The Transactions of the Korea Information Processing Society
TKIPS Vol.14, No.3, pp.149~160
eISSN: 3022-7011
https://doi.org/10.3745/TKIPS.2025.14.3.149

A Survey of Phishing Campaign Trends and the Classification of Detection Techniques

Ji-Hoon Park† · Sang-Hoon Choi†† · Ki-Woong Park†††

ABSTRACT

Recent phishing attacks have demonstrated notable cost-effectiveness and efficiency through the use of phishing kits and AI technologies, resulting in a substantial rise in phishing efforts. Furthermore, phishing is evolving in complexity as perpetrators leverage psychological weaknesses, enhanced generation methods, and evasion tactics. Credentials obtained via phishing are frequently utilised for subsequent assaults, causing more harm. To safeguard users against phishing assaults, it is essential to examine how contemporary phishing strategies circumvent existing detection systems and to derive insights from the newest research developments. This study categorises phishing detection solutions into four types: URL and domain-based, web page component-based, visual similarity-based, and message content-based detection, while highlighting the challenges faced by each approach.

Keywords : Phishing Campaign, Phishing Detection, Machine Learning, LLM, Feature Extraction

피싱 캠페인의 동향 조사 및 피싱 탐지기법에 대한 분류

박 지 훈† · 최 상 훈†† · 박 기 웅†††

요 약

최근 피싱 공격은 피싱 키트와 AI의 활용으로 인해 저비용 고효율의 특성을 가지고 있으며, 이러한 이유로 피싱 캠페인의 횟수가 점점 증가하고 있다. 또한, 피싱은 피해자의 심리와 다양한 생성 기법 및 우회 기법으로 인해 점점 고도화되고 있으며, 피싱을 통해 탈취한 자격 증명은 부차적인 공격에 이용되어 더 큰 피해로 확산된다. 따라서, 피싱으로부터 피해자를 보호하기 위해 현재 피싱 공격이 어떻게 기존의 탐지기법을 우회하는지 분석하고, 이에 관한 최신 연구 동향의 파악을 통해 인사이트를 도출하는 것이 중요하다. 본 논문에서는 피싱 탐지기법을 URL 및 도메인, 웹 페이지 구성요소, 시각적 유사성, 메시지 콘텐츠로 분류하고, 각 탐지기법이 보유한 도전 과제에 관해 소개한다.

키워드 : 피싱 캠페인, 피싱 탐지, 머신러닝, LLM, 특징 추출

1. 서 론

피싱이란 개인정보와 낚시의 합성어로, 공격자가 전자적인 전송매체를 통해 피해자 스스로 개인정보를 공개하도록 유도하는 사이버 위협 중 하나이다. 피싱은 신뢰할 수 있는 대상으로 가장하거나, 시각적인 교란, 피해자의 감정을 표적으로 하는 등 인적 요소를 포함할 수 있다[1]. 이와 같은 사회공학적 특징을 가진 피싱은 1990년대부터 시작되어 현재까지도 유효한 공격임을 시사한다[2, 3].

현재 피싱 캠페인의 규모와 피해 대상은 지속적으로 증가하고 있으며, 이러한 현상의 주요 원인으로 기술 발전에 의한 피싱 콘텐츠의 자동화와 피싱 캠페인을 수행하는 데 필요한 노력 감소를 식별할 수 있다[4, 5]. 특히, 피싱 캠페인을 위해 사용되는 이메일 및 SMS 등의 메시지, 자격 증명이나 개인정보 입력이 가능한 웹 페이지, 상호작용 할 수 있는 URL과 도메인은 생성형 AI, 머신러닝 등을 사용하여 자동화가 가능하므로 공격에 필요한 노력을 감소시킨다[6]. 또한, 피싱 캠페인을 수행할 수 있도록 백그라운드를 제공하는 피싱 키트는 서

※ 이 논문은 2024년 한국정보처리학회 ACK 2024의 우수논문으로 "피싱 탐지 기법 조사 및 분류: 비정상적 상호작용"의 제목으로 발표된 논문을 확장한 것임.
※ 이 논문은 과학기술정보통신부의 재원으로 정보통신기획평가원(IITP)의 정보보호핵심원천기술개발(Project No. RS-2024-00438551, 30%), 국방ICT융합연구(Project No. 2022-11220701, 30%), 정보통신방송혁신인재양성사업(Project No. 2021-0-01816, 30%), 한국 연구재단(NRF) 중견후속연구사업(Project No. RS-2023-00208460, 10%)의 지원을 받아 수행된 연구임.
† 준 회 원 : 세종대학교 SysCore Lab. 석사과정
†† 비 회 원 : 세종대학교 SysCore Lab. 박사후 연구원
††† 종신회원 : 세종대학교 정보보호학과 교수
Manuscript Received : March 06, 2025
Accepted : March 18, 2025
* Corresponding Author : Ki-Woong Park(woongbak@sejong.ac.kr)

비스형 피싱(Phishing-as-a-Service) 플랫폼을 통해 거래되고 있으며, 피싱 키트에 포함된 피싱 이메일 템플릿, 피싱 웹 사이트 생성, 피싱 타겟 목록 제공과 같은 다양한 기능을 가지므로, 공격 기술에 대한 전문성이 없는 공격자의 피싱 캠페인을 더 수월하게 만든다[7]. 공격자는 이러한 사전 작업을 통해 피싱을 수행하고 피해자의 상호작용을 유도할 수 있으며 이에 대한 작업 흐름을 〈Fig 1〉과 같이 표현한다.

피싱은 인간의 심리를 악용하는 비기술적인 요소와 도메인 생성, 웹 페이지 구성요소와 디자인 변조, 텍스트 변형 기법 등의 기술적인 요소로 이루어져 있어, 알아차리기 어려운 특징이 있다. 공격자가 피싱 캠페인을 위해 SNS 및 공개된 웹 사이트로부터 개인정보 수집이 가능하다는 점과 대규모 언어 모델(LLM)을 활용한 피싱 콘텐츠 생성 및 서비스형 피싱에 의해 피싱 키트가 지속적으로 업데이트된다는 점을 통해 앞으로도 더 발전된 공격이 가능하다는 것을 알 수 있다[8, 9]. 특히, 이러한 현상을 통해 불특정 다수를 대상으로 했던 기존의 피싱 캠페인은 특정 타겟을 대상으로 하는 개인화된 공격으로 전환되거나, 매번 새롭게 생성되는 URL 및 도메인과 난독화된 웹 페이지 구성요소, 대체 또는 섭동이 추가된 메시지 콘텐츠 생성 기법은 기존의 탐지 메커니즘을 우회하는 등 점점 고도화되고 있다[10, 11, 12].

우리는 피싱 탐지기법을 조사하기에 앞서, 최근 피싱 캠페인이 증가하는 현상과 더 고도화된 피싱 캠페인이 수행되는 것을 확인하였다. 이에 따라, 정교한 피싱 캠페인에 대응할 수 있도록 피싱 탐지기법 또한 변화하는 것을 식별하였으며, 기존의 피싱 탐지를 저해하는 요소와 해당 문제점을 개선하는 최신 연구에 대한 인사이트의 필요성을 확인하였다.

따라서, 우리는 피싱 탐지기법을 URL 및 도메인 기반, 웹 페이지 구성요소 기반, 시각적 유사성 기반, 메시지 콘텐츠 기반으로 분류하고 각 항목에 관한 연구 동향을 분석한다. 본 논문의 2장에서 피싱의 개요를 기술하고, 3장에서는 피싱 탐지기법을 4가지로 분류하여 각 탐지기법을 기술적으로 분석한다. 4장에서는 각 탐지기법이 가진 한계점과 그 이유에 관해 분석하며, 5장에서 결론 도출과 향후 연구 방향을 제시한다.

2. 피싱 개요

피싱 캠페인은 일반적으로 명의를 도용한 범죄를 일으키거나 초기 시스템에 침투하기 위해 자격 증명을 수집할 목적을 가지며, 맬웨어 또는 랜섬웨어를 유포하기 위해 수행된다. 이러한 피싱 캠페인은 불특정 다수에게 대량으로 수행되거나, 특정 타겟을 대상으로 수행된다. 특히, 특정 타겟을 대상으로 하는 경우 공격의 효과성을 높이기 위해 공개된 정보나 불법적인 경로로 수집한 자격 증명을 이용할 수 있다[13, 14].

피싱의 피해는 개인정보의 탈취 및 유출에서 끝나지 않고, 공격자들이 추가 공격을 수행할 수 있는 기회를 제공한다. 특히, 피싱을 통해 유출된 개인정보는 다크넷의 포럼에서 유포되거나 마켓 플레이스에서 거래되어, 다른 공격자에 의해 다양한 방식으로 악용될 수 있다[15]. 주요한 악용 방식으로는 명의를 도용한 사기, 클라우드 등의 시스템 침해가 있으며, 악용에 따른 후속 공격으로 개인 또는 기업의 재정적인 손실, 추가적인 계정 및 개인 정보 유출을 야기한다[16, 17].

2.1 URL 및 도메인 생성

피싱 URL은 합법적인 기관 등에 사용되는 URL이 아닌, 공격자가 자격 증명을 탈취하기 위해 생성한 피싱 웹 페이지로 리디렉션되는 악성 URL을 의미한다.

성공적인 피싱 캠페인을 위해, 공격자는 다양한 도메인 생성 기법을 사용할 수 있으며, 생성된 URL은 무작위로 생성된 의미 없는 문자열로 구성되거나, 의미가 있는 문자열 또는 합법적인 도메인 이름을 변형한 형태로 구성되어 기존의 블랙리스트 방식을 우회한다[18]. 또한, 신규 생성된 URL은 피싱 웹 사이트 공유 플랫폼에 업데이트되기 전까지 피싱인지 아닌지 알 수 없으므로 제로데이의 특성을 가진다[19, 20].

2.2 웹 페이지 구성요소

피싱 캠페인은 피해자가 이메일 또는 SMS 메시지 내에 존재하는 URL을 클릭하여 피싱 웹 사이트에 접속하도록 유도하는 과정으로 이루어진다. 특히, 피싱 웹 페이지는 피해자가 자격 증명을 입력할 수 있는 폼으로 구성되거나 기술 지원 및 오류 메시지로 위장한 사기성 팝업을 노출시켜 맬웨어를 설치하도록 유도할 수 있다. 이러한 피싱 웹 페이지는 HTML과 JavaScript, CSS를 통해 구성되며 피해자가 자격 증명을 입력할 때, 입력된 자격 증명이 공격자가 지정한 서버로 전송되는 코드가 포함된다.

공격자는 이메일, SMS 내에 URL을 사용하는 것 외에도,

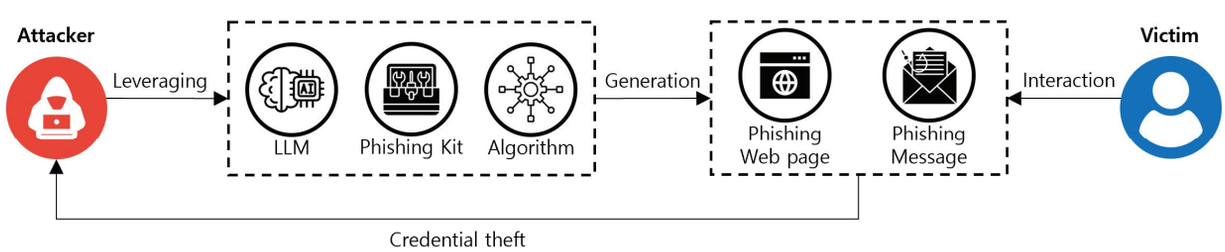

Fig 1. Flow of a phishing campaign

HTML 파일을 첨부하여 피싱을 수행할 수 있다. 특히, 첨부되는 HTML 파일은 EXE, ISO, ZIP, DOCX와 같은 파일보다 덜 의심스럽고, 소스 코드가 난독화되므로 필터링 등의 보안 조치를 우회할 수 있는 특징이 있다[21, 22, 23].

피싱에 사용되는 웹 페이지는 LLM을 사용하여 합법적인 웹 페이지의 HTML 소스 코드, JavaScript, CSS 및 DOM tree 구성요소를 복제할 수 있다[24]. 또한, 피싱 키트는 빠르고 대량으로 피싱 웹 페이지를 생성할 수 있는 이점이 있어, 공격자가 짧은 시간 내에 공격을 수행하고 목적을 달성할 수 있게 된다[25].

2.3 시각적 유사성

공격자가 생성한 피싱 웹 페이지는 합법적인 웹 페이지와 동일하거나, 매우 높은 유사성을 갖도록 모티브가 되는 합법적인 웹 페이지로 가장한다. 특히, 공격자는 피싱 웹 페이지를 구성할 때 잘 알려진 업체의 로고, 디자인을 사용하거나 시각적인 특징을 이용할 수 있으며, 생성된 피싱 웹 페이지는 인간의 시각적으로 구별하기 어려우므로, 피해자를 쉽게 속일 수 있다[26, 27].

공격자는 피싱 웹 페이지를 더 정교하게 구성하기 위해 글꼴, 레이아웃, 색상을 모방한 피싱 로그인 페이지를 생성하여 자격 증명 입력을 유도할 수 있으며, 유사한 도메인이 포함된 URL을 사용하여 공격의 효과성을 높일 수 있다[28].

2.4 메시지 콘텐츠 생성

피싱 메일은 2024년을 기준으로 전년 대비 856% 증가하였으며, 이에 대한 주요인으로는 공격자가 피싱 캠페인의 시도를 위해 ChatGPT 등의 LLM을 악용하는 것이 있다[29].

LLM은 인간과 유사한 텍스트를 생성하는 데 특화되어 있으며, 공개된 웹 사이트로부터 개인의 관심사와 같은 정보를 포함하여 개인화된 스피어피싱 이메일을 생성할 수 있다. 특히, LLM을 사용한 피싱 메일 생성 방법은 프롬프트 엔지니어링과 자동화를 통해 기존 피싱 메일을 생성하는 데 소모되는 비용을 비약적으로 감소시킬 수 있으므로, 공격자는 저비용 고효율로 피싱 캠페인을 수행할 수 있다[8].

피싱 캠페인의 효율성 증가 외에도, 공격자는 피싱 캠페인의 성공률을 높이기 위해 설득, 흥미, 호기심, 공포, 급박함 등의 감정적인 요소가 포함된 텍스트를 이용하여 자격 증명을 입력하도록 유도한다[30]. 이러한 인적 요소가 포함된 메시지 콘텐츠는 일반적으로 필터링되지 않으며, 적대적 텍스트 생성 기법의 경우 기존의 탐지기법을 우회하므로 이메일, SMS에서 사용되는 메시지 콘텐츠 기반의 탐지를 어렵게 한다[31, 32].

3. 피싱 탐지기법의 기술적 분류

본 장에서는 피싱 캠페인에서 탐지할 수 있는 요소를 URL 및 도메인, 웹 페이지 구성요소, 시각적 유사성, 메시지 콘텐츠로 분류한다. 우리는 피싱 요소별 대상과 특징 추출 방법 그리고 최신 연구들에 대해 분석하였고 결과는 〈Table 1〉과 같다.

3.1 URL 및 도메인

URL은 메시지 전송매체를 통해 직접 전송되거나 이메일 내 텍스트, 웹 사이트 및 웹 페이지 등 다양한 콘텐츠 내에 포함되어 피해자를 피싱 웹 사이트로 리디렉션시키기 위해 사용된다. 특히, 피싱 캠페인에서의 URL은 공격자와 피해자 간 상호작용에서 시작 부분이 되므로, 탐지의 중요성이 증대된다.

이러한 특징을 가진 피싱 URL은 도메인 생성 알고리즘(DGA)을 통해 블랙리스트에 없는 도메인을 대량으로 생성할 수 있으므로 기존의 접근법을 우회한다[33]. 또한, Anti Phishing Working Group(APWG)은 2021년 1분기부터 피싱 웹 사이트의 83%가 HTTPS 프로토콜을 사용하고 있음을 보고하여 HTTP 프로토콜을 통해 피싱 URL을 식별하는 것이 더 이상 유효하지 않음을 시사한다[34].

기존 접근법의 한계점을 개선하고자 URL과 도메인에서 피싱을 효과적으로 탐지하기 위해 URL 구성요소가 가진 특징을 분석하였으며, 예시는 〈Fig 2〉와 같다. 이러한 URL 및 도메인의 구성요소를 기반으로 정상 URL 및 피싱 URL의 연관성을 분석한 연구와 이러한 특징을 기반으로 모델을 학습하여 피싱인지 아닌지 탐지하는 연구 또한 진행되었다[35].

1) 단어 임베딩 기반의 탐지

기존의 피싱을 탐지하기 위해 블랙리스트 방식이 사용되었다. 하지만, 최근 발생하는 피싱 DGA을 통해 블랙리스트에 없는 URL을 무작위로 생성하여 탐지를 어렵게 한다. 특히 DGA를 통해 도메인을 생성할 때, 인기 있는 브랜드 또는 서비스와 매우 유사한 도메인을 생성하여 등록함으로써 피해자가 정상적인 기관과 상호작용하고 있다고 믿게 만든다. 이러한 DGA를 통해 생성된 URL 및 도메인이 피싱인지 아닌지 탐지하기 위해, 최근 Word2Vec을 활용한 방법이 연구되고 있다[36].

Word2Vec은 단어를 벡터화하여 단어 간의 의미적, 문법적인 관계를 나타내는 데에 뛰어나며, 이를 도메인분석에 적용한 Dom2Vec이 연구되었다[37]. 해당 연구에서는 Alexa의 합법적인 도메인 데이터 세트와 Corebot 및 Gozi와 같은 DGA 패밀리로부터 생성된 도메인[38], Phishistorm[39]의 피싱 및 합법적인 URL 데이터 세트를 통해 학습을 수행한다.

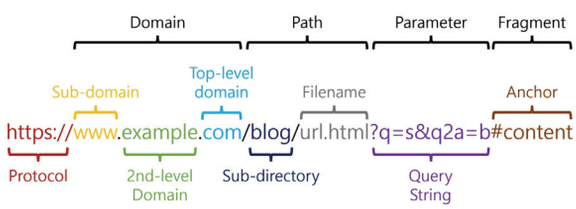

Fig 2. Parts of a URL example

Table 1. Taxonomy of phishing detection technique research

Category	Target	Method	Research	Accuracy
URL and Domain	URL Features	Static Analysis	[35]	97.52%
			[64]	98.72%
			[80]	90%
		Word2Vec	[36]	90%
			[37]	95%
		Jaccard Index	[39]	94.91%
		CNN	[46]	98.2%
		CNN, Hybrid	[78]	95.41%
		AutoEncoder	[47]	91.24%
		ResNet, AutoEncoder	[48]	98%
		BPE Tokenization	[76]	99.67%
		CGRU	[77]	99.61%
		LSTM	[81]	98.50%
	Certificate, Domain Features	Static Analysis	[41]	98%
		Bi-LSTM	[42]	89%
Webpage Content	HTML Source Code	Dynamic Analysis	[50]	98.25%
		TF-IDF	[52]	98.48%
		CNN	[53]	93%
			[54]	98.1%
		CNN, Hybrid	[78]	93.64%
		BERT, MLP	[55]	97.18%
		Static Analysis	[80]	90%
	HTML DOM tree	CNN	[57]	98.84%
		CNN, OCR	[58]	99.4%
		Word2Vec, Hybrid	[78]	90.30%
Visual Similarity	Logo Image, Screenshot	CNN	[60]	81.03%
			[63]	85%
		Static Analysis	[64]	98.72%
	Logo Text	OCR	[65]	98%
			[66]	99.13%
Message Content	Word	TF-IDF, Tokenization	[70]	98.25%
			[83]	99.2%
		Morphological Analysis	[71]	96.85%
		Word2Vec, Tokenization	[72]	99%
	Context	ChatGPT	[68, 84]	87%

Dom2Vec은 dom2words 단계에서 입력된 도메인을 알려진 단어의 집합으로 분할하거나 단어가 없는 경우엔 토큰 집합으로 분할하고, Word2Vec에 임베딩을 수행하여 최소, 최대, 평균을 포함하는 5개의 풀링 계층으로 입력된다. 특히, 단어 빈도와 역문서 빈도(TF-IDF)를 도메인에 적용하여 도메인 내에서 등장하는 단어의 빈도와 전체 도메인 리스트에서의 희소성을 고려하여 특정 단어에 대한 중요도를 찾는다.

제안된 접근법은 LightGBM 기반 모델을 통해 DGA와 Phishistorm URL에 대한 피싱 탐지를 수행하였으며, 높은 재현율과 정확도를 달성하였다. 도메인 이름 내에서 의미적인 관계를 식별해내는 Dom2Vec은 DGA의 반복적인 패턴과 특정 문자의 사용을 쉽게 감지할 수 있다. 하지만, 타이포스쿼팅과 같은 방법에 의해 합법적인 도메인의 이름이 일부 치환되는 경우와 서브도메인을 사용한 경우, 특수문자가 포함된 경우는 탐지 성능이 낮아질 수 있다.

2) 도메인 이름 및 인증서를 활용한 탐지 방법

피싱 캠페인에 사용되는 URL은 기존 HTTP 프로토콜 대신 HTTPS 프로토콜을 사용하는 추세로, 피해자가 합법적인 웹사이트와 상호작용하고 있다고 믿게 만든다[40]. 특히, 공격자

는 만료된 인증서, 침해되거나 알려지지 않은 root CA로부터 발급받은 인증서, 인증서 폐기 정보가 없는 인증서를 통해 HTTPS를 악용하므로, 이러한 의심스러운 인증서를 식별하기 위한 연구가 진행되었다[41].

이러한 고도화된 피싱 전략에 대응하기 위해, URL의 도메인 이름에 추가로 인증서를 활용하여 탐지하는 Unmasking Phishers가 연구되었다[42]. 해당 연구팀은 피싱 웹 사이트 아카이브인 PhishTank[43]와 인기 있는 상위 도메인 목록을 제공하는 Tranco[44], 인증서의 Subject Alternative Name (SAN, 주체 대체 이름)을 수집하기 위한 Censys[45]를 사용하여 머신러닝 기반의 시스템을 제안하였다.

제안된 시스템은 인증서, SAN에서 중요한 특징을 추출하였으며, 특히 Bi-LSTM 모델을 통해 도메인 이름을 벡터화하고 새로운 특징을 생성하여 탐지 성능을 향상시킨다. 도메인 이름 및 인증서로부터 추출된 특징을 통해 랜덤 포레스트(RF), K-최근접 이웃(K-NN), 서포트 벡터 머신(SVM) 분류기를 평가하였으며, 이 중 RF는 89%의 정확도로 피싱 웹 사이트를 식별하여 가장 뛰어났다. 또한, 특징 엔지니어링의 구성요소에서 Bi-LSTM기반의 특징과 도메인 길이 및 복잡성, 인증서 유효기간의 길이가 중요함을 보인다.

3) 오버샘플링 및 앙상블을 활용한 피싱 탐지

시간의 흐름에 따라 복잡해지고 증가하는 피싱 공격은 신속한 탐지, 높은 탐지 효율성 및 정확도가 요구된다. 공격자는 피해자의 자격 증명을 수집하기 위해, 피해자를 피싱 웹 사이트로 유도하고 이러한 과정에서 URL이 사용된다. 이러한 URL은 DGA를 통해 대량 생성이 가능하고 블랙리스트를 우회할 수 있어, 이에 대응하기 위한 URL 기반의 탐지 연구가 진행되었다[46, 47].

목록에 없는 피싱 URL이 끊임없이 증가하는 문제를 해결하기 위해 URL에서 피싱 공격과 관련된 패턴을 신속하게 식별하는 RNT-J 모델이 설계되었다[48]. 해당 연구에서는 URL의 서브도메인 유무, URL 내 .(dot)의 개수, URL 길이와 같은 특징이 포함된 Kaggle의 피싱 데이터 세트를 사용하였으며, 정상적인 URL 데이터 세트와 피싱 URL 데이터 세트의 불균형 문제로 인해 오분류되는 것을 방지하고자 SMOTE 기법을 사용하여 데이터 세트의 균형을 맞춘다[49].

또한, RNT-J 모델은 분류 성능을 향상시키기 위해 오토 인코더와 ResNet으로부터 추출된 특징을 결합하여 피싱과 관련된 패턴을 효과적으로 식별하고 지속적인 파인 튜닝을 통해 탐지 정확도를 향상시킨다. 이와 같이 다양한 프로세스를 거치는 RNT-J는 높은 정확도를 달성하고 처리 시간을 단축시킴으로써 실시간 탐지에 효과적임을 보인다.

3.2 웹 페이지 구성요소

피싱 웹 페이지는 파일 첨부를 악용한 HTML 스머글링 또는 URL이 포함된 메시지 콘텐츠를 통해 피해자와 상호작용한다[22]. 또한, 피싱 웹 사이트는 피해자가 자격 증명을 입력할 수 있는 로그인 페이지와 입력된 자격 증명을 공격자의 서버로 전송하기 위한 코드가 내장된다. 이러한 요소들은 공격자에 의해 난독화되어 피싱인지 아닌지 탐지를 어렵게 한다[50].

피싱을 위한 웹 페이지는 LLM을 통해 생성 및 복제할 수 있으며, 서비스형 피싱에 의해 서비스되어 지속적인 위협이 된다[7]. 특히, 모티브가 존재하는 피싱 웹 페이지는 원본 브랜드 및 기업의 웹 페이지와 유사한 디자인을 사용하여 피해자를 더 쉽게 속일 수 있으므로, 웹 페이지의 구성요소로부터 피싱 여부를 탐지하기 위해 다양한 특징들이 사용된다[51].

웹 페이지를 구성하고 구축하는데 사용되는 HTML 소스코드와 레이아웃을 구성하는 CSS, 웹 페이지를 트리 구조로 표현하는 DOM tree, 내부 및 외부 Javascript의 개수는 정상 및 피싱 웹 페이지에서 차이를 보이기 때문에 웹 페이지 구성요소를 기반으로 한 피싱 탐지에서 사용된다[52].

1) MLP 및 NLP 기반의 HTML 구성요소 탐지

웹 페이지를 생성할 수 있는 새로운 기술로 인해, 피싱 웹 페이지의 구성요소 분석을 통한 탐지가 중요해지고 있다. 특히, HTML 콘텐츠의 특징을 기반으로 피싱을 탐지하기 위해 HTML 문서 내 문자와 단어의 의미를 분석한 연구가 진행되었다[53, 54].

공격자는 피해자의 자격 증명을 탈취하기 위해 HTML 콘텐츠를 이용하므로, 이는 정상 웹 페이지와 피싱 웹 페이지를 구분하기 위한 중요한 특징이다. 이러한 HTML 콘텐츠의 특징을 분석하기 위해 기학습된 다층 퍼셉트론과 자연어 처리 모델을 통합한 MultiText-LP가 연구되었다[55].

해당 연구에서는 Alexa로부터 얻은 양성 데이터와 OpenPhish[56]에서 수집한 피싱 데이터를 통해 데이터 세트를 구축하고, HTML을 구성하는 페이지 제목 및 콘텐츠와 같은 텍스트 데이터와 하이퍼링크, CSS 및 JavaScript, 로그인 폼, 페이지 레이아웃과 콘텐츠의 특징을 추출한다.

제안된 MultiText-LP의 구성요소 중 하나인 기학습된 자연어처리 모델은 텍스트적인 특징을 처리하고, 이 외 숫자와 링크 및 페이지 레이아웃과 같은 특징은 다층 퍼셉트론 모델에 의해 처리되어 복잡한 패턴을 학습한다. 이후 모델의 출력으로 생성된 임베딩을 결합함으로써 MultiText-LP에 입력된 특징이 피싱인지 아닌지 분류한다.

2) HTML DOM 구성요소 분석을 활용한 탐지

일반적인 피싱 웹 사이트는 합법적인 URL의 패턴, 디자인을 모방하지만, HTML의 DOM tree까지 복제하지 않는 차이가 존재한다. 특히, 합법적인 웹 페이지의 경우 HTML DOM tree에서 head, body, table과 같은 요소가 잘 조직화 되어있는 반면, 피싱 웹 페이지는 더 단순하고 불규칙한 구조를 가진다. 이러한 HTML DOM tree 구조의 복잡성에서 나타나는 차이점은 피싱 탐지에서 중요한 요소로 작용하므로, 합법적인

웹 페이지와 피싱 웹 페이지의 HTML DOM tree 구조를 비교하여 피싱을 탐지하는 연구가 수행되었다[57].

하지만, 공격자는 HTML DOM의 구성요소인 CSS와 JavaScript를 통해 시각적인 요소와 동작 또한 모방할 수 있어 시그니처 기반의 피싱 탐지기법을 우회하므로, 이를 해결하기 위해 광학 문자 인식(OCR)과 휴리스틱이 사용된, 트랜스포머 기반의 인코더가 연구되었다[58].

해당 연구에서는 공격자가 탐지를 회피하기 위해 타이틀 태그 텍스트 사이에 문자를 입력하고 HTML DOM tree 텍스트 내 문자의 유니코드 표현을 변조하는 것을 식별하였다. 또한, 폼 필드 변조를 통해 원본 웹 페이지의 동작을 모방할 수 있고 이미지의 픽셀, 밝기, 색상에 미세한 변조가 가해지는 것을 확인하였다.

변조된 특징을 분석하기 위해 웹 사이트에 대해 다양한 시그니처를 보유한 UCI 피싱 데이터 세트[59]와 PhishTank[44]를 사용하였으며, HTML DOM 구성요소의 변조를 탐지하기 위한 타이틀 태그 텍스트, HTML DOM tree 텍스트, 폼 필드, 이미지 및 로고의 특징을 활용한다. 해당 특징을 통해, 타이틀 태그 텍스트 내의 문자를 결합하여 단어를 생성하는 작업, OCR을 통해 텍스트의 원래 유니코드 표현을 추출하여 원본 유니코드 표현으로 복원하는 작업, 캔버스 태그의 픽셀 데이터 재생성 및 원본 이미지와 복원하는 작업을 수행한다. 이러한 휴리스틱 기반의 특징을 임베딩하여 트랜스포머 인코더의 입력으로 사용함으로써 피싱 웹 페이지를 탐지한다.

3.3 시각적 유사성

공격자는 유명하거나 알려진 웹 페이지와 동일하게, 또는 부분적으로 일치하도록 피싱 웹 페이지를 디자인하여 피해자의 자격 증명 입력을 유도한다. 이러한 특성을 가진 피싱 웹 페이지는 모티브가 되는 페이지와 시각적으로 유사해야 하므로 안정적인 특징이 된다.

시각적 유사성을 기반으로 피싱 웹 페이지를 탐지하기 위해 스크린샷과 로고가 사용된다[60]. 특히, 피싱 웹 페이지는 피해자가 시각적으로 인식하는 것에 영향을 미치도록, 합법적인 브랜드와 웹 페이지의 모양, 로고 등을 복제하는 것을 목표로 한다. 이러한 시각적 유사성은 피해자가 정상적인 상호작용을 하고 있다고 속일 수 있으며, 적대적 생성 모델을 통해 탐지를 우회할 수 있도록 고도화되고 있으므로 탐지의 필요성이 중요해지고 있다[61].

1) 객체 검출을 활용한 시각적 유사성 기반 탐지

고도화된 피싱 캠페인으로 인해, 기존 URL 및 HTML 기반 탐지 방법을 우회할 수 있는 문제가 발생하였으며, 이에 대응하기 위한 노력으로 로고 이미지와 스크린샷을 사용한 시각적 유사성 기반 탐지 방법이 연구되었다[27, 62].

기존 시각적 유사성 기반의 피싱 접근 방식은 HTML 구성요소도 사용되었으나 HTML 코드는 변조가 가능하고 다른 HTML 코드를 사용하여 시각적으로 유사한 페이지를 생성할 수 있으므로, 렌더링된 웹 사이트의 스크린샷을 통해 피싱 탐지 연구가 진행되었다.

이러한 시각적 유사성 기반 탐지의 정확도를 향상시키고 적대적 샘플에 대한 내성을 갖는 모델의 연구가 진행되었으며, 해당 연구에서는 객체 검출을 통해 웹 페이지 전체 스크린샷과 로고를 동시에 활용한 피싱1 웹 페이지 탐지를 수행한다[63].

제안된 접근법은 최초 URL을 입력으로 받고 해당 웹 페이지에 존재하는 스크린샷을 캡처한다. 이후 스크린샷이 객체 감지 모델에 입력되고 웹 페이지 마다 다른 로고 영역을 리사이징을 수행한다. 리사이징된 로고와 입력된 스크린샷의 특징을 추출한 후 각각의 특징을 결합하며, 예측되는 URL을 출력한다. 이후 최초 입력된 URL과 예측 URL을 비교하고, 일치하지 않는다면 피싱으로 분류한다.

2) 광학 문자 인식을 사용한 시각적 유사성 비교

공격자는 크롤링을 통해 원본이 되는 페이지와 유사한 피싱 웹 사이트를 생성하고, 디자인을 변조하여 기존의 시각적 유사성 기반 탐지를 우회할 수 있으므로, 이를 탐지하기 위해 OCR 기술을 사용한 연구가 진행되었다[64, 65].

OCR은 텍스트 추출, 언어 및 문자 체계 인식, 페이지 레이아웃 분석이 가능하므로, 시각적 유사성 기반의 피싱 탐지에 이점이 있다. 또한, OCR을 사용한 시각적 유사성 기반의 피싱 탐지는 다른 탐지기법의 한계점인 소스 코드 난독화 및 구성요소 변조에 강하기 때문에, 원본 웹 페이지에 의존하는 피싱 웹 페이지 탐지에 특화되어 있다.

이와 같은 특징을 가진 OCR과 웹 페이지 내 주요 색상 특징을 추출 기법을 사용하여 피싱 웹 페이지를 탐지하는 Phish-Sight가 연구되었다[66]. 해당 연구는 OpenPhish[57]의 피싱 데이터 세트와 Alexa의 합법적인 데이트 세트로부터 URL에 해당하는 웹 페이지의 스냅샷을 생성한다. 이후 Python의 Colorgram을 사용하여 RGB 색상 팔레트를 추출하고, OCR 도구인 tesseract를 사용하여 타겟팅된 브랜드명을 추출하고 모델에 학습시킴으로써 피싱 여부를 탐지한다.

3.4 메시지 콘텐츠

APWG의 2024년 3분기 보고서에 의하면 이메일, SMS, SNS의 메시지 기능 등을 통해 전송되는 피싱 메시지가 증가하고 있으며, 개인화된 공격 또한 상승하는 추세로 나타난다[5].

또한, 공격자는 피싱의 성공률을 높이기 위해 인간의 심리적인 요소를 악용하는 사회공학 기법을 사용하거나 메시지를 변형하는 등의 우회 기법을 활용하므로, 기존의 필터링만으로는 피싱을 효과적으로 대응할 수 없다[67]. 또한, 피싱 캠페인을 수행하기 위해 메시지를 생성하는 작업은 LLM을 통해 자동화되어 저비용 고효율의 공격이 가능하다[8].

이러한 피싱 메시지는 피싱 캠페인에서 가장 첫 번째로 수

행되는 단계이며, 다양한 매체를 통해 피해자에게 전송되어 악의적인 상호작용을 유도할 수 있다. 특히, 피해자의 자격 증명이 탈취된 이후 초기 액세스 브로커에 의해 거래 또는 공유되거나, 자격 증명의 탈취 및 유출 이후 금전적인 피해, 기업에 대한 공격과 같은 부수적인 피해를 일으킬 수 있으므로 피싱 메시지 콘텐츠를 분석하는 것이 중요하다[15, 16, 17].

1) LLM을 활용한 스피어피싱 탐지

LLM은 피싱 캠페인의 패러다임을 변화시키고 있다. 특히, LLM은 정교하고 개인화된 피싱 이메일을 자동을 생성하는 데에 특화되어있어, 특정 대상에 맞게 설득력있는 메시지를 생성하고 배포하기 위한 노력을 최소화한다. 반면에, LLM은 제공된 텍스트의 맥락적인 뉘앙스를 이해할 수 있으므로, LLM을 사용하여 메시지 콘텐츠에 포함된 악의적인 의도를 포착하기 위한 연구가 수행되었다[68].

해당 연구에서는 이메일의 텍스트 내용을 추출하고 해당 이메일이 악의적인 의도가 있는지 확인하는 쿼리를 모델에 입력한다. 초기 이메일과 함께 입력되는 쿼리는 공격자가 피싱 캠페인을 위해 사용하는 상호성, 일관성, 사회적 증거, 호감도, 권위, 희소성과 같은 설득 전략을 기반으로 구성된다[69]. 이후 모델의 응답으로 출력된 벡터를 프롬프트된 문맥 문서 벡터로 정의하였으며, 벡터 데이터 세트로 구축함과 동시에 해당 벡터를 지도 학습 모델에 입력으로 사용한다.

제안된 LLM은 앙상블 기법을 통해 GPT-3.5, GPT-4, Gemini Pro를 결합하여 환각 현상과 편향성을 최소화하였다. 또한, 가상 인물의 개인 식별 정보(PII)를 활용한 스피어피싱 메일을 자동으로 생성하는 시스템을 통해 데이터 세트를 구축하였으며, 기존 연구와 비교를 통해 LLM을 활용한 추론 기반의 문맥 벡터화가 피싱 탐지에 뛰어나다는 것을 입증하였다.

2) SMS 콘텐츠 기반의 피싱 탐지

메시지 콘텐츠는 단어, 관용어, 구문 및 약어와 같은 비공식언어와 같은 요소로 구성될 수 있다. 특히 메신저 피싱의 경우 이메일과 달리 그 길이가 짧으며, 상대적으로 활용할 수 있는 정보가 적으므로 악성 메시지를 탐지하기 위한 특징의 개수 역시 적은 특징이 있다[70].

또한, 메신저 피싱 데이터 세트에서는 합법적인 데이터보다 피싱 데이터가 매우 적은 데이터 불균형 문제가 발생한다. 이러한 문제를 해결하기 위해 편향 판별 분석을 사용하여 데이터 불균형 문제를 해결하고, 정상 메시지를 단일 메시지 단위에서 확장한 대화 데이터를 사용하여 피싱 여부를 탐지하는 연구가 진행되었으며, 문장의 형태소 분석을 통해 동사와 명사의 특징을 활용한다[71].

이외에도, 스미싱 메시지의 98%가 이전에 보낸 메시지의 변형이라는 통계에 기인하여 메시지의 단어, 문자, 메시지 구조에 특수문자가 포함된 적대적 메시지에 저항하는 연구가 진행되었다[72].

해당 연구는 의도적인 문법 오류, leet(야민정음), 단어 내 특정 문자 대체 및 삭제에 의한 텍스트 회피기법을 차용하여 텍스트 회피 공격 도구(EVA)를 설계하였다[73, 74]. 또한, EVA를 통해 원본 메시지에 섭동이 주입된 적대적 메시지를 생성하고 기호, 문자, 공백과 같은 섭동이 포함된 변형 메시지를 온디바이스 스미싱 분류기에 학습하여 강건성을 평가한다.

4. 피싱 탐지기법의 한계점 분석

우리는 피싱 탐지를 위해 사용되는 요소를 URL 및 도메인, 웹 사이트 구성요소, 시각적 유사성, 메시지 콘텐츠로 분류하였으며, 각 분류별 피싱 탐지의 기존 연구와 최신 동향을 분석하였다.

본 장에서는 현재 연구되고 있는 피싱 연구의 동향에서 분류된 탐지기법이 가진 도전 과제와 한계점을 분석하고 평가한다.

4.1 URL 및 도메인 분석의 한계

URL 및 도메인 분석은 피싱 URL의 짧은 수명으로 인해 피싱 탐지에 많은 제약을 받는다[75, 76]. 특히, 피싱 URL은 오픈 소스 인텔리전스(OSINT) 및 피싱 공유 플랫폼에 등록되지만, 이러한 피드는 피싱 URL의 활성화에서 위협 피드에 등록될 때까지의 시간 간격이 존재하여 등록 시점에 이미 비활성화되어있을 수 있다. 또한, 공격자는 피싱 웹 페이지를 자체적으로 중지시켜 접근을 불가능하게 만드는 방법을 사용하며, 특정 위치 및 IP만 사용하여 접근할 수 있도록 설정할 수 있다.

이러한 피싱 URL의 트렌드를 통해 공격자는 다수의 URL 및 도메인을 미리 생성하고 짧은 시간만 활성화함으로써 탐지를 회피한다는 것을 알 수 있다.

이 외에도, 여러 피싱 데이터 세트는 피드가 중복될 수 있어 다양한 데이터 세트로부터 샘플을 수집할 경우 데이터 세트 간의 중복 문제를 고려하여 전처리하는 과정이 필요하다[77].

4.2 웹 페이지 구성요소 분석의 한계

기존 블랙리스트 및 화이트리스트를 사용한 피싱 탐지기법은 자동화된 생성 기법에 의해 효과가 제한되어, 이러한 문제를 해결하기 위한 HTML 콘텐츠 분석 연구가 이루어졌다. 특히, HTML 콘텐츠의 JavaScript 난독화, 의심스러운 로그인 폼 등의 요소는 악의적인 행위에 이용될 수 있으므로 중요한 특징이 된다.

HTML 콘텐츠만을 사용하여 피싱을 탐지하는 것은 단일 웹 페이지의 특징만 분석하므로 높은 정확도를 가진다. 하지만, 단일 특징 분석 기반의 모델은 적대적 웹 페이지 생성을 통해 탐지를 우회할 수 있다[78].

또한, HTML 콘텐츠에 대한 데이터 세트는 희소하고 최신화되지 않아 다른 연구와의 비교가 어려운 점이 존재한다. 또한, 순수 HTML 콘텐츠에만 초점을 맞춘 기존 연구는 타 분류기법에 비해 현저히 적으므로, 제안된 모델의 데이터 세트 벤

치마킹을 통해 성능을 평가하는 것은 한계가 존재한다[55].

4.3 시각적 유사성 분석의 한계

로고 및 스크린샷, 브랜드, UI 등의 디자인 요소를 사용한 시각적 유사성 분석의 주요 한계는 적대적 생성 기법에 의한 탐지 회피가 있다[28, 79]. 이러한 회피 요소는 디자인의 글꼴 및 대소문자 변경, 배경 색 제거와 같은 변조 기법과 섭동 주입을 사용한 변조 기법으로 분류된다. 또한, 적대적으로 생성된 시각적 요소는 기존 모델의 탐지를 회피할 수 있으므로, 모델 설계 시 강건성을 고려하여야 한다.

또한, 시각적 유사성은 모티브가 되는 디자인 요소에 의존하여 원본 디자인이 변경될 경우 재학습을 수행해야 한다.

이외에도, 시각적 유사성 기반의 탐지기법은 데스크톱과 모바일 환경에서 나타나는 레이아웃, 콘텐츠의 차이와 컴퓨팅 리소스에서의 격차가 존재한다. 따라서, 서로 다른 플랫폼에서 얻은 스크린샷은 비교할 수 없으므로 각각의 플랫폼에 동일한 데이터 세트를 사용할 수 없는 한계가 존재한다[80, 81].

4.4 메시지 콘텐츠 분석의 한계

이메일, SMS, SNS의 메시지 기능 등을 통해 전송되는 메시지는 피싱 캠페인에서 첫 번째로 수행되는 위협이므로 초기 대응이 중요하다. 현재 이메일의 콘텐츠를 분석하기 위한 연구는 자연어처리에 기반한 접근법에 의존하여 특정 단어 개수, 철자 오류 감지와 같은 특징을 활용한다[68]. 이러한 접근법은 개인화된 스피어피싱을 탐지하기 어려울 수 있으며, 탐지 결과에 대한 설명 가능성이 제한된다. 따라서, 개인화된 피싱 캠페인을 탐지하기 위해 문맥에 포함된 피싱 의도를 파악하는 것이 중요하며, 분류 결과에 따른 설명을 제공할 수 있어야 한다.

모바일 환경에서 발생하는 트래픽의 증가에 따라, 모바일 메시지 콘텐츠 분석 또한 중요해지고 있다. 특히, SMS는 제한된 크기를 가지고 있어 짧고 간결한 메시지로 구성되는 특징이 있으며, 이러한 특성을 가진 SMS에서 문장의 구문과 패턴, 긴급성, 의심스러운 링크와 같은 텍스트 특징 분석을 위해 머신러닝이 사용된다[82]. 하지만, 공격자는 피싱을 수행하기 전 피해자의 신뢰를 얻기 위해 대화를 지속할 수 있으므로, 단일 메시지 내에서 표면적인 텍스트 특징을 분석하는 것에는 한계가 있다. 따라서, 대화 기반의 문맥에서 텍스트를 분석하는 것이 중요하다[83].

5. 결 론

우리는 현재 LLM과 피싱 키트가 제공하는 저비용 고효율의 효과로 인해 피싱 캠페인이 증가하고 있음을 식별하였다. 따라서, 사이버 위협의 시작이 되는 피싱 탐지의 필요성을 식별하여 광범위한 피싱 탐지기법의 조사 및 분류를 수행하였으며, 본 연구에서는 피싱 탐지기법을 URL 및 도메인, 웹 페이지 구성요소, 시각적 유사성, 메시지 콘텐츠로 분류하고 적용된 기술과 한계점을 조사하였다.

피싱 탐지기법의 공통적인 한계점은 정상 데이터와 피싱 데이터의 차이로 인해 발생하는 데이터 불균형이 있다. 이를 완화하기 위해 데이터 샘플링 기법 또는 다양한 데이터 세트 사용의 필요성이 있으며, 다양한 데이터 세트를 사용할 때 중복 데이터에 대한 대책이 중요하다.

피싱 캠페인을 위해 생성되는 도메인, 웹 페이지 구성요소, 디자인, 메시지는 변조 및 난독화, 적대적 생성의 대상이 되어 기존 탐지를 우회할 수 있는 문제점이 존재한다. 이러한 제한사항을 통해 피싱 탐지기법에 있어 신속성과 정확성이 요구되는 다양한 도전 과제가 남아있음을 확인할 수 있었다. 향후 연구로, 기존 단어 및 단일 메시지로 피싱을 탐지하는 접근법에서 메시지 내 피싱의 의도가 있는지 판단하기 어려운 문제를 해결하고자 한다. 특히, 메시지 전체에서 나타나는 문맥적인 특징과 공격자와의 상호작용 과정에서 식별할 수 있는 특징을 활용하여 데이터 세트를 구축하고자 한다.

References

[1] M. Jari, “M. An Overview of Phishing Victimization: Human Factors, Training and the Role of Emotions,” in *12th International Conference on Computer Science and Information Technology*, 2022, pp.217-228.

[2] R. Dhamija, J. D. Tygar and M. Hearst, “Why phishing works,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2006, pp.581-590.

[3] Gary Smith, Top Phishing Statistics for 2025: Latest Figures and Trends [Internet], https://www.stationx.net/phishing-statistics.

[4] Eliot Baker and Maxime Cartier, Phishing Trends Report (Updated for 2024) [Internet], https://hoxhunt.com/guide/phishing-trends-report.

[5] APWG, PHISHING ACTIVITY TRENDS REPORT 3rd Quarter 2024 [Internet], https://docs.apwg.org/reports/apwg_trends_report_q3_2024.pdf.

[6] M. Schmitt and I. Flechais, “Digital Deception: Generative artificial intelligence in social engineering and phishing,” *Artificial Intelligence Review*, Vol.57, No.12, pp.1-23, 2024.

[7] Andreea Chebac, What Is Phishing-as-a-Service (PhaaS) and How to Protect Against It [Internet], https://heimdalsecurity.com/blog/what-is-phishing-as-a-service-phaas.

[8] J. Hazell, “Spear phishing with large language models,” *arXiv Preprint arXiv:2305.06972*, 2023.

[9] S. S. Roy, P. Thota, K. V. Naragam and S. Nilizadeh, “From Chatbots to Phishbots?: Phishing Scam Generation in Commercial Large Language Models,” in *IEEE Symposium on Security and Privacy*, 2024, pp.36-54.

[10] A. R Tapsoba, T. F. Ouédraogo, M. B. Diallo and W. B. S. Zongo, "Toward Real Time DGA Domains Detection in Encrypted Traffic," in *Proceedings of the 7th International Conference on Networking, Intelligent Systems and Security*, 2024, pp.1-9.

[11] I. Skula and M. Kvet, "URL and Domain Obfuscation Techniques - Prevalence and Trends Observed on Phishing Data," in *IEEE 22nd World Symposium on Applied Machine Intelligence and Informatics*, 2024, pp.000283-000290.

[12] F. Chen, T. Wu, V. Nguyen, S. Wang, H. Hu, A. Abuadbba and C. Rudolph, "Adapting to Cyber Threats: A Phishing Evolution Network (PEN) Framework for Phishing Generation and Analyzing Evolution Patterns using Large Language Models," *arXiv Preprint arXiv:2411.11389*, 2024.

[13] M. Nadeem, S. W. Zahra, M. N. Abbasi, A. Arshad, S. Riaz and W. Ahmed, "Phishing Attack, Its Detections and Prevention Techniques," *International Journal of Wireless Security and Networks*, Vol.1, No. 2, pp.13-25, 2023.

[14] D. A. B. Villalva, J. Onaolapo, G. Stringhini and M. Musolesi, "Under and over the surface: a comparison of the use of leaked account credentials in the Dark and Surface Web," *Crime Science*, Vol.7, No.1, pp.17, 2018.

[15] J. Hughes et al., "The Art of Cybercrime Community Research," *ACM Computing Surveys*, Vol.56, No.6, pp.1-26, 2024.

[16] Mandiant Intelligence, Why Are You Texting Me? UNC3944 Leverages SMS Phishing Campaigns for SIM Swapping, Ransomware, Extortion, and Notoriety [Internet], https://cloud.google.com/blog/topics/threat-intelligence/unc3944-sms-phishing-sim-swapping-ransomware/?hl=en.

[17] PF. P. E. Putra, A. Zulfikri, G. Arifin and R. M. Ilhamsyah, "Analysis of Phishing Attack Trends, Impacts and Prevention Methods: Literature Study," *Brilliance: Research of Artificial Intelligence*, Vol.4, No.1, pp.413-421, 2024.

[18] A. Rizi, "Divergent Detection: A Comprehensive Study of DGAS Using FNNS for Original, Noise-Modified, and Linear Recursive Sequences(LRS)," Ph.D. Dissertation, Dakota State University, SD, USA, 2024.

[19] M. N. Feroz and S. Mengel, "Examination of data, rule generation and detection of phishing URLs using online logistic regression," in *IEEE International Conference on Big Data*, 2014, pp241-250.

[20] C. M. R. da Silva, E. L. Feitosa and V. C. Garcia, "Heuristic-based strategy for Phishing prediction: A survey of URL-based approach," *Computers & Security*, Vol.88, pp.101613, 2020.

[21] K. Kaushik, G. Sharma, P. Narooka, G. Chhabra and A. Vishnoi, "HTML Smuggling: Attack and Mitigation," in *16th International Conference on Security of Information and Networks*, 2023, pp.1-5.

[22] Kaspersky, HTML Attachments: A Gateway for Malware? [Internet], https://www.kaspersky.com/resource-center/threats/malicious-html-attachments.

[23] Niranjan Gedge and Sijo Jacob, The Anatomy of HTML Attachment Phishing: One Code, Many Variants [Internet], https://www.trellix.com/blogs/research/the-anatomy-of-html-attachment-phishing.

[24] S. S. Roy, K. V. Naragam and S. Nilizadeh, "Generating phishing attacks using chatgpt," *arXiv preprint arXiv:2305.05133*, 2023.

[25] Olga Svistunova and Anton Yatsenko, Phishing-kit market: what's inside "off-the-shelf" phishing packages [Internet], https://securelist.com/phishing-kit-market-whats-inside-off-the-shelf-phishing-packages/106149.

[26] K. L. Chiew, K. S. C. Yong and C. L. Tan, "A survey of phishing attacks: Their types, vectors and technical approaches," *Expert Systems with Applications*, Vol.106, pp.1-20, 2018.

[27] A. K. Jain and B. B. Gupta, "Phishing detection: analysis of visual similarity based approaches," *Security and Communication Networks*, Vol.2017, No.1, pp.5421046, 2017.

[28] F. Ji et al., "Evaluating the Effectiveness and Robustness of Visual Similarity-based Phishing Detection Models," *arXiv preprint arXiv:2405.19598*, 2024.

[29] SOCRadar, Phishing in 2024: 4,151% Increase Since Launch of ChatGPT; AI Mitigation Methods [Internet], https://socradar.io/phishing-in-2024-4151-increase-since-chatgpt.

[30] A. Ferreira and S. Teles, "Persuasion: How phishing emails can influence users and bypass security measures," *International Journal of Human-Computer Studies*, Vol.125, pp.19-31, 2019.

[31] T. Saka, R. Jain, K. Vaniea and N. Kökciyan, "Phishing Codebook: A Structured Framework for the Characterization of Phishing Emails," *arXiv preprint arXiv:2408.08967*, 2024.

[32] B. Liang, H. Li, M. Su, P. Bian, X. Li and W. Shi, "Deep Text Classification Can be Fooled," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 2018, pp.4208-4215.

[33] A. Cucchiarelli, C. Morbidoni, L. Spalazzi and M. Baldi, "Algorithmically generated malicious domain names detection based on n-grams features," *Expert Systems with Applications*, Vol.170, pp.114551, 2021.

[34] APWG, Phishing Activity Trends Report 1st Quarter 2021 [Internet], https://docs.apwg.org/reports/apwg_trends_report_q1_2021.pdf.

[35] M. A. Tamal, K. Islam, T. Bhuiyan, A. Sattar and N. U. Prince,

"Unveiling Suspicious Phishing Attacks: Enhancing Detection with an Optimal Feature Vectorization Algorithm and Supervised Machine Learning," *Frontiers in Computer Science*, Vol.6, No.2024, pp.1428013, 2024.

[36] L. T. Aravena, P. Casas, J. Bustos-Jiménez and M. Findrik, "More than Words is What you Need - Detecting DGA and Phishing Domains with Dom2Vec Word Embeddings," in *8th Network Traffic Measurement and Analysis Conference*, 2024, pp.1-4.

[37] L. T. Aravena, P. Casas, J. Bustos-Jiménez, G. Capdehourat and M. Findrik, "Dom2Vec - Detecting DGA Domains Through Word Embeddings and AI/ML-Driven Lexicographic Analysis," in *19th International Conference on Network and Service Management*, 2023, pp.1-5.

[38] C. Morbidoni, L. Spalazzi, A. Teti and A. Cucchiarelli, "Leveraging n-gram neural embeddings to improve deep learning DGA detection," in *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, 2022, pp.995-1004.

[39] S. Marchal, J. Francois, R. State and T. Engel, "PhishStorm: Detecting Phishing with Streaming Analytics," *IEEE Transactions on Network and Service Management*, Vol.11, No.4, pp.458-471, 2014.

[40] H. Hadan, N. Serrano and L. J. Camp, "A holistic analysis of web-based public key infrastructure failures: comparing experts' perceptions and real-world incidents," *Journal of Cybersecurity*, Vol.7, No.1, pp.tyab025, 2021.

[41] M. AlSabah, M. Nabeel, Y. Boshmaf and E. Choo, "Content-agnostic detection of phishing domains using certificate transparency and passive dns," in *Proceedings of the 25th International Symposium on Research in Attacks, Intrusions and Defenses*, 2022, pp.446-459.

[42] M. Haraldsdóttir, S. Homayoun, E. Lynge and C. D. Jensen, "Unmasking phishers: ML for malicious certificate detection," *Computers & Industrial Engineering*, Vol.198, pp.110652, 2024.

[43] Cisco Talos Intelligence Group, PhishTank [Internet], https://phishtank.org.

[44] V. Le Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczyński and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation," in *26th Annual Network and Distributed System Security Symposium*, 2019.

[45] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey and J. A. Halderman, "A Search Engine Backed by Internet-Wide Scanning," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015, pp.542-553.

[46] S. Y. Yerima and M. K. Alzaylaee, "High accuracy phishing detection based on convolutional neural networks," in *3rd International Conference on Computer Applications & Information Security*, 2020, pp.1-6.

[47] A. Assefa and R. Katarya, "Intelligent phishing website detection using deep learning," in *8th International Conference on Advanced Computing and Communication Systems*, 2022, pp.1741-1745.

[48] F. S. Alsubaei, A. A. Almazroi and N. Ayub, "Enhancing phishing detection: A novel hybrid deep learning framework for cybercrime forensics," *IEEE Access*, Vol.12, pp.8373- 8389, 2024.

[49] Tan, Choon Lin, Phishing Dataset for Machine Learning [Internet], https://www.kaggle.com/datasets/shashwatwork/phishing-dataset-for-machine-learning.

[50] P. Zhang et al., "Crawlphish: Large-scale analysis of client-side cloaking techniques in phishing," in *42nd IEEE Symposium on Security and Privacy*, 2021, pp.1109-1124.

[51] R. Zieni, L. Massari, and M. C. Calzarossa, "Phishing or not phishing? A survey on the detection of phishing websites," *IEEE Access*, Vol.11, pp.18499-18519, 2023.

[52] A. Aljofey et al., "An effective detection approach for phishing websites using URL and HTML features," *Scientific Reports*, Vol.12, No.1, pp.8842, 2022.

[53] C. Opara, B. Wei, and Y. Chen, "HTMLPhish: Enabling phishing web page detection by applying deep learning techniques on HTML analysis," in *International Joint Conference on Neural Networks*, 2020, pp.1-8.

[54] C. Opara, Y. Chen, and B. Wei, "Look before you leap: Detecting phishing web pages by exploiting raw URL and HTML characteristics," *Expert Systems with Applications*, Vol.236, pp.121183, 2024.

[55] F. Çolhak, Mİ. Ecevit, B. E. Uçar, R. Creutzburg and H. Dağ, "Phishing website detection through multi-model analysis of HTML content," *arXiv preprint arXiv:2401.04820*, 2024.

[56] OpenPhish, OpenPhish [Internet], https://www.openphish.com.

[57] J. H. Yoon, S. J. Buu and H. J. Kim, "Phishing webpage detection via multi-modal integration of HTML DOM graphs and URL features based on graph convolutional and transformer networks," *Electronics*, Vol.13, No.16, pp.3344, 2024.

[58] A. Memon and A. A. Manjotho, "APFormer: Anti-phishing transformer for website-phishing detection via joint feature learning," in *International Conference on Engineering & Computing Technologies*, 2024, pp.1-5.

[59] Rami Mohammad and Lee McCluskey, Phishing Websites [Internet], https://archive.ics.uci.edu/dataset/327/phishin

g+websites.

[60] S. Abdelnabi, K. Krombholz, and M. Fritz, "Visualphishnet: Zero-day phishing website detection by visual similarity," in *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, 2020, pp.1681-1698.

[61] L. Wu et al., "Editing text in the wild," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp.1500-1508.

[62] I. F. Lam, W. C. Xiao, S. C. Wang and K. T. Chen, "Counteracting phishing page polymorphism: An image layout analysis approach," in *Advances in Information Security and Assurance: Third International Conference and Workshops*, 2009, pp.270-279.

[63] M. Wang, L. Song, L. Li, Y. Zhu and J. Li., "Phishing webpage detection based on global and local visual similarity," *Expert Systems with Applications*, Vol.252, No.1, pp.124120, 2024.

[64] R. S. Rao and A. R. Pais, "Two level filtering mechanism to detect phishing sites using lightweight visual similarity approach," *Journal of Ambient Intelligence and Humanized Computing*, Vol.11, No.9, pp.3853-3872, 2020.

[65] M. Dunlop, S. Groat and D. Shelly, "Goldphish: Using images for content-based phishing analysis," in *Fifth international conference on internet monitoring and protection*, 2010, pp.123-128.

[66] P. Pandey and N. Mishra, "Phish-Sight: A new approach for phishing detection using dominant colors on web pages and machine learning," *International Journal of Information Security*, Vol.22, No.4, pp.881-891, 2023.

[67] P. Burda, L. Allodi and N. Zannone, "Cognition in social engineering empirical research: A systematic literature review," *ACM Transactions on Computer-Human Interaction*, Vol.31, No.2, pp.1-55, 2024.

[68] D. Nahmias, G. Engelberg, D. Klein and A. Shabtai, "Prompted contextual vectors for spear-phishing detection," *arXiv preprint arXiv:2402.08309*, 2024.

[69] R. B. Cialdini, "The Science of Persuasion," *Scientific American*, Vol.284, No.2, pp.76-81, 2001.

[70] M. Gupta, A. Bakliwal, S. Agarwal and P. Mehndiratta, "A comparative study of spam SMS detection using machine learning classifiers," in *Eleventh International Conference on Contemporary Computing*, 2018, pp.1-7.

[71] S. Kim, J. Park, H. Ahn and Y. Lee, "Detection of Korean phishing messages using biased discriminant analysis under extreme class imbalance problem," *Information*, Vol.15, No.5, pp.265, 2024.

[72] J. W. Seo et al., "On-device smishing classifier resistant to text evasion attack," *IEEE Access*, 2024.

[73] J. Li, S. Ji, T. Du, B. Li and T. Wang, "Textbugger: Generating adversarial text against real-world applications," in *Proceedings of the 26th Annual Network and Distributed System Security Symposium*, 2019.

[74] D. Pruthi, B. Dhingra and Z. C. Lipton, "Combating adversarial misspellings with robust word recognition," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp.5582-5591.

[75] D. Hriday, A. Kulkarni, V. Balachandran and T. Das, "Phish-Blitz: Advancing phishing detection with comprehensive webpage resource collection and visual integrity preservation," in *17th International Conference on COMmunication Systems & NETworkS*, 2025.

[76] P. Maneriker, J. W. Stokes, E. G. Lazo, D. Carutasu, F. Tajaddodianfar and A. Gururajan, "Urltran: Improving phishing url detection using transformers," in *IEEE Military Communications Conference*, 2021, pp.197-204.

[77] P. Vamsi, U. Muthaiah and C. H. Roshan Vardhan, "Defending the digital frontier: URL-based phishing detection extension," in *International Conference on Computational Intelligence in Data Science*, 2024, pp.65-76.

[78] R. J. van Geest, G. Cascavilla, J. Hulstijn and N. Zannone, "The applicability of a hybrid framework for automated phishing detection," *Computers & Security*, Vol.139, No.1, pp.103736, 2024.

[79] Q. Hao, N. Diwan, Y. Yuan, G. Apruzzese, M. Conti and G. Wang, "It doesn't look like anything to me: Using diffusion model to subvert visual phishing detectors," in *33rd USENIX Security Symposium*, 2024, pp.3027-3044.

[80] C. Amrutkar, Y. S. Kim and P. Traynor, "Detecting mobile malicious webpages in real time," *IEEE Transactions on Mobile Computing*, Vol.16, No.8, pp.2184-2197, 2016.

[81] R. S. Rao, T. Vaishnavi and A. R. Pais, "PhishDump: A multi-model ensemble based technique for the detection of phishing sites in mobile devices," *Pervasive and Mobile Computing*, Vol.60, pp.101084, 2019.

[82] M. R. Al Saidat, S. Y. Yerima and K. Shaalan, "Advancements of SMS spam detection: A comprehensive survey of NLP and ML techniques," *Procedia Computer Science*, Vol.244, pp.248-259, 2024.

[83] S. Verma, V. Ayala-Rivera and A. O. Portillo-Dominguez, "Detection of Phishing in Mobile Instant Messaging Using Natural Language Processing and Machine Learning," in *11th International Conference in Software Engineering Research and Innovation*, 2023, pp.159-168.

[84] Y. Chen et al., "A survey of large language models for cyber threat detection," *Computers & Security*, Vol.145, pp.104016, 2024

박 지 훈

https://orcid.org/0009-0004-1434-6288
e-mail : qkrwlgns1325@naver.com
2022년 국립순천대학교 컴퓨터교육과(학사)
2023년~현 재 세종대학교 SysCore Lab. 석사과정
관심분야 : 피싱 탐지, 사이버 위협 인텔리전스, 클라우드 시스템 보안

최 상 훈

https://orcid.org/0000-0002-9549-0887
e-mail : csh0052@gmail.com
2014년 대전대학교 정보보안학과(학사)
2016년 대전대학교 전산정보보안학과(석사)
2023년 세종대학교 정보보호학과(박사)
2023년~현 재 세종대학교 SysCore Lab. 박사후 연구원
관심분야 : 하이퍼바이저, 시스템 모니터링, 시스템 메모리, 악성코드 분석, 딥러닝

박 기 웅

https://orcid.org/0000-0002-3377-223X
e-mail : woongbak@sejong.ac.kr
2005년 연세대학교 Computer Science(학사)
2007년 KAIST Electrical Engineering(석사)
2009년 Microsoft Research, Graduate Research Fellowship
2012년 KAIST Electrical Engineering(박사)
2012년 국가보안기술연구소 연구원
2016년 대전대학교 정보보안학과 교수
2016년~현 재 세종대학교 정보보호학과 교수
관심분야 : 클라우드 시스템 보안, 초고속 보안 시스템, 시스템 인스펙션, 디지털 포렌식