반도체디스플레이기술학회지 제23권 제3호(2024년 9월)
Journal of the Semiconductor & Display Technology, Vol. 23, No. 3. September 2024.

NLP와 PunyCode 변환 기법을 활용한 스미싱 메시지 탐지 시스템 설계 및 구현

국중진[*†]·이건희[*]·김주환[**]·오재혁[**]·박재한[**]·박정은[**]
[*†]상명대학교 전자정보시스템공학과, [**]상명대학교 정보보안공학과

Design and Implementation of a Smishing Message Detection System Using NLP and PunyCode Conversion Techniques

Joongjin Kook[*†], Keonhee Lee[*], Juhwan Kim[**], Jaehyeok Oh[**], Jaehan Park[**] and Jungeun Park[**]
[*†]Dept. of Electronics and Information System Engineering, Sangmyung University,
[**]Dept. of Information Security Engineering, Sangmyung University

ABSTRACT

As smartphones become more integral to daily life, security concerns, particularly regarding smishing, have risen significantly. With the increasing frequency and variety of smishing attacks, current detection methods struggle to provide effective solutions, with most commercial algorithms achieving around a 70% detection rate. This paper proposes a novel approach to enhancing smishing detection accuracy by utilizing Natural Language Processing to analyze message syntax and semantics, combined with URL Punycode conversion and whitelist techniques. The approach focuses on improving detection through comprehensive message analysis, aiming to address the limitations of existing preventive methods. The proposed system offers conceptual improvements in smishing detection strategies, providing a more robust framework for addressing evolving security challenges.

Key Words : English Key Word: NLP , PunyCode, WhiteList, BlackList, Semiconductor

1. 서 론

스마트폰이 일상 생활에서 필수적인 도구로 자리 잡으면서, 이를 겨냥한 다양한 보안 위협도 급증하고 있다. 그 중에서도 스미싱(Smishing)은 문자 메시지를 이용한 피싱 공격의 일종으로, 사용자에게 의심스럽지 않은 링크를 클릭하게 유도하여 악성 코드를 설치하거나, 개인정보를 탈취하는 방식이다. 스미싱 공격은 무료 쿠폰, 은행 또는 공공기관 사칭, 주식 투자 제안 등 다양한 형태로 위장하여 사용자를 속이며, 최근 몇 년간 그 발생 빈도와 피해 규모가 급격히 증가하고 있다.

2021년을 기준으로 스미싱 공격의 발생 건수는 17,841건에 달했으며, 이로 인한 금전적 피해도 1,265억 원을 넘어섰다[1]. 이와 같은 상황에서 스미싱 탐지를 위한 효과적인 기술적 대응이 필수적이며, 이를 위한 연구가 활발히 진행되고 있다[2, 3]. 그러나 이동통신사 등에서 제공하는 기존의 스미싱 탐지 시스템들은 주로 키워드 기반의 단순한 필터링 기법이나 기존의 악성 URL 데이터베이스를 활용한 탐지 방법에 의존하고 있어, 스미싱 메시지의 다양한 패턴을 충분히 반영하지 못하는 한계가 있다. 특히, 새로운 유형의 스미싱 메시지나 정교하게 위장된 URL에 대한 탐지율이 낮아, 탐지 시스템의 성능이 제한적이다.

[†]E-mail: kook@smu.ac.kr

본 연구는 이러한 한계를 극복하기 위해 자연어 처리(Natural Language Processing, NLP)를 활용한 스미싱 메시지 정확도 향상 기법을 제안한다. 제안된 기법은 스미싱 메시지의 구문과 의미를 심층적으로 분석하여, 메시지 내에 포함된 의심스러운 키워드와 패턴을 탐지하는 데 중점을 둔다. 또한, 스미싱 메시지에서 자주 발견되는 악성 URL을 보다 정확히 식별하기 위해 Punycode 변환 및 화이트리스트와 블랙리스트 비교 방법을 사용한다.

이를 통해 기존의 탐지 시스템에서 놓칠 수 있었던 새로운 유형의 스미싱 메시지와 변종 공격을 효과적으로 탐지할 수 있도록 설계되었다.

2. 본 론

스미싱 메시지를 효과적으로 탐지하기 위해서는 메시지 내의 패턴과 특징을 정확히 분석하는 것이 중요하다. 본 연구에서는 자연어 처리(NLP)를 활용하여 스미싱 메시지의 구문과 의미를 분석하고, 이를 통해 의심스러운 키워드와 악성 URL을 탐지하는 방법을 제안한다. 본론에서는 이와 관련된 연구 방법을 상세히 설명한다.

2.1 스미싱 메시지 주요 패턴 분석

Fig. 1과 같은 스미싱 메시지는 사용자에게 신뢰감을 주면서도 특정 행동을 유도하는 몇 가지 주요 패턴을 가지고 있다. 이러한 패턴을 이해하고 분석하는 것은 스미싱 탐지의 핵심이다.

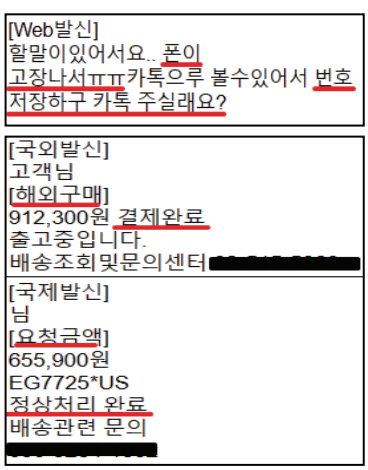

Fig. 1. Examples of smishing SMS.

첫 번째로, 스미싱 메시지에는 의심스러운 키워드가 자주 포함된다. "무료", "긴급", "당첨", "인증"과 같은 단어들은 사용자가 즉각적으로 반응하도록 유도한다. 이러한 단어들은 메시지의 긴급성을 강조하며, 사용자가 경계심을 늦추고 행동을 취하도록 만든다.

두 번째로, 특수문자의 비정상적 사용이 흔하다. 메시지에 특수문자를 삽입하거나 단어를 변형함으로써 필터링을 회피하려는 시도가 자주 발견된다. 예를 들어, "무료"를 "무/료"나 "무!료"로, "인증"을 "인&증"으로 표현하여 정상적인 단어처럼 보이게 한다.

세 번째로, 스미싱 메시지는 특정 행동을 유도한다. 링크 클릭이나 전화 연락을 요구하며, "지금 클릭하세요", "즉시 확인 필요"와 같은 문구로 긴박감을 조성한다. 이는 사용자가 즉각적인 행동을 하도록 압박감을 주기 위한 심리적 전략이다.

마지막으로, 악성 URL이 포함되어 있는 경우가 많다. 이러한 URL은 정상적인 URL처럼 보이도록 위장되거나 축약된 형태로 제공되어, 사용자가 의심 없이 클릭하도록 유도한다. 예를 들어, "bit.ly/2a3BcD"와 같은 축약 URL이나 정상 URL과 유사한 "microsoft.com"과 같은 위장 URL이 사용된다.

이처럼, 스미싱 메시지는 다양한 패턴을 통해 사용자를 속이려 하며, 메시지의 유형 또한 다양하다. 이러한 패턴의 분석은 스미싱 탐지의 중요한 부분을 차지한다. 본 연구는 이러한 패턴을 정밀하게 분석하고, 이를 통해 스미싱 메시지를 보다 효과적으로 탐지할 수 있는 방법을 제시한다.

2.2 NLP 를 이용한 스미싱 메시지 탐지 방법

2.2.1 의심스러운 키워드 탐지

스미싱 메시지에서 의심스러운 키워드를 탐지하는 것은 스미싱 탐지 과정에서 매우 중요한 첫 단계이다. 스미싱 메시지에는 사용자의 즉각적인 반응을 유도하기 위한 특정 키워드가 빈번하게 포함되며, 이러한 키워드를 효과적으로 식별하고 분석하는 것이 필요하다. 본 연구에서는 NLP기법을 사용하여 메시지 내의 의심스러운 키워드를 탐지하고 분류하는 과정을 다음과 같이 제시한다.

먼저, 통신 빅데이터 플랫폼과 Kaggle에서 제공된 실제 스미싱 메시지 1344개의 데이터셋을 기반으로, 메시지 내 자주 사용되는 단어들을 NLP 기술을 활용하여 불용어 처리 및 토큰화를 수행한 후, 단어의 빈도수를 측정하였고 이를 분류하였다. 분류 과정에서는 스미싱 메시지에서 흔히 사용되는 광고성 단어, 사기 유도 단어, 피싱 의심 단어를 대상으로 빈도 분석을 진행하였다. 이러한 단어들은 추가적으로 특수문자 단어, 광고성 단어, 지인 사칭 단어, 국제 문자 단어와 같은 분류 기준에 따라 세부적으로 분류되어 Fig.2와 같은 스미싱 의심 단어 테이블로 작성하였다.

	token	document_fr
1	web발신	109
2	국제발신	26
3	국외발신	19
4	바랍니다	18
5	입니다	14
6	고객님	11
7	광고	11
8	있습니다	11
9	안녕하세요	11
10	완료	11
11	선착순	10
12	하루	10
13	드리겠습...	10

	token	document_fr
1	'발신'	155
2	'확인'	44
3	'완료'	37
4	'원'	33
5	'무료'	30
6	'고객'	29
7	'국제'	27
8	'발송'	25
9	'거부'	22
10	'광고'	22
11	'무의'	20

Fig. 2. Normalization and tokenization of suspicious keywords.

실제 스미싱 메세지가 수신되면 사용된 주요 단어들을 토큰화하여 추출한다. 토큰화는 메시지를 단어 단위로 분할하여 분석하는 과정으로, 이 과정에서 메시지 내의 각 단어가 별개의 단위로 분리되며 분석에 필요한 형태로 변환된다. 이후 단어의 의미를 쉽게 분석하고 일관되게 처리하기 위해 정규화를 수행한다. 정규화 과정에서는 정규 표현식을 사용하여 불용어(분석에 불필요한 단어)를 제거하고, 변형된 단어들을 표준화한다. 예를 들어, "무!료"와 "무료"는 동일한 의미로 처리된다.

이후 단어 빈도 분석을 통해 메시지에서 자주 등장하는 단어들을 식별한다. "무료", "긴급", "당첨", "인증"과 같은 단어들은 대표적인 의심스러운 키워드로, 이러한 단어들이 포함된 메시지는 스미싱 가능성이 높다고 판단된다. 이 과정에서 빈도수가 높은 단어들을 선정하고, 앞서 분류된 스미싱 의심 단어들과 비교한다. 이 과정을 통해 빈도수가 높은 단어가 스미싱 의심 단어와 일치할 경우, 해당 메시지는 스미싱 메시지로 분류된다.

2.3 악성 URL 탐지 알고리즘

2.3.1 악성 URL 분석

스미싱 메시지의 데이터셋 중 URL 링크가 포함된 스미싱 메세지는 84%를 차지한다. 이러한 스미싱 메시지는 Fig. 3과 같은 형태를 포함하고 있으며, URL은 사용자가 클릭했을 때 악성 코드를 설치하거나 피싱 사이트로 유도될 위험이 크다. 따라서, 스미싱 메시지 탐지에서 악성 URL을 정확하게 검출하는 것은 매우 중요하다. 본 연구에서는 스미싱 메시지에서 악성 URL을 탐지하기 위한 효과적인 알고리즘을 제안한다.

기존의 탐지 시스템은 주로 VirusTotal[6]과 같은 외부 데이터베이스를 활용하여 URL을 검출하지만, 그 탐지율은 200건의 스미싱 메시지 중 141건 약 70%에 불과하다. 이는 많은 악성 URL이 탐지되지 않고 사용자를 위협할 수 있음을 의미한다. 본 논문에서는 더 높은 정확도를 제공할 수 있는 새로운 탐지 알고리즘이 필요함을 시사한다.

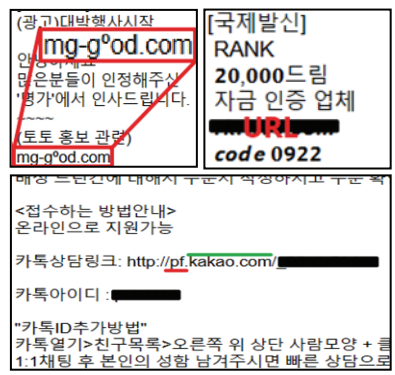

Fig. 3. Smishing SMS with malicious URL.

2.4 악성 URL 검출 알고리즘

미검출 스미싱 메세지를 검출하기 위해서 추가적으로 악성 URL 공격형태의 대표적인 형태를 분할하고 이에 따른 인식률을 높이는 방법을 사용했다. 탐지 정확도를 높이기 위해서는 공격유형별로 각기 다른 탐지 기술이 요구된다. 따라서 본 논문에서는 악성 URL 탐지성능을 향상시키기 위해 Fig. 5에 소개된 가장 높은 비중을 보이는 IDN Homograph Attacks, Domain Spoofing, Subdomain Hijacking 공격에 대해 별도의 탐지 방법 Punycode 변환과 화이트리스트 대조를 적용하였다.

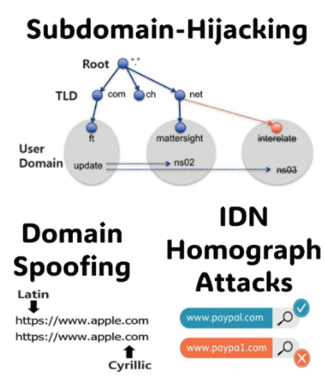

Fig. 4. Techniques for Detecting Malicious URLs.

우선, IDN Homograph Attacks는 라틴어 및 문자와 같은 다양한 문자를 활용한 정상 URL과 시각적으로 동일하게 보이는 URL이다. 도메인은 실제로 네트워크에서 사용될 때 Punycode라는 형식으로 변환된다. Punycode는 ASCII만을 사용하여 비-ASCII 문자를 인코딩 하는 방법이다. xn--로

시작하는 Punycode를 디코딩하여 실제 유니코드 문자로 변환하고 디코딩 된 문자들을 시각적으로 유사한 문자가 포함되었는지 확인한다. 라틴문자 “a” 키릴문자 “а”, 라틴 문자 “o” 그리스 문자 ”ο”, 라틴 문자”i” 키릴문자”і” 등과 같은 예시가 있다. 유사문자의 종류를 유사 문자 매핑 테이블로 Fig. 5와 같은 형태로 구성한 뒤 Punycode로 변환된 URL을 대조하여 악성 URL을 탐지한다.

Similar Character	Mapped Character	Description
`а` (U+0430)	`a` (U+0061)	Cyrillic `а` vs. Latin `a`
`е` (U+0435)	`e` (U+0065)	Cyrillic `е` vs. Latin `e`
`о` (U+043E)	`o` (U+006F)	Cyrillic `о` vs. Latin `o`
`і` (U+0456)	`i` (U+0069)	Cyrillic `і` vs. Latin `i`
`ѕ` (U+0455)	`s` (U+0073)	Cyrillic `ѕ` vs. Latin `s`
`р` (U+0440)	`p` (U+0070)	Cyrillic `р` vs. Latin `p`
`l` (U+006C)	`I` (U+0049)	Lowercase `l` vs. Uppercase `I` (Latin)
`0` (U+0030)	`O` (U+004F)	Digit `0` vs. Uppercase `O` (Latin)

Fig. 5. Similar Character Mapping Table.

다음으로, Domain Spoofing은 공격자가 합법적인 도메인을 위조하거나 비슷한 도메인을 만들어 사용자들을 속이는 공격이다. 이를 검출하기 위해 정상 URL 화이트리스트를 작성하여 이 리스트 외의 도메인에 접근하지 못하도록 제한한다.

마지막으로Subdomain Hijacking은 악의적인 사용자가 서브도메인의 제어권을 탈취하고 이를 악용하는 공격이다. Subdomain Hijacking에 대한 방지 방법은 ‘Subjack’[7], ‘Subover [8]’와 같은 도구를 사용하거나 방화벽 등 다양한 방법이 제시되지만 애플리케이션 환경을 고려해 공인된 도메인들을 모아 화이트 리스트를 작성 하여 적용 하였다. Fig.6은 화이트 리스트의 예시를 나타낸다.

No.	Domain	Allowed Domain
1	`korea.go.kr`	`https://www.gov.kr/`
2	`nhbank.com`	`https://www.naver.com/`
3	`samsung.com`	`https://www.kbcard.com/`
4	`google.com`	`https://www.google.com/`
5	`naver.com`	`https://pay.naver.com/`
6	`daum.net`	`https://banking.kbstar.com/`
7	`hanwha.com`	`https://mail.google.com/`
8	`sktelecom.com`	`https://www.apple.com/`
9	`lg.com`	`https://www.samsung.com/`
10	`kbfg.com`	`https://www.wooribank.com/`

Fig. 6. Whitelist Domain Table.

상용 탐지 시스템이 검출하지 못한 30%의 스미싱 메시지를 검출하기 위해, 대표적인 악성 URL 공격 유형을 분할하여 이에 따른 인식률을 높이는 방법을 적용하였다. 탐지 정확도를 높이기 위해 공격 유형별로 상이한 탐지 기술이 요구되며, 이에 따라 본 논문에서는 IDN Homograph Attacks, Domain Spoofing, Subdomain Hijacking과 같은 주요 공격 유형에 대해 각각의 맞춤형 탐지 기법을 적용하였다.

3. 실 험

3.1 실험 환경

본 연구에서는 제안된 스미싱 메시지 탐지 알고리즘의 성능을 평가하기 위해 400건의 메시지를 사용하여 실험을 진행하였다. 이 중 200건은 스미싱 메시지, 나머지 200건은 정상 메시지로 구성되었으며, 데이터셋은 다양한 패턴과 변형된 URL을 포함하도록 설계되었다. 메시지 데이터셋은 스미싱 메시지 탐지의 모든 측면을 평가할 수 있도록 구성되었으며, 이를 통해 제안된 알고리즘의 성능을 다각적으로 평가하였다.

알고리즘 검증을 위해 Android 기반의 모바일 애플리케이션을 개발하였다. 이 애플리케이션은 스미싱 메시지 수신 시 NLP를 활용해 텍스트와 URL을 분석하고, 이를 바탕으로 악성 여부를 판단하는 기능을 갖추고 있다. 메시지 수신 후, NLP 분석 모듈이 텍스트를 분석하여 의심스러운 키워드와 패턴을 탐지하며, URL은 VirusTotal API, Punycode 변환, 화이트리스트 및 블랙리스트 대조 과정을 통해 검증된다.

실험은 Python 기반의 NLP 라이브러리(NLTK, scikit-learn)와 URL 분석 도구를 사용하여 진행되었으며, 애플리케이션은 실제 외부로부터 스미싱 메시지를 수신하여 실시간으로 탐지 성능을 확인하였다.

3.2 실험 및 결과

3.2.1 의심 키워드의 토큰화 및 정규화

스미싱 메시지에서 자주 사용되는 의심스러운 키워드를 탐지하기 위해, NLP 기반의 토큰화 및 정규화 과정을 활용하였다. 이 과정에서 먼저 메시지 내의 텍스트를 단어 단위로 분할하여, "무료", "긴급", "당첨"과 같은 주요 키워드를 식별하였다. 이들 키워드는 스미싱 메시지에서 사용자가 즉각적으로 반응하도록 유도하기 위해 자주 사용되는 단어들이다.

정규화 과정에서는 특수문자나 불필요한 공백을 제거하여 키워드의 변형된 형태를 동일하게 처리할 수 있도록 하였다. 예를 들어, "무!료", "긴/급"과 같은 변형된 키워드들도 이 과정에서 "무료", "긴급"으로 표준화되어 탐지

정확도가 향상되었다. 이러한 과정을 통해 스미싱 메시지의 주요 패턴을 보다 정밀하게 분석할 수 있었으며, 이를 통해 의심스러운 키워드를 포함한 메시지들을 효과적으로 탐지할 수 있었다.

3.2.2 악성 URL 탐지

스미싱 메시지의 대부분은 악성 URL을 포함하고 있으며, 이러한 URL을 정확히 탐지하는 것은 스미싱 탐지의 핵심 요소이다. 본 연구에서는 Punycode 변환과 화이트리스트 및 블랙리스트 비교를 통해 악성 URL을 탐지하는 방법을 제안하였다.

Fig. 7. Detection result of smishing SMS containing suspicious keywords.

Punycode 변환은 시각적으로 정상적인 URL처럼 보이도록 위장된 악성 URL을 식별하는 데 사용되었고, Domain Spoofing과 Subdomain Hijacking과 같은 다양한 URL 공격 기법에 대응하기 위해, 정상 URL 목록(화이트리스트) 활용하였다. 예를 들어, 합법적인 도메인과 매우 유사한 도메인이나 특정 서브도메인을 탈취하여 악용하는 URL도 Fig. 8과 같이 이 과정에서 효과적으로 탐지되었다.

Fig. 8. Detection result of smishing SMS with malicious URL.

3.2.3 Android 기반 검증 애플리케이션

제안된 탐지 알고리즘을 실제 환경에서 검증하기 위해 Android 기반 애플리케이션을 개발하였다. 이 애플리케이션은 사용자가 스미싱 메시지를 수신할 때, NLP 모듈을 통해 메시지의 텍스트와 URL을 실시간으로 분석하고, 악성 여부를 판단하는 기능을 갖추고 있다.

실험에서는 사용자가 스미싱 메시지를 수신할 경우, 애플리케이션이 즉시 메시지 내의 의심스러운 키워드와 URL을 분석하여 위험성을 평가하였다. 예를 들어, URL이 축약된 형태이거나 정상적인 URL처럼 보이는 경우에도, 애플리케이션은 VirusTotal API와 Punycode 변환을 통해 URL의 악성 여부를 검사하였다.

실험 결과, 이 애플리케이션은 200건의 스미싱 메시지 중 167건을 성공적으로 탐지하였으며, 이는 약 83.5%의 탐지율을 기록하였다. 제안된 알고리즘이 실시간 탐지 상황에서도 효과적으로 작동함을 입증하였다.

이와 같은 결과는 스미싱 메시지를 수신한 즉시 분석하고 탐지하는 기능이 강화되었음을 보여주며, 실제 환경에서 사용자 보안을 강화할 수 있는 효과적인 방법임을 시사한다.아래의 Fig. 9는 애플리케이션 동작 화면이다.

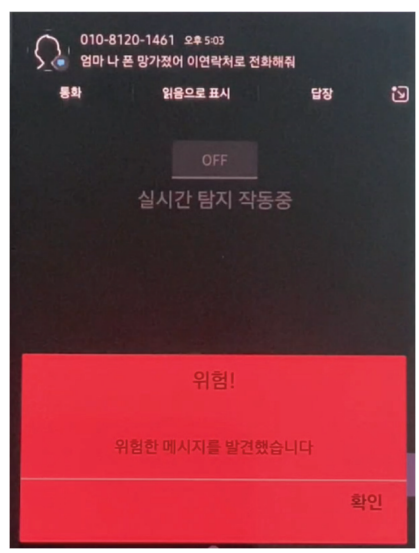

Fig. 9. Android application that detects smishing SMS.

4. 결 론

본 연구에서는 스미싱 메시지의 탐지 정확도를 향상시키기 위해 세 가지 주요 방법을 제안하고, 이를 Android 기반의 애플리케이션에 적용하여 실험을 통해 성능을 평가하였다. 제안된 방법들은 패턴 탐지 모듈의 최적화, 스

팸/피싱 메세지 데이터셋의 재검토, 그리고 의심 키워드의 토큰화 및 정규화 과정의 개선을 포함하며, 이들 모두가 스미싱 탐지의 정확성을 높이는 데 기여하였다.

실험 결과, 제안된 방법들을 종합적으로 적용한 탐지 시스템은 기존의 탐지율(약 70%)에 비해 약 167건, 83.5%로 탐지 정확도가 크게 향상되었음을 확인할 수 있었다. 특히, 다양한 패턴과 최신 스미싱 기법을 반영한 데이터셋, 그리고 변형된 단어에 대한 정규화 과정을 통해 실질적인 탐지 성능이 개선되었다. 이로써, 제안된 방법들이 실제 환경에서도 스미싱 메시지를 보다 효과적으로 탐지할 수 있음을 입증하였다.

본 연구의 결과는 스미싱 메시지 탐지 시스템의 정확성 향상을 위한 효과적인 접근 방식을 제시하며, 모바일 환경에서의 스미싱 공격에 대한 대응력을 크게 강화할 수 있음을 보여준다. 향후 연구에서는 더욱 다양한 유형의 스미싱 메시지를 포함한 확장된 데이터셋과, 실시간 탐지 성능을 최적화하기 위한 추가적인 연구가 필요할 것이다.

참고문헌

1. Choi, M., "Current Status, Types, Trends, and Implications of Voice Phishing," Korean Social Trends 2022, pp. 307-315, 2022.
2. Choi, J.Y., Oh, S.J., Roh, G.W., and Jeong, W.H., "Utilizing Pre-trained Language Models for Effective Smishing Detection," Proceedings of the Korean Institute of Information Scientists and Engineers (KIISE) Conference, Busan, Dec. 2023.
3. Jeong, S.H., Do, H.J., Cho, J.E., Park, Y.E., Kim, J.W., and Choi, J.Y., "Development of a Smishing Detection Mobile Application Using Text Mining," Proceedings of the Korean Institute of Communications and Information Sciences (KICS) Conference, Gangwon, Feb. 2022.
4. Ulfath, R.E., Sarker, I.H., Chowdhury, M.J.M., and Hammoudeh, M., "Detecting Smishing Attacks Using Feature Extraction and Classification Techniques," Proceedings of the International Conference on Big Data, IoT, and Machine Learning, pp. 677–689, Dec. 2021.
5. Verma, S., Ayala-Rivera, V., and Portillo-Dominguez, A.O., "Detection of Phishing in Mobile Instant Messaging Using Natural Language Processing and Machine Learning," Proceedings of the 2023 11th International Conference in Software Engineering Research and Innovation (CONISOFT), León, Guanajuato, Mexico, Nov. 2023. DOI: 10.1109/CONISOFT58849.2023.00029.
6. https://www.virustotal.com/gui/home/upload
7. https://github.com/haccer/subjack
8. https://book.hacktricks.xyz/v/kr/pentesting-web/domain-subdomain-takeover

접수일: 2024년 8월 29일, 심사일: 2024년 9월 10일, 게재확정일: 2024년 9월 14일