보이지 않는 딥웹의 위협과 탐지에 관한 연구

경아, 하윤경, 배은석, 김윤태, 이원영, 강용범, 이상진*
포테이토넷, *고려대학교 정보보호대학원
everska@potatonet.ai, *sangjin@korea.ac.kr

A Study on the threat and detection of invisible DeepWeb

Kyoung A Shin, Yun Kyung Hah, Eun Suk Bae, Yoon Tae Kim, Won Young Lee, Yong Beom Kang, Sang Jin Lee*
PotatoNet, *School of Cybersecurity, Korea Univ.

요 약

본 논문은 최근 국내에서 발생하고 있는 피싱과 랜섬웨어부터 통신사 유심정보 유출과 개인정보 유출까지 나라 안팎이 시끄럽다. 코로나로 진행된 디지털 진화와 AI 공격의 지능화 때문이라 하지만, 다시 한번 우리의 보안이 적절했는지에 대해 뒤돌아보고자 한다.
· ·

· · 딥웹은 검색엔진으로 검색되지 않아 선제적 예방은 어렵지만 사용자는 수신한 링크를 클릭하여 접속이 가능하다는 점에서 더 위협적이다. 눈에 보이지않는 딥웹은 · 하였다.

Ⅰ. 서 론

본 논문에서는 최근 국내에서 발생하고 있는 피싱과 랜섬웨어부터 통신사 유심정보 유출과 개인정보 유출까지 나라 안팎이 시끄럽다. 코로나로 진행된 디지털 진화와 AI 공격의 지능화 때문이라 하지만, 다시 한번 우리의 보안이 적절했는지 살펴보고자 한다.

· ·

딥웹은 검색엔진으로 검색되지 않아 선제적 예방은 어렵지만 사용자는 수신한 링크를 클릭하여 접속이 가능하다는 점에서 더 위협적이다. 눈에 보이지않는 딥웹은 · 있다.[1]

Ⅱ. 본론

본 논문에서는 · 웹은 현대 사회에서 정보의 공유와 소통을 가능하게 하는 매우 중요한 플랫폼이다. 디지털 혁신이 가속화되면서 웹의 발전은 눈부시게 이루어졌지만, 동시에 웹을 악용한 사이버 공격 역시 점점 더 심각한 사회적 문제로 대두되고 있다. 실제로 전체 사이버 공격의 85%가 웹을 통해 발생하며, 웹은 주요 공격 경로가 되고 있다.

웹은 크게 두 가지 영역으로 나뉜다. 하나는 일반 사용자가 쉽게 접근할 수 있고 검색엔진에도 노출되는 '표면웹'이고, 다른 하나는 사용자에게 보이지 않고 검색엔진에도 검색되지 않는 '딥웹[4]'이다(그림 5-1-1). 많은 웹 위협이 딥웹에 숨어있음에도 불구하고, 표면웹만을 중심으로 악성 웹을 탐지하려 하기 때문에 실제로 많은 위협을 놓치고 있다. 딥웹은 포트 번호, 하위 URL, 파라미터 등 다양한 변수를 통해 더욱 교묘하게 숨겨져 있어, 표면웹에 노출된 정보만으로는 악성 사이트를 탐지하는 데 한계가 있다.

대다수 악성 URL 탐지 서비스는 무료 보안서비스로서 과거에 탐지된 악성 URL 목록, 즉 '블랙리스트' 기반으로 악성 여부를 점검한다. 하지만 블랙리스트 방식은 평균 5시간이면 사라지는 악성 사이트를 제대로 잡아내지 못한다. 또한, 블랙리스트 방식은 신규 악성웹이나 딥웹에 숨겨진 악성웹을 탐지할 수 없기 때문에, 근본적으로 한계를 가질 수밖에 없다.

전 세계적으로 웹 위협 탐지율이 매우 낮다는 연구 결과도 있다[1]. 예를 들어, 바이러스토탈에 등록된 탐지 엔진의 정확도가 5% 미만에 그치고 있으며, 웹 위협의 약 39%는 나타났다가 사라지기를 반복하여 탐지가 어렵다[2]. 악성웹이 나타나는 시점은 주로 공격 시점이며, 공격 대상인 사용자가 해당 웹 URL이 전달되는 순간이기 때문에, 실시간으로 사용자의 웹 접속을 점검해야만 효과적으로 탐지할 수 있다.

딥웹은 일반적으로 검색엔진으로 접근할 수 없는 웹의 숨겨진 영역을 뜻한다. 딥웹은 다시 다크웹과 다크웹이 아닌 영역(협의의 딥웹)으로 나눌 수 있다. 본 표준에서는 다크웹이 아닌, 일반 웹브라우저로 접근 가능한 숨겨진 영역인 협의의 딥웹을 '딥웹'으로 정의한다. 다시 딥웹에는 인증 절차를 거쳐야만 접근할 수 있는 정상 딥웹(이메일, 온라인 뱅킹, 회사 내부 데이터베이스 등)과, 인증 없이 숨겨져 있으면서 악의적으로 활용될 수 있는 악성 딥웹이 존재한다. 딥웹은 단순히 HTML이나 자바스크립트로 코드를 숨긴(예: display:none 등) 것과는 다르다. 코드상으로 확인이 가능한 영역은 딥웹으로 보지 않는다.

악성 딥웹은 주로 웹서버에 숨겨진 웹쉘 등으로 나타나며, 해킹된 웹서버를 통해 악성 코드를 유포하거나 추가 공격의 거점으로 활용된다. 사용자의 웹 접속을 실시간으로 점검하면 이러한 악성 딥웹을 효과적으로 탐지할 수 있고, 도메인 정보를 분석하면 선제적으로 악성 딥웹 사이트를 찾아낼 수도 있다.
한편, 다크웹은 Tor, I2P 등 특수 브라우저와 .onion 등 특수 도메인을 사용해 여러 겹의 암호화와 익명성을 보장하는 영역이다. 다크웹은 언론인 보호나 정치적 탄압 회피 등 합법적 목적으로도 이용되지만, 불법 거래나 범죄 활동에 악용되는 경우가 많다. 다크웹에서는 악성 딥웹 URL이 범죄 거래나 특정 커뮤니티 공격에 이용되기도 한다. 다만, 다크웹 자체는 일반적인 공격 경로로 많이 사용되지는 않는다.
이처럼 딥웹과 다크웹은 기존의 보안 탐지 체계로는 충분히 대응하기 어려운 새로운 위협을 내포하고 있다. 이에 따라, 본 표준에서는 일반 브라우저로 접근 가능한 웹의 숨겨진 영역, 즉 딥웹에서 사이버 공격에 사용되는 악성 URL을 효과적으로 수집하고 탐지하기 위한 요구사항을 정의하고자 한다. 이는 기존 블랙리스트 기반의 한계를 극복하고, 실시간·선제적으로 웹 위협에 대응하기 위한 새로운 접근법이라 할 수 있다.

웹은 정보의 공유와 소통을 가능하게 하는 중요한 플랫폼이다. 디지털 혁신은 웹의 발전을 가속화했지만, 이와 동시에 웹 서비스를 이용한 사이버 공격이 점점 더 심각한 사회 문제로 대두되고 있다.
사이버 공격의 85%는 웹을 통해 발생한다. 웹은 사용자에게 보이고 검색엔진에 의해 검색되는 표면웹과 사용자에게 보이지 않고 검색엔진에 검색되지 않는 딥웹[4]으로 구성된다(그림 5-1-1).
웹 위협의 상당수는 딥웹에 숨어있으나, 다수의 보안업체들은 표면웹을 탐색하며 악성웹을 찾으려 하기 때문에 탐지에 실패하고, 이로 인해 관련 산업은 한계에 봉착하고 있다. 탐지 실패의 원인은 포트번호, 하위 URL, 파라미터까지 다양한 변수를 이용하여 딥웹에 숨어있기 때문에, 표면웹에 노출된 웹사이트 정보만으로 예측하기 어렵다.
표면웹에 보이지 않는 악성웹을 찾기 위한 방법은 공격 목표인 사용자의 웹 접속을 점검하는 것이다. 그러나 무료 보안서비스에서는 평균 5시간이면 사라지는 블랙리스트 방식을 적용하고 있다. 이 블랙리스트 방식은 웹의 악성 여부를 점검하는 것이 아니라, 과거에 탐지된 악성 URL과 일치하는가를 확인하는 방식이다. 따라서 블랙리스트 방식으로는 악성웹이나 신규 악성웹을 탐지할 수 없을 뿐만 아니라, 딥웹에 숨은 악성웹을 찾을 수 있는 기회까지 놓치게 된다.
전세계적으로 웹 위협의 탐지율이 저조하다는 연구가 발표되었다[1]. 바이러스토탈에 등록된 탐지 엔진 72.5%의 정확도는 5% 미만이라고 한다. 또한 웹 위협은 39%가 나타났다 사라지기를 반복한다[2]. 악성웹이 나타나는 시점은 공격 시점이며 공격 목표인 사용자에게 악성웹이 전달되는 시점이다. 악성웹이 나타나는 시점에 실시간으로 웹접속을 점검해야 악성웹을 찾을 수 있다.
딥웹의 은닉 특성으로 인하여 표면웹에서 악성웹을 찾을 수 없고, 유료 고객의 실시간 위협 채널에 한정적인 악성웹 수집만 가능하자, 이를 극복하고자 민간과 공공에서는 위협을 공유하고 활용하는 협력체계를 구성하였다. 그럼에도 악성 딥웹에 대한 선제적 대응 부재와 블랙리스트 기반의 웹 점검 방식은 APT, 피싱 공격과 같은 사이버 공격에 효과적으로 대처하지 못하는 한계를 드러내고 있다.
광의의 딥웹은 일반 검색엔진으로 접근할 수 없는 웹의 숨겨진 영역을 의미하며, 다크웹과 다크웹이 아닌 협의의 딥웹으로 나눌 수 있다. 본 표준에서는 다크웹이 아닌 협의의 딥웹을 딥웹이라 지칭하고자 한다.
딥웹은 일반적인 웹브라우저를 통해 접근 가능한 표면웹과 동일한 웹의 숨겨진 영역이다. 딥웹에서 정상 딥웹과 악성 딥웹 간에는 명확한 차이가 존재한다. 정상 딥웹에는 접근 권한을 검증하는 인증 절차를 통해 보호되는 개인 이메일 계정, 온라인 뱅킹 페이지, 회사 내부 데이터베이스, 의료기록 시스템, 학술 논문 데이터베이스 등이 포함된다. 정상 딥웹은 직접적으로 내용 확인은 어려우나 합법적이고 정당한 목적으로 일반 대중의 공개적인 접근을 제한하고 있으며, 개인정보 보호나 보안상의 이유로 검색엔진에 노출되지 않는다.
인증 절차가 없는 “숨겨진 영역에 위치하는 딥웹”은 표면웹을 통해 접근이 불가능 하다.

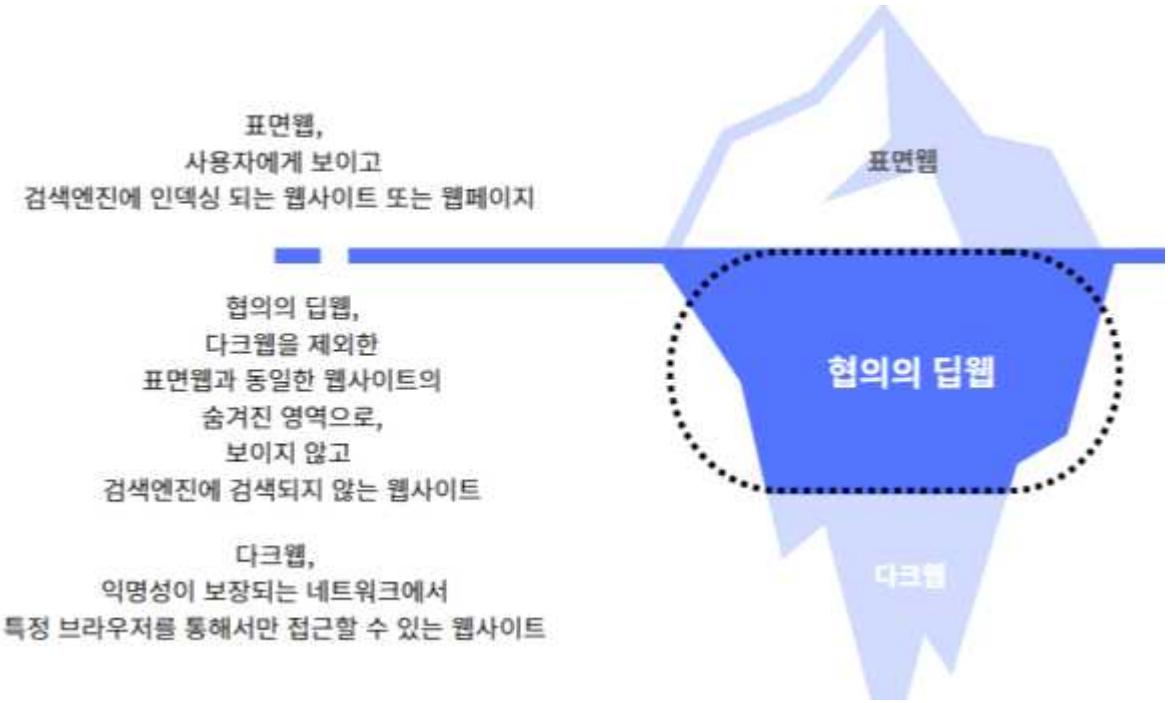
접근이 불가능한 것은 웹페이지와 연결된 하이퍼링크가 없기 때문이다. 이렇게 링크가 단절된 특성이 악성 딥웹의 대표적인 특성이다.
정상적인 웹은 링크를 통해 상호 유기적으로 연결된 구조를 형성하나, 링크가 단절되어 연결이 끊긴 웹은 비정상적인 딥웹으로서, 이러한 비정상적인 딥웹에서 악성 웹을 찾아내는 원리가 디지털 체인 이론이다.

그러나 악성 딥웹의 은닉성이 사이즈가 0인 이미지 또는 display : none의 iframe 컨텐츠와는 구별된다. 즉 HTML과 자바스크립트를 통해 식별가능한 컨텐츠는 브라우저 상에 보이지 않는다 하여도 코드를 통해 확인이 가능하기 때문에 딥웹이라고 볼 수 없다.
??악성웹의 시작은 웹서버에서 출발한다. 웹서버의 악성 딥웹을 탐지하고 특히 딥웹에 숨은 웹쉘 탐지에 효과적이다.

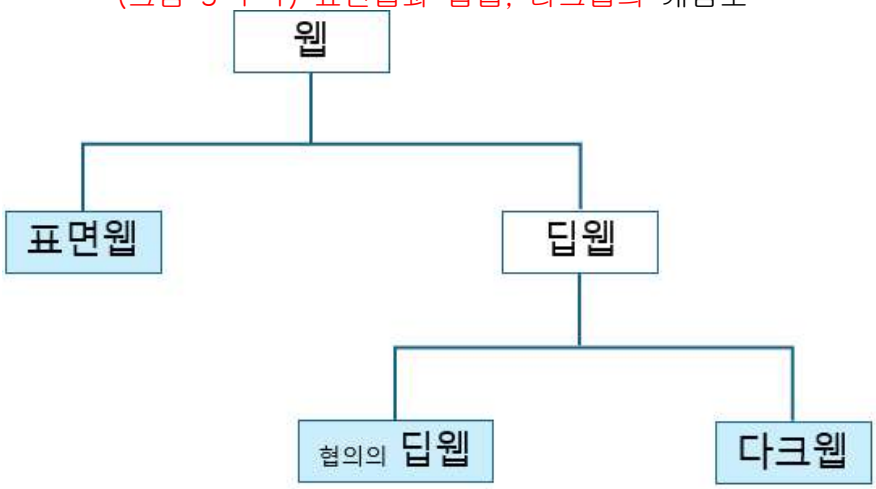
악성 딥웹 탐지 기술은 사용자의 웹 점검을 통해 악성웹을 효과적으로 탐지할 수 있을 뿐만 아니라, 해킹된 웹서버의 악성 딥웹을 탐지하고, 도메인 정보를 이용하여 악성 딥웹 사이트를 찾아낼 수 있다. 특히 악성 딥웹 사이트 탐지는 사용자의 웹 요청과 무관하게 선제적인 악성웹 대응이 가능하다.

다크웹은 Tor, I2P, 덕덕고와 같은 특수 브라우저를 이용하여 여러 겹의 레이어로 암호화된 네트워크를 거치면서 추적이 어려워 익명성을 보장한다. 다크웹의 도메인은 .onion 같은 특별한 도메인을 사용한다. 합법적인 용도(언론인 보호, 정치적 탄압 회피 등)도 있지만, 불법적인 거래나 활동이 이루어지는 경우가 많다.
본 표준에서 다루는
?? 사이버 공격 악성 URL이 다크웹에서도 발견된다.
악성 딥웹은 직접적으로 해당 URL이 공격용이지만 다크웹 URL은 직접적으로 공격에 사용되지 않는다. 다크웹 접속을 위한 특수 브라우저를 이용해야 한다는 점과 .onion이라는 도메인 이름의 특이성으로 인하여 쉽게 노출되기 때문에 사이버 공격에서 거의 사용되지 않는다.

다크웹에서는 악성 딥웹의 URL이 범죄 거래 또는 다크웹의 특정 커뮤니티를 공격하기 위해 사용되기도 한다. 다크웹의 특정 포럼이나 채팅방 등에 게시하여 악성코드에 감염시키나 개인정보가 유출되는 등의 피해를 일으킨다. 이러한 방법은 대상 커뮤니티의 신뢰도를 떨어뜨리고, 참여 구성원을 혼란시키거나, 경쟁 포럼을 약화시키려는 의도이다.
다크웹에서 불법 거래를 위해 사용되는 웹쉘 URL은 사용자를 공격하는 악



(그림 5-1-1) 표면웹과 딥웹, 다크웹의 개념도



(그림 5-1-2) 표면웹과 딥웹, 다크웹의 계층도

성 URL은 아니지만, 웹쉘 공격 이후 사용자 및 시스템 공격으로 이어질 수 있어 웹쉘도 포함하고자 한다.

본 표준에서는 사이버공격에 사용되는 악성웹(URL)을 딥웹에서 수집하고 탐지하기 위한 요구사항을 정의한다. 즉 일반 브라우저로 접근가능한 표면웹과 동일한 웹의 보이지 않는 숨겨진 영역에 존재하는 악성 URL을 대상으로 한다.

5.2 딥웹 유형

딥웹은 정상딥웹과 악성딥웹으로 나뉜다.

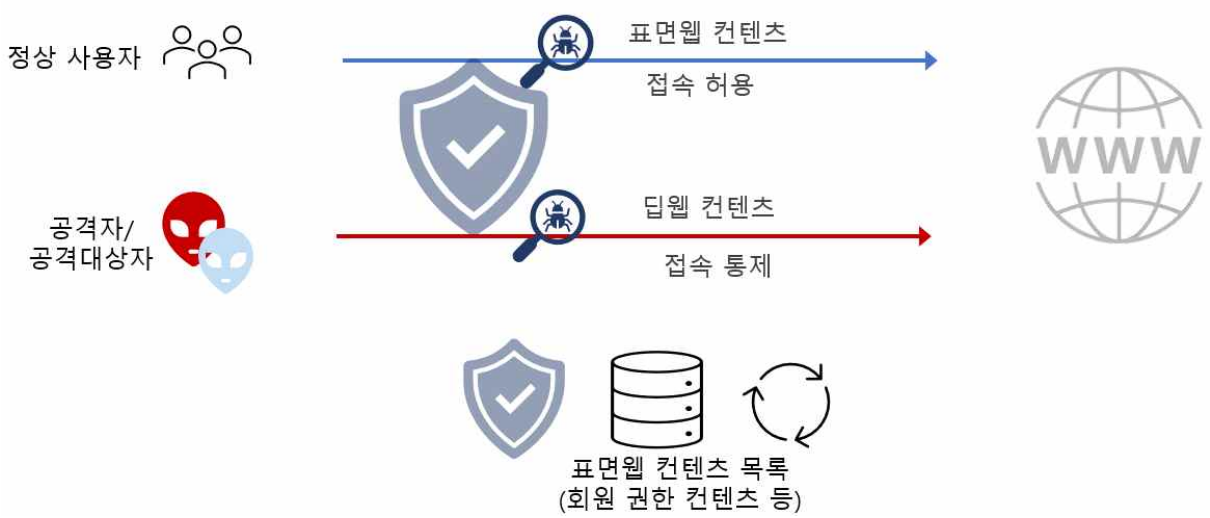
정상 딥웹은 인증 절차를 거쳐 비공개 커뮤니티나 유료 결제 등 특정 조건을 충족한 사용자만 접근이 가능한 웹이다. 비개방성으로 인하여 불법 정보를 공유할 수 있으나 이런 경우 구성원 간에 합의 하에 진행되는 것이다. 정상 딥웹은 구성원 간 합의 하에 운영되며 내부 링크 구조를 갖는 것이 특징이다.

악성 딥웹은 불특정 다수를 대상으로 사용자 몰래 공격을 수행하며 인증 없이 일반브라우저로 접근 가능한 점에서 정상딥웹과 구별된다. 은닉 특성이 없는 단순 피싱 공격은 제외하지만, 피싱 사이트로 유도하는 딥웹 기반 공격은 악성 딥웹에 포함된다.

딥웹은 사용자 또는 일반적인 검색엔진에 공개되지 않는 웹사이트이며[4], 딥웹의 규모는 표면웹의 550배로 추정된다[5]. 공격자는 검색엔진의 크롤러에 의해 수집이 안되는 딥웹의 은닉성을 이용하여 개인정보를 유출하거나 악성코드를 유포하기 위해 악용된다.

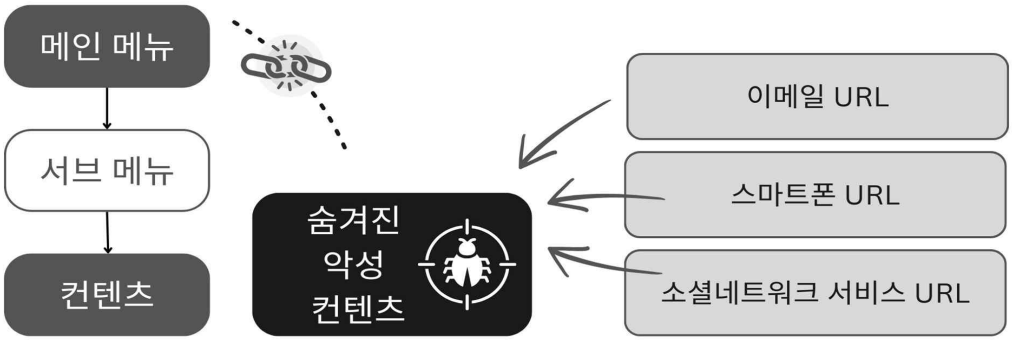
Broken link는 단절된 링크라는 공통점을 가진다. 그러나 Broken link는 (외부) 링크는 있으나 컨텐츠가 없는 경우를 뜻한다. 링크를 따라 컨텐츠에 접근할 때 해당 컨텐츠가 사라진 경우이다. 악성 딥웹은 컨텐츠는 있는데 링크가 없는 경우이다. 악성 딥웹은 사이버 공격을 위해 악성 컨텐츠는 존재하나, 동일 웹사이트에 있는 다른 컨텐츠와 내부 링크가 없는 상태이다. 내부 링크가 단절되어 내부 컨텐츠로부터의 접근이 불가능하다. 즉 웹의 컨텐츠와 컨텐츠를 연결하는 내부 하이퍼링크가 없어 정상적인 웹사이트의 홈페이지로부터 접근이 불가능하다. 그러나 악성 컨텐츠에 대한 웹주소를 알면 바로 접근 가능하고 공격이 실행될 수 있다는 점에서 더 위협적이다.

번 대 상태코드가 나오기도 한다. 악성은 확정할 수 없으나 딥웹 대상으로 포



(그림 5-2-2) 웹 요청에 따른 은닉 컨텐츠 웹 접근 통제 방안

함하기로 한다.



(그림 5-2-1) 링크 단절에 의해 표면웹에서 보이지 않는 악성 딥웹,
이메일이나 문자, SNS를 통해 공격 URL로 악용되는 억성 딥웹

<표 5-2> 상태코드에 따른 링크단절 및 링크 확인불가의 딥웹 식별 요구사항

식별 코드	구분	HTTP 상태코드[1]	컨텐츠	링크
Req-I1	웹 페이지와 연결된 링크가 없는 딥웹	200	내용 있음	링크없음
Req-I2	웹의 컨텐츠가 없는 딥웹	200	내용 없음 -빈페이지 -에러메시지만	링크없음
Req-I3	리다이렉트 딥웹	300 번대	내용 확인 불가/ 사이트 접근불가	링크없음
Req-I4	오류로 감춰진 딥웹	400~500	내용 확인 불가	확인불가

딥웹은 링크 단절 외에도 상태코드에 따라 링크 정보를 확인할 수 없는 상태 역시 딥웹으로 추정할 수 있다. 300~500의 상태코드에서는 컨텐츠를 은닉함으로써 링크 정보를 확인할 수 없게 된다. 탐지 우회를 목적으로 컨텐츠를 은닉하는 경우라고 볼 수 있다. 권한이 없거나 매개변수가 불충분하여 400

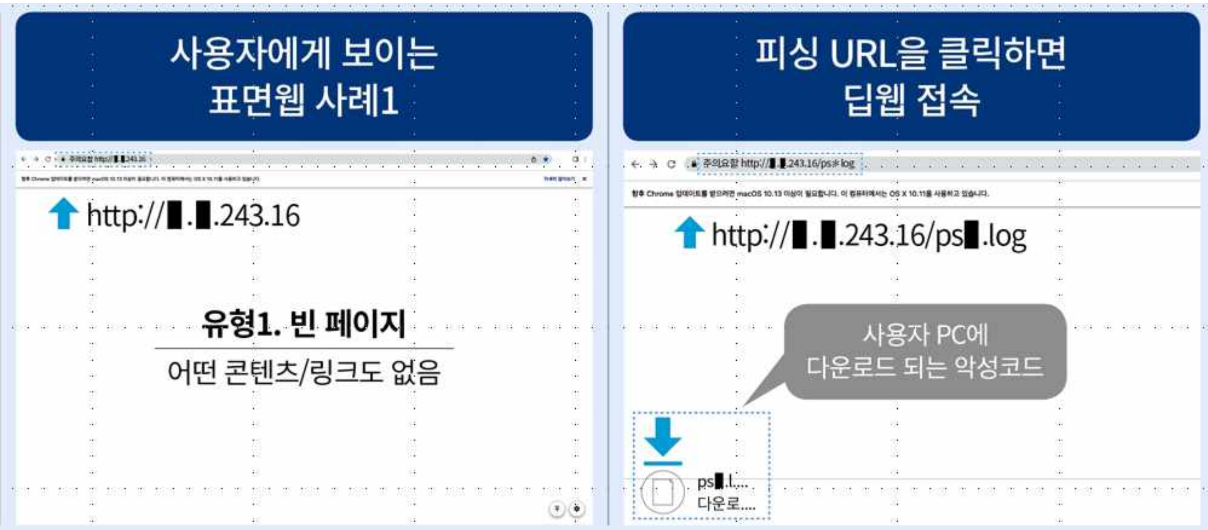
5.2.2 웹 컨텐츠가 없는 딥웹

웹사이트 내부에 허용된 사용자 또는 권한이 있는 사용자가 접속할 수 있는 하위 페이지 또는 링크 외 웹 페이지를 뜻한다. 이는 웹사이트 개발 도중 생성된 테스트 페이지 또는 공격자가 해당 웹사이트를 해킹하고 내부에 공격자가 만든 웹 페이지를 숨겨둔 경우이다. 특정 링크로의 접속을 통해 딥웹 컨텐츠로 접속할 수 있다.

웹페이지에 링크가 없고 컨텐츠가 존재하지만 스크린샷에는 출력 내용이 없거나, 코드는 자바스크립트로만 구성되어 있다. 이렇게 웹 컨텐츠가 없는 경우, 메인페이지를 제외한 그 어떤 경로로도 이동이 불가능하다.

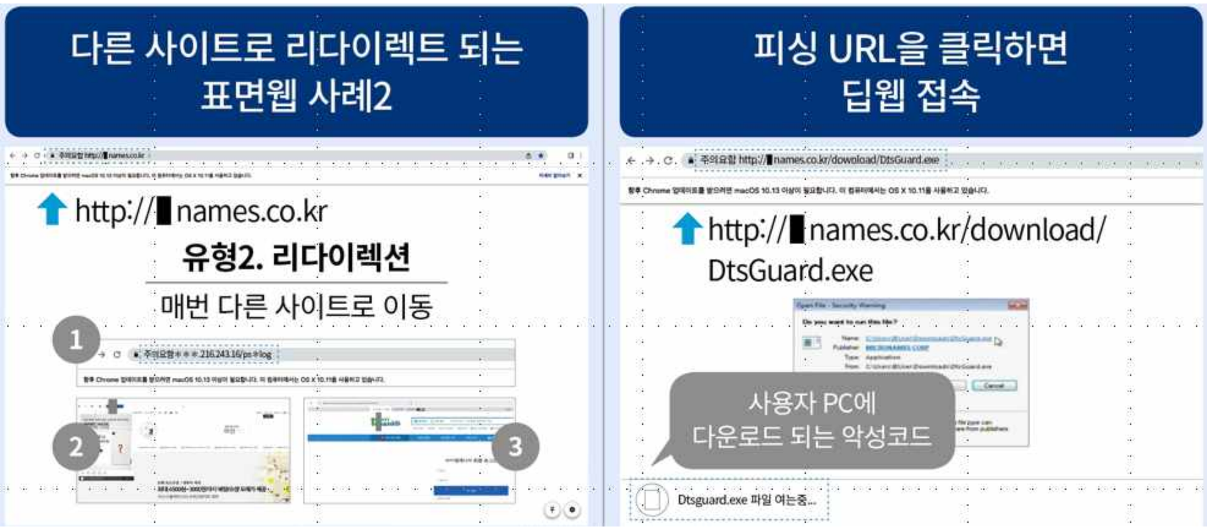
5.2.3 리다이렉트 딥웹

사용자가 웹사이트에 접근할 시 다른 웹사이트로 자동 리다이렉트가발생하는 경우이다. 그런데 홈페이지에 접근하면 리다이렉트 되나, 서브페이지로



(그림 5-3-2) 메인페이지의 컨텐츠가 은닉된 딥웹의 예시

접근하면 악성코드가 다운로드 되기도 한다.서브페이지 URL을 아는 경우 · · · · · 있다.
직접 접근이 가능하다.



(그림 5-3-3) 리다이렉트 딥웹의 예시

5.2.4 오류로 감춰진 딥웹

웹사이트에 접속 시 상태 코드 200은 정상 접속 상태를 나타내지만, 그 외 300~500의 상태 코드는 정상이 아닌 상태를 나타낸다. 리다이렉트(301, 302)를 포함하여 접속 권한이 없거나(401, 403), 웹 페이지가 없거나(404), 서버 오류(500) 등이다.
이러한 웹서버 오류는 의도적으로 탐지를 회피하기 위하여 사용된다. 특정 시간 대에 악성 활동을 하는 경우 외에는 오류 상태를 반환한다. 상태코드 403은 접근권한이 없음을 의미한다. 사용자별 서비스 웹은 사용자 코드에 따라 서비스 내용이 달라지며, 별도의 서비스 웹의 메인페이지를 두지 않고 403으로 접근권한 없음으로 처리하기도 한다. 이를 악용하여 공격자는 악성웹의 메인페이지를 403 오류로 악성 페이지를 위장한다. · · · · · · ·
· ·
· ·
· ·
· ·
· ·

ACKNOWLEDGMENT

Put sponsor acknowledgments.

참 고 문 헌

. . . . . .
. . . . . .
. . . . . .
. . . . . .
. . . . . .
. . . . . .
. . . . . .
. . .있다.

Ⅲ. 결론

본 논문에서는 . . . . . .
. . . . . .
. . . . . .

[1] Davies R. W." The Data Encryption standard in perspective,"Computer Security and the Data Encryption Standard, pp. 129-132.

[2] Miles E. Smid, "From DES to AES," 2000, (http//www.nist.gov/aes).

[3] Shamir, A. "On the security of DES," Advances in Cryptology, Proc.Crypto '85, pp. 280-285, Aug. 1985.

[4] NIST, "Announcing the Advanced Encryption Standard(AES),"FIPS PUB ZZZ, 2001, (http://www.nist.gov/aes).

[5] Daemen, J., and Rijmen, V. “AES Proposal: Rijndael, Version2.,” Submission to NIST, March 1999.

1)