀
〈관심분야〉 정보보호 정책, 보안 감사, 개인정보보호 정책, 보안 거버넌스

이 경 호 (Kyung-ho Lee) 종신회원
1989년 8월: 서강대학교 수학과 학사
1997년 8월: 서강대학교 정보통신대학원 석사 졸업
2009년 8월: 고려대학교 정보보호대학원 박사 졸업
2017년 2월~2019년 2월: 고려대학교 정보전산처장
2011년~현재: 고려대학교 정보보호대학원 교수
〈관심분야〉 정보보호 정책, 개인정보보호 정책, 위험관리, 머신러닝, 블록체인