사용자 편의성 및 탐지 정확도 증대를 위한 웹 서비스 기반 피싱 사이트 탐지 방법

민준홍, 김서진, *원종현, 서준호, 남수만
청주대학교 디지털보안학과, *인공지능소프트웨어학과
{minjh04217, kimseojin0307, chwon318, junho2515, smnam}@cju.ac.kr

A Web Service-Based Phishing Site Detection Method for Enhanced User Convenience and Detection Accuracy

Jun-hong Min, Seo-jin Kim, Chong-hyun Won, Jun-ho Seo, Su-man Nam
Cheongju Univ. Department of Digital security and *Department of AI software

요 약

인터넷 인프라 확산으로 피싱 공격이 급증하는 상황에서, 기존의 블랙리스트 및 시그니처 기반 탐지 방식은 난독화되거나 신규 도메인을 사용하는 지능형 피싱을 막는 데 한계가 있다. 본 연구는 이러한 문제를 해결하고 탐지 정확도를 높이고자 웹 서비스 기반 피싱 사이트 탐지 기법을 제안한다. 이 기법은 CDP(Chrome DevTools Protocol)를 활용한 실시간 네트워크 흐름 분석과 LLM(Large Language Model)을 이용한 난독화 해독 및 코드 내 위험 함수 탐지 기능을 결합한 것이 특징이다. 이를 통해 신규 및 고난도 난독화 공격까지 효과적으로 탐지하며, Phishtank 데이터셋을 대상으로 한 검증에서, 실험환경 기준 탐지율 100%를 달성하여 우수한 성능을 입증했다.

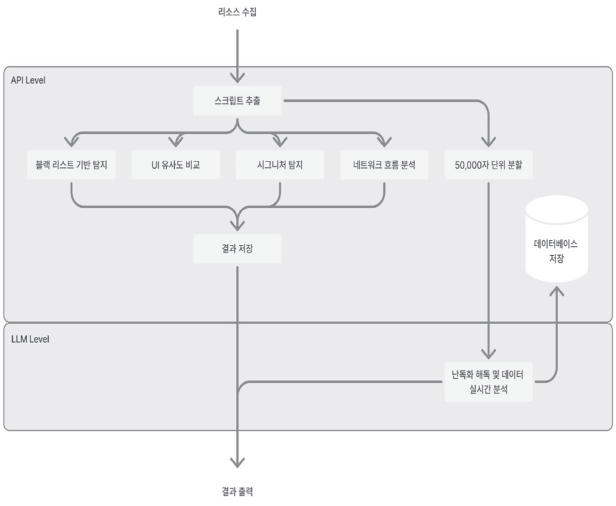
Ⅰ. 서론

안랩의 2025년도 2분기 피싱 문자 트렌드 보고서에 따르면 URL 삽입이 67.37%로 가장 많이 사용되며 1위 사칭 산업군이 금융권으로 확인된다[1]. 이러한 URL 삽입형 피싱 문자는 사용자를 피싱 사이트로 유도하여 실질적 피해로 이어질 수 있다. 피싱 사이트를 탐지하는 기존의 기법으로는 알려진 URL·도메인 목록과 요청 주소를 정규화해 대조하는 방식인 블랙리스트 기반 탐지와 피싱에서 자주 쓰이는 고유 코드 패턴을 미리 규칙으로 만들어 대조하는 방식인 시그니처 탐지가 존재하나, 이미 알려진 피싱 사이트나 도메인만 탐지할 수 있고 신규 도메인이나 난독화된 형태의 JavaScript는 탐지가 어려워 공격자가 쉽게 탐지를 회피할 수 있다[2]. 본 논문은 LLM과 CDP를 사용하여 이미 알려진 피싱 사이트 뿐만 아니라 신규 피싱 사이트도 탐지하는 방법을 제안한다. 제안 방법에서는 LLM을 사용하여 난독화된 형태의 JavaScript를 해독하고 원본 형태의 JavaScript를 알아내 코드 내 위험 함수를 탐지한다. 또한, CDP를 사용해 내부의 네트워크 도메인 기능을 활성화하고 네트워크 요청을 수집 및 분석한다. 이를 통해 변화를 모니터링하여 비정상적인 통신 여부를 탐지한다. 그리하여, 본 제안 방법은 기존 방식의 한계점을 보완하여 알려진 피싱 사이트와 신규 피싱 사이트의 인식을 향상하는 데 기여한다.

Ⅱ. 본론

2.1 개요

본 논문에서는 기존 방식과 달리 네트워크 흐름 분석과 LLM 기반 분석을 동시에 분석하여 효과적으로 피싱 사이트를 인식한다. 제안 방법은 LLM을 사용하여 난독화된 형태의 JavaScript를 해독하며, 위험 함수를 탐지한다. 또한 CDP를 사용하여 네트워크 흐름을 탐지한다. 이를 위해 실제 웹사이트에 접속하여 동적 콘텐츠를 분석하기 위해 셀레니움(Selenium)을 활용하였고 브라우저 환경에서 JavaScript, AJAX와 같은 동적 요소를 불러온다. 수집된 요소들을 통합하여 파일을 생성하고, 이를 기반으로 블랙리스트 기반 탐지, UI 유사도 비교, 시그니처 탐지, 네트워크 흐름 분석에 더불어 LLM을 활용한 난독화 해독 및 코드 분석을 수행하여 피싱 사이트 여부를 판단한다. [그림 1]은 본 논문에서 제시하는 탐지 기법의 절차이다.

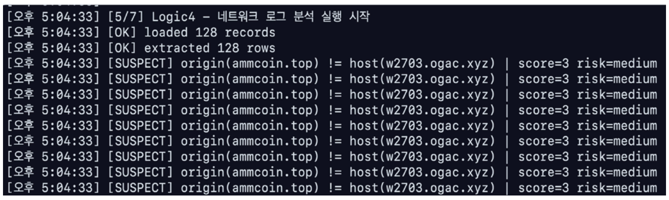


[그림 1] 제안 방법의 절차

2.2 제안 기법의 탐지 방법

네트워크 흐름 분석 방식은 CDP를 사용하여 브라우저 내부의 네트워크 도메인 기능을 활성화한다. 이를 통해 브라우저에서 발생되는 모든 네트워크 요청을 수집 및 분석하고, 사용자가 입력한 민감 정보가 외부로 전송되는지 추적한다. 또한 요청의 타임스탬프, HTTP 메서드, 출발지(origin), 목적지(host), 응답 상태 코드 등의 변화를 [그림 2]와 같이 모니터링하여

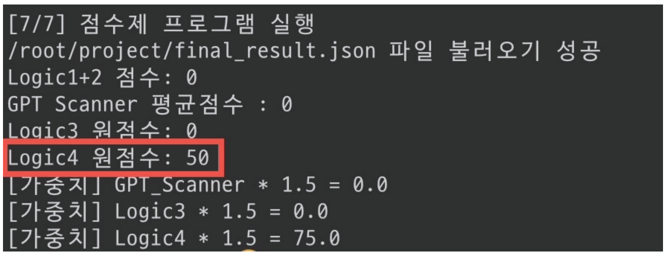
비정상적인 통신 여부를 탐지한다.



```
[오후 5:04:33] [5/7] Logic4 - 네트워크 로그 분석 실행 시작
[오후 5:04:33] [OK] loaded 128 records
[오후 5:04:33] [OK] extracted 128 rows
[오후 5:04:33] [SUSPECT] origin(ammcoin.top) != host(w2703.ogac.xyz) | score=3 risk=medium
[오후 5:04:33] [SUSPECT] origin(ammcoin.top) != host(w2703.ogac.xyz) | score=3 risk=medium
[오후 5:04:33] [SUSPECT] origin(ammcoin.top) != host(w2703.ogac.xyz) | score=3 risk=medium
[오후 5:04:33] [SUSPECT] origin(ammcoin.top) != host(w2703.ogac.xyz) | score=3 risk=medium
[오후 5:04:33] [SUSPECT] origin(ammcoin.top) != host(w2703.ogac.xyz) | score=3 risk=medium
[오후 5:04:33] [SUSPECT] origin(ammcoin.top) != host(w2703.ogac.xyz) | score=3 risk=medium
[오후 5:04:33] [SUSPECT] origin(ammcoin.top) != host(w2703.ogac.xyz) | score=3 risk=medium
[오후 5:04:33] [SUSPECT] origin(ammcoin.top) != host(w2703.ogac.xyz) | score=3 risk=medium
[오후 5:04:33] [SUSPECT] origin(ammcoin.top) != host(w2703.ogac.xyz) | score=3 risk=medium
```

[그림 2] 네트워크 흐름 분석 과정

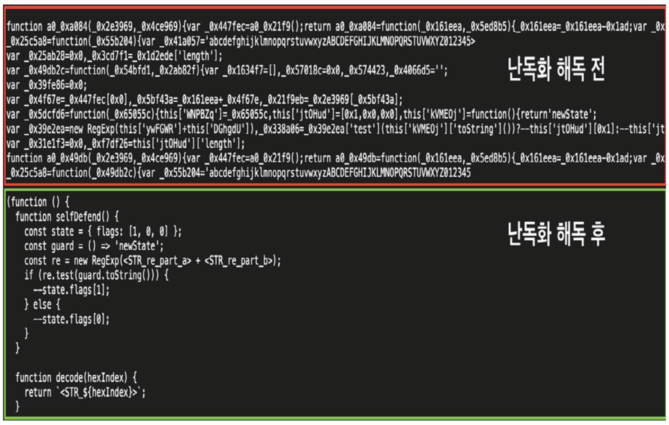
이를 통해 [그림 3]은 기존의 방식으로 탐지하지 못한 피싱 사이트를 탐지한다.



```
[7/7] 점수제 프로그램 실행
/root/project/final_result.json 파일 불러오기 성공
Logic1+2 점수: 0
GPT Scanner 평균점수 : 0
Logic3 원점수: 0
Logic4 원점수: 50
[가중치] GPT_Scanner * 1.5 = 0.0
[가중치] Logic3 * 1.5 = 0.0
[가중치] Logic4 * 1.5 = 75.0
```

[그림 3] 네트워크 흐름 이상 탐지

본 논문에서 LLM은 난독화 해독과 코드 내 위험 함수 탐지에 활용된다. 사용된 LLM은 [그림 4]와 같이 받아온 데이터의 난독화를 복호화할 때 입력받은 JS/HTML에서 DOM, script 경계, 주석, 공백, 평문을 제외하고 이해하기 어려운 형태를 분석하여 해독한다. 이를 바탕으로 [그림 5]와 같이 리다이렉트 유도, 지갑 연결 위장, 브랜드 사칭과 같은 위험 가능성이 있는 특정 함수를 중점으로 탐지하여 해당 사이트가 피싱 사이트인지 아닌지를 인식한다.

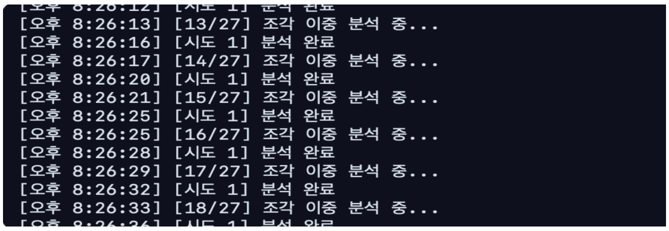


```
function a0_0xa084(_0x2e3969,_0x4ce969){var _0x447fec=a0_0x21f9();return a0_0xa084=function(_0x161eea,_0x5ed8b5){_0x161eea=_0x161eea-0x1ad;var _0x
_0x25c5a8=function(_0x55b204){var _0x41a057='abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ012345>
var _0x25ab28=0x0,_0x3cd7f1=_0x1d2ede['length'];
var _0x49db2c=function(_0x54bfd1,_0x2ab82f){var _0x1634f7=[],_0x57018c=0x0,_0x574423,_0x4066d5='';
var _0x39fe86=0x0;
var _0x4f67e=_0x447fec[0x0],_0x5bf43a=_0x161eea+_0x4f67e,_0x21f9eb=_0x2e3969[_0x5bf43a];
var _0x5dcfd6=function(_0x65055c){this['WNPBZq']=_0x65055c,this['jtOHud']=[0x1,0x0,0x0],this['kVMEOj']=function(){return'newState';
var _0x39e2ea=new RegExp(this['ywFGWR']+this['DGhgdU']),_0x338a06=_0x39e2ea['test'](this['kVMEOj']['toString']())?--this['jtOHud'][0x1]:--this['jt
var _0x31e1f3=0x0,_0xf7df26=this['jtOHud']['length'];
function a0_0x49db(_0x2e3969,_0x4ce969){var _0x447fec=a0_0x21f9();return a0_0x49db=function(_0x161eea,_0x5ed8b5){_0x161eea=_0x161eea-0x1ad;var _0x
_0x25c5a8=function(_0x49db2c){var _0x55b204='abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ012345
```

```
(function () {
  function selfDefend() {
    const state = { flags: [1, 0, 0] };
    const guard = () => 'newState';
    const re = new RegExp(<STR_re_part_a> + <STR_re_part_b>);
    if (re.test(guard.toString())) {
      --state.flags[1];
    } else {
      --state.flags[0];
    }
  }

  function decode(hexIndex) {
    return `<STR_${hexIndex}>`;
  }
```

[그림 4] 난독화 해독 전/후 비교

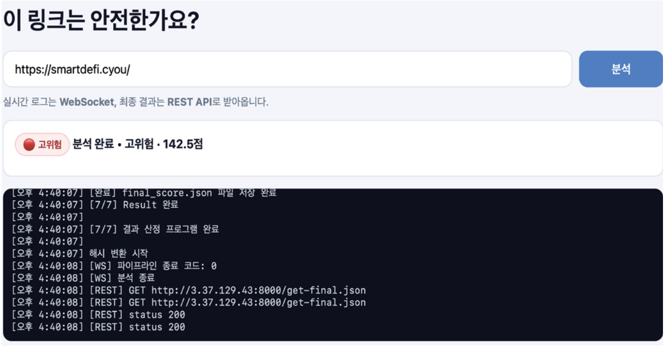


```
[오후 8:26:13] [13/27] 조각 이중 분석 중...
[오후 8:26:16] [시도 1] 분석 완료
[오후 8:26:17] [14/27] 조각 이중 분석 중...
[오후 8:26:20] [시도 1] 분석 완료
[오후 8:26:21] [15/27] 조각 이중 분석 중...
[오후 8:26:25] [시도 1] 분석 완료
[오후 8:26:25] [16/27] 조각 이중 분석 중...
[오후 8:26:28] [시도 1] 분석 완료
[오후 8:26:29] [17/27] 조각 이중 분석 중...
[오후 8:26:32] [시도 1] 분석 완료
[오후 8:26:33] [18/27] 조각 이중 분석 중...
```

[그림 5] LLM 기반 분석 과정

2.3 제안 기법의 구현

이후 각 탐지 결과를 종합한 최종 위험 점수를 [그림 6]과 같이 산정하고 등급화한다. 또한 [그림 7]처럼 탐지된 URL은 데이터베이스에 저장하며 이후 동일 URL 입력 시 즉시 결과 제공이 가능하다. 최종 탐지 점수에 따라 색상을 적용하여 시각적으로 즉각 판단할 수 있게 하였으며, 사용자가 스스로 판단할 수 있게 추가 정보를 제공한다.



이 링크는 안전한가요?

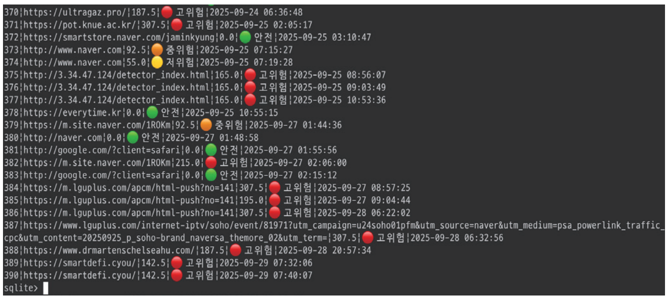
https://smartdefi.cyou/

실시간 로그는 WebSocket, 최종 결과는 REST API로 받아옵니다.

🔴 고위험 분석 완료 • 고위험 · 142.5점

```
[오후 4:40:07] [완료] final_score.json 파일 저장 완료
[오후 4:40:07] [7/7] Result 완료
[오후 4:40:07]
[오후 4:40:07] [7/7] 결과 산정 프로그램 완료
[오후 4:40:07]
[오후 4:40:07] 해시 변환 시작
[오후 4:40:08] [WS] 파이프라인 종료 코드: 0
[오후 4:40:08] [WS] 분석 종료
[오후 4:40:08] [REST] GET http://3.37.129.43:8000/get-final.json
[오후 4:40:08] [REST] GET http://3.37.129.43:8000/get-final.json
[오후 4:40:08] [REST] status 200
[오후 4:40:08] [REST] status 200
```

[그림 6] 제안 시스템의 결과 화면



```
370¦https://ultragaz.pro/¦187.5¦🔴 고위험¦2025-09-24 06:36:48
371¦https://pot.knue.ac.kr/¦307.5¦🔴 고위험¦2025-09-25 02:05:17
372¦https://smartstore.naver.com/jaminkyung¦0.0¦🟢 안전¦2025-09-25 03:10:47
373¦http://www.naver.com¦92.5¦🟠 중위험¦2025-09-25 07:15:27
374¦http://www.naver.com¦55.0¦🟡 저위험¦2025-09-25 07:19:28
375¦http://3.34.47.124/detector_index.html¦165.0¦🔴 고위험¦2025-09-25 08:56:07
376¦http://3.34.47.124/detector_index.html¦165.0¦🔴 고위험¦2025-09-25 09:03:49
377¦http://3.34.47.124/detector_index.html¦165.0¦🔴 고위험¦2025-09-25 10:53:36
378¦https://everytime.kr¦0.0¦🟢 안전¦2025-09-25 10:55:15
379¦https://m.site.naver.com/1ROKm¦92.5¦🟠 중위험¦2025-09-27 01:44:36
380¦http://naver.com¦0.0¦🟢 안전¦2025-09-27 01:48:58
381¦http://google.com/?client=safari¦0.0¦🟢 안전¦2025-09-27 01:55:56
382¦https://m.site.naver.com/1ROKm¦215.0¦🔴 고위험¦2025-09-27 02:06:00
383¦http://google.com/?client=safari¦0.0¦🟢 안전¦2025-09-27 02:15:12
384¦https://m.lguplus.com/apcm/html-push?no=141¦307.5¦🔴 고위험¦2025-09-27 08:57:25
385¦https://m.lguplus.com/apcm/html-push?no=141¦195.0¦🔴 고위험¦2025-09-27 09:04:44
386¦https://m.lguplus.com/apcm/html-push?no=141¦307.5¦🔴 고위험¦2025-09-28 06:22:02
387¦https://www.lguplus.com/internet-iptv/soho/event/81971?utm_campaign=u24soho01pfm&utm_source=naver&utm_medium=psa_powerlink_traffic_
cpc&utm_content=20250925_p_soho-brand_naversa_themore_02&utm_term=¦307.5¦🔴 고위험¦2025-09-28 06:32:56
388¦https://www.drmartenschelseahu.com/¦187.5¦🔴 고위험¦2025-09-28 20:57:34
389¦https://smartdefi.cyou/¦142.5¦🔴 고위험¦2025-09-29 07:32:06
390¦https://smartdefi.cyou/¦142.5¦🔴 고위험¦2025-09-29 07:40:07
sqlite> 
```

[그림 7] 데이터베이스 저장 결과

Ⅲ. 결론

피싱 사이트는 사회적 관심사를 이용한 심리적 허점, 개인정보 탈취를 통한 주변 지인 위장 등으로 매년 꾸준히 증가하고 있다. 이를 해결하기 위해 기존의 블랙리스트 기반 탐지나 시그니처 탐지 방식이 사용되었으나, 신규 도메인이나 난독화를 통해 쉽게 회피가 가능하다는 한계가 있었다. 이에 본 논문에서는 이러한 기존 탐지 기법의 한계를 보완하기 위해 LLM 기반 탐지 방식과 CDP를 사용한 네트워크 기반 탐지를 결합하여 제안했다. 제안된 탐지 방식은 기존 기법을 우회하는 신규 피싱 사이트와 난독화된 공격까지 효과적으로 탐지하는 우수한 성능을 보였다. 더불어, 신규 피싱 사이트와 난독화까지 효과적으로 탐지하는 성능을 보였으며, 제안 기법의 실효성을 검증하기 위해 상용 솔루션인 NordVPN과 Virus Total, 본 논문의 탐지 기법을 Phishtank 데이터셋 기반의 약 89개의 피싱사이트를 대상으로 비교 실험한 결과 NordVPN은 58.4%, Virus Total은 77.5%를 기록했지만, 본 논문의 기법은 탐지율 100%를 기록하였다. 더불어, 정상 사이트 18개를 모두 정확히 판단하여 오탐률 0%를 보이며, 실제 서비스 적용을 위한 높은 신뢰성까지 확보하였다. 이로써 본 논문의 탐지 기법의 기존 방식 대비 우수한 성능을 입증하였다. 이를 통해, 제안된 기법의 정보보안 분야에서의 활용 가능성을 입증했다는 점에서 학문적·실무적 기여를 함과 동시에 사용자가 웹사이트 접속 이전에 피싱 여부를 사전 판별할 수 있도록 지원함으로써, 실제 보안 서비스 및 기업 보안 솔루션에 적용할 수 있는 잠재력을 보여준다.

참고문헌

[1] 안랩, 2025년 2분기 피싱 문자 트렌드 보고서 발표, 2025.

[2] Yeon-Gi Jung, “A Study on Blocking Criminal Websites Based on Internet Domain Registration Data of Phishing Sites.” Journal of Digital Forensics , 9(2), 79-96, 2015.

[3] PhishTank, (https://phishtank.org/)