情報保護學會論文誌
第 22 卷 第 4 號, 2012. 8

# 피싱사이트 실시간 탐지 기법*

사 준 호,† 이 상 진‡
고려대학교 정보보호연구원

# Real-time Phishing Site Detection Method*

Joon ho Sa,† Sangjin Lee‡
Korea University, CIST

요 약

최근 대다수 피싱사이트는 원시사이트(피싱사이트가 사칭하는 기관의 공식 웹사이트)와 유사하게 보이기 위해 원시사이트의 이미지, 게시글 등 컨텐츠를 링크하여 화면에 표시한다. 본 논문은 이러한 유형의 피싱사이트에 사용자가 접속하는 경우 피싱사이트의 URL이 HTTP referer 헤더필드를 통해 원시사이트로 유입되는 특성을 이용하여 피싱사이트를 실시간 탐지하는 시스템을 제안한다. 제안된 시스템은 원시사이트에 유입된 HTTP 트래픽을 아웃오브패스(out-of-path) 방식으로 수집하여 분석함으로써 홈페이지 실운영 환경에 대한 영향을 최소화하였으며, 원시사이트를 참조한 웹 사이트의 URL에 대해 휴리스틱 분석을 실시함으로써 피싱사이트를 실시간으로 탐지할 수 있도록 설계하였다. 제안된 시스템을 피싱사이트 표적이 되고 있는 국내 모 기관 홈페이지에 적용한 결과 6일 동안 40개의 피싱사이트를 탐지하였다.

ABSTRACT

Nowadays many phishing sites contain HTTP links to victim web-site's contents such as images, bulletin board etc. to make the phishing sites look more real and similar to the victim web-site. We introduce a real-time phishing site detection system which makes use of the characteristic that the phishing sites' URLs flow into the victim web-site via the HTTP referer header field when the phishing site is visited. The detection system is designed to adopt an out-of-path network configuration to minimize effect on the running system, and a phishing site source code analysis technique to alert administrators in real-time when phishing site is detected. The detection system was installed on a company's web-site which had been targeted for phishing. As result, the detection system detected 40 phishing sites in 6 days of test period.

**Keywords**: Phishing, Phishing Site Detection, Traffic Analysis

## I. 서 론

피싱사이트는 정부기관, 금융회사 등 신뢰할 수 있는 기관의 웹 사이트를 사칭하여 인터넷 사용자의 주민번호, 비밀번호, 신용카드 정보 등 개인정보를 탈취하는 악의적인 웹 사이트를 의미한다. 피셔(phisher, 피싱사이트 사기범)는 신뢰할 수 있는 기관들의 웹 사이트와 유사하게 생긴 피싱사이트를 개설하고 이메일, SMS 등을 발송하여 피해자들이 피싱사이트에 접속하여 개인정보를 입력하도록 유도한다.

APWG(Anti-Phishing Working Group)에 따르면 2011년 상반기 동안 전 세계에서 매월

접수일(2012년 2월 6일), 수정일(2012년 3월 27일), 게재확정일(2012년 3월 27일)
* 본 연구는 지식경제부 및 한국산업기술평가관리원의 산업원천기술개발사업의 일환으로 수행되었습니다. [10035157, 실시간 분석을 위한 디지털 포렌식 기술 개발]
† 주저자, bigcheese.james@gmail.com
‡ 교신저자, sangjin@korea.ac.kr

28,000~38,000여개 피싱사이트가 발견되었으며, 발견된 피싱사이트 중 72.7%는 금융 및 결제 영역을 대상으로 하고 있었다[1]. 이는 금전적 이익을 목적으로 한 피싱사이트 공격이 광범위하게 발생되고 있음을 보여준다.

반면 국내의 경우, 전자금융거래 시 보안카드, OTP 등 매체를 이용한 이중인증(two-factor authentication)이 보편화 되어 있어 피싱사이트로 인한 피해로 부터 상대적인 안전지대로 인식되고 있었다. 그러나 2011년 상반기부터 정부 및 금융기관 등 웹 사이트를 사칭하여 보안카드 전체 숫자와 신용카드 정보, 개인정보 등을 유출하고 획득한 정보를 이용해 공인인증서를 발급하거나 신용카드 카드론 서비스를 이용해 불법 예금이체를 시도하는 피싱사이트 사기범죄가 증가하기 시작하였다. 경찰청 사이버테러대응센터에 따르면 2011년 상반기 동안 경찰청과 KISA에 신고 접수 등으로 차단된 피싱사이트의 수는 총 125건에 달하며 피싱사이트들이 사칭한 기관은 경찰청 46건, 검찰청 34건, 농협 등 금융기관 25건, 한국인터넷진흥원(개인정보침해신고센터) 20건에 달하였다[2].

본 논문에서는 이러한 피싱사이트들 중 대다수가 원시사이트와 유사하게 보이기 위해 원시사이트의 이미지, 게시글 등 컨텐츠를 피싱사이트에 링크하는 것을 이용하여 원시사이트에 유입되는 HTTP referer 헤더필드 트래픽 분석을 통해 피싱사이트를 실시간 탐지하는 시스템을 설계하고자 한다. 시스템은 데이터 수집, 분석 및 리포팅 등 일련의 과정을 실시간으로 자동화하여 실시함으로써 피싱사이트 조기 발견이 가능하며, 사후에는 수집된 로그의 분석을 통해 피셔의 IP주소 확보에도 중요한 단서를 제공해 줄 것으로 생각한다.

본 논문의 2장에서는 피싱사이트 탐지를 위한 기존의 연구동향들을 살펴보고, 3장에서는 원시사이트의 컨텐츠를 링크한 피싱사이트의 특징과 이를 이용한 피싱사이트의 탐지 개념을 설명하였다. 4장에서는 시스템 구현 방법에 대해 설명하고, 5장에서는 구현된 탐지시스템을 국내 모 기관 웹 서버 환경에 적용하여 얻은 실험결과를 설명하였다.

## II. 기존 연구 동향

블랙리스트 기반 피싱사이트 탐지 기술은 국내외에서 가장 보편적으로 사용되고 있는 탐지기법이다. 블랙리스트 기술을 사용하는 솔루션으로 국내에는 한국인터넷진흥원의 웹체크[3]가 있으며, 해외에는 구글의 안전브라우징(Safe Browsing[4]), 마이크로소프트의 인터넷익스플로러 피싱필터(Internet Explorer Phishing Filter[5]) 등이 있다. Vienna 대학 연구진은 Phishing Tank를 통해 수집한 피싱사이트 샘플집단을 안전브라우징과 인터넷익스플로러 피싱필터에 실험하여 탐지율을 측정한 결과 안전브라우징이 90.23%의 높은 탐지율을 보였다고 설명하였다[6]. 블랙리스트 기반 탐지는 구현이 용이하며 오탐율이 낮으나, 안티바이러스 솔루션들이 직면하고 있는 문제점과 같이 알려지지 않은 피싱사이트를 탐지할 수 없다는 한계점이 존재한다. 특히 국내의 경우 해외에 비해 자체적인 피싱사이트 블랙리스트 수집 및 공유체계가 미비하여 블랙리스트를 이용한 피싱사이트 탐지에 어려움이 있다.

Johns Hopkins 대학과 구글의 연구진은 URL 구조를 분석하여 피싱사이트를 식별하는 방안을 소개하였다[7]. 로지컬회귀법과 구글의 페이지랭크(Page Rank), 도메인 기반 특징 분석, URL 키워드 비교 방법으로 2,508개 URL을 대상으로 실험한 결과 95.8%의 정탐율과 1.2%의 오탐율을 보임으로써 탐지방법의 효과성을 증명하였다. 하지만 이 방안은 공격자가 URL 패턴을 조금만 변경하여도 탐지율이 크게 저하될 수 있는 한계점이 존재한다.

Pittsburgh 대학과 Carnegie Mellon 대학 연구진은 웹 사이트 페이지 소스코드의 특징을 분석하여 피싱사이트를 식별하는 방안을 연구하였다[8]. 페이지 소스코드에 대해 TF-IDF 휴리스틱 산법, 도메인 수명, 로고 이미지와 도메인명의 불일치, 의심스러운 URL, 의심스러운 HTML 링크, IP주소, URL 중 "."(마침표)의 개수, 로그인 폼 존재여부로 구성된 8가지 항목을 분석하는 방법으로 100개 정상 웹 사이트와 100개 피싱사이트에 대해 실험한 결과 89% 정탐율과 1%의 오탐율을 보였다. 이 연구는 페이지 소스코드 분석을 통한 피싱사이트 탐지의 효과성을 증명하였으나, 페이지 분석에 소요되는 시간이 길며, TF-IDF 산법의 특성상 영어를 제외한 아시아국가 언어의 웹 사이트 분석에 한계점이 존재한다.

A. Kumar는 원시사이트 웹 로그에 HTTP referer 헤더필드가 IP주소로 되어 있거나 원시사이트의 기관명을 포함하는지 주기적으로 검토하여 피싱사이트를 탐지하는 아이디어를 제시하였다[10]. 그러나 APWG에 따르면 2011년 상반기 동안 발견된 피

싱사이트 중, 피싱사이트 주소가 도메인명 없이 IP주소로만 이루어진 피싱사이트는 전체의 4.5% 미만, 원시사이트 기관명을 포함하는 피싱사이트는 전체의 80% 내외이며[1], 피싱사이트의 평균 수명은 54시간 37분(약 2.3일)으로 나타나고 있다[11]. 따라서 관리자가 직접 웹 로그를 분석함으로써 발생하는 시간공백으로 인해 대응이 늦어지는 경우 피싱사이트는 이미 피해자가 접속하여 범죄에 악용된 이후이거나, referer URL에 IP주소와 기관명이 포함되는 특징만을 분석하여 판단하는 경우 다수 피싱사이트가 탐지되지 않는 문제점이 존재한다.

이외에도 .com, .net 등 최상위 도메인(Top Level Domain)의 존 파일(zone file)을 후이즈(whois) 정보와 함께 데이터베이스화하여 유사도메인을 사용하는 피싱사이트를 검출하는 방안도 존재한다[12]. 이 방안은 도메인에 일정 패턴이 존재하는 피싱사이트 조기 발견에 매우 효과적이나 도메인에 일정 패턴이 존재하지 않거나 존 파일을 외부에 공개하지 않는 도메인(.cn, .ru, .in 등)을 사용하는 피싱사이트의 탐지에는 한계가 있다.

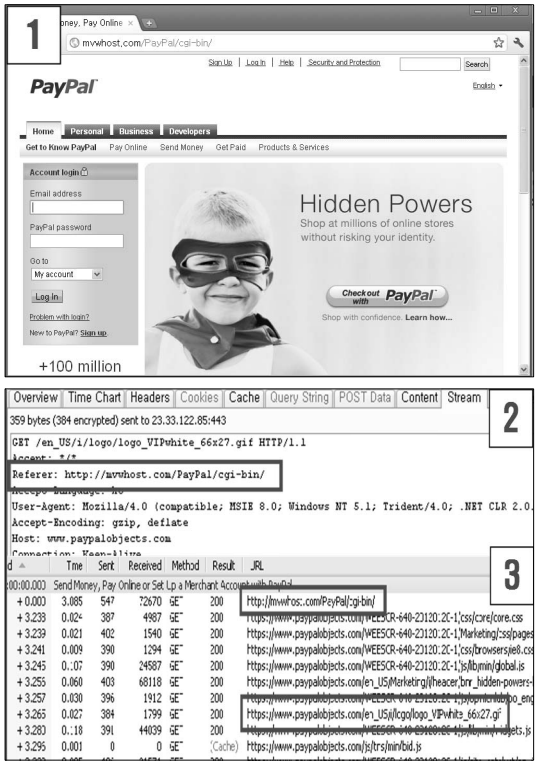
## III. 피싱사이트 특징 및 탐지 개념

### 3.1 원시사이트를 링크한 피싱사이트의 특징

피셔들은 피싱사이트를 원시사이트와 유사하게 보이기 위해 원시사이트의 이미지(.swf, .jpg 등)를 피싱사이트에 링크하여 표시하거나, 게시판 등 동적 컨텐츠를 링크하여 게시글, 메뉴 등을 클릭하였을 때에도 해당 내용이 원시사이트처럼 표시되게 한다. 피싱사이트가 내·외부 컨텐츠를 참조하여 화면에 표시하는 과정은 사용자 웹 브라우저에 의해 자동화하여 처리되는 과정으로, 사용자 화면에는 하나의 통일된 웹 사이트로만 보이게 된다([그림 1]의 1). 사용자가 접속 중인 사이트의 컨텐츠들이 실제로 어떠한 소스로 부터 전송되는지 확인하기 위해서는 페이지의 소스코드를 분석하거나, 또는 마우스 커서 지정 시 웹 브라우저 하단에 표시되는 링크 URL을 하나씩 관찰해야 한다.

### 3.2 원시사이트를 링크한 피싱사이트의 특징

RFC2616(HTTP/1.1)[13]에 정의된 HTTP referer 헤더필드는 클라이언트가 서버로 요청을 전송할 때 해당 요청이 어떠한 소스 URL에서 참조되었




[그림 1] 1) 피싱사이트 화면 2) 원시사이트로 전송되는 HTTP 요청의 헤더 내용 3) 피싱사이트 접속 시 원시사이트로 전송되는 HTTP 요청

는지 알려주는 기능을 제공한다. 이는 일반 웹 브라우저(인터넷익스플로러, 파이어폭스, 크롬 등)들에서 기본으로 준수하고 있으며, 링크를 통해 전송되는 모든 HTTP 요청의 헤더필드에 "Referer: 절대URI|상대URI" 형식으로 전송된다. 따라서 1절에서와 같이 원시사이트를 링크한 피싱사이트에 사용자가 접속하여 해당 링크를 클릭하거나 또는 원시사이트의 이미지 등이 피싱사이트에 링크되어 자동으로 표시되는 경우 피싱사이트 URL이 아래 과정을 통해 원시사이트로 전송된다.

1) 웹 브라우저는 링크가 지정하고 있는 원시사이트의 컨텐츠를 요청하기 위한 HTTP 요청헤더를 작성
2) 링크가 피싱사이트에서 참조되었으므로 HTTP 요청헤더의 referer 헤더필드에는 피싱사이트 URL을 기록([그림 1]의 2)
3) HTTP 요청 전송 시 피싱사이트 URL이 HTTP referer 헤더필드의 값으로 원시사이트

에 전송됨(〔그림 1〕의 3)

위의 특징을 이용하여 원시사이트에서 웹 서버로 유입되는 HTTP 요청 트래픽을 수집하고 HTTP referer 헤더필드의 URL 값들을 분석함으로써 원시사이트를 링크한 피싱사이트의 실시간 탐지가 가능하다.

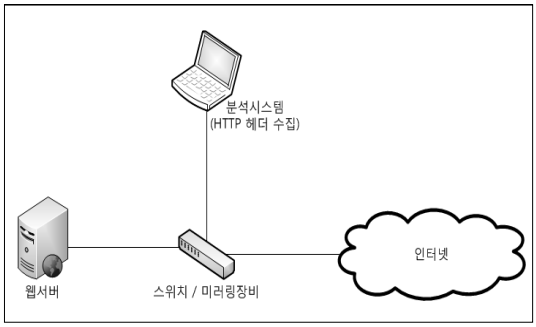
## IV. 피싱사이트 탐지시스템 설계

탐지시스템은 원시사이트로 유입되는 HTTP 트래픽을 분석함으로써 실시간으로 피싱사이트를 탐지하기 위한 목적으로 제작하였다. 따라서 웹 사이트 규모가 큰 경우 이에 해당하는 대용량 네트워크 트래픽을 실시간으로 분석할 수 있어야 하며, 트래픽은 응용계층의 HTTP 헤더를 분석하기 때문에 시스템에 부하가 크게 발생할 수 있으므로 효율적인 화이트리스트(white-list) 관리를 통한 프로그램 성능 및 로그 관리가 중요하다.

이를 위해 원시사이트로 유입되는 네트워크 트래픽은 〔그림 2〕와 같이 일련의 전처리 과정을 거친 뒤 로그에 증적되며 필요시 관리자에게 이메일, SMS 등을 통해 전달된다.

### 4.1 HTTP 트래픽 수집 및 전처리 과정

실운영 시스템에 대한 영향을 최소화하기 위해 〔그림 3〕과 같이 미러링(mirroring) 포트 또는 TAP장비를 이용해 아웃오브패스 방식으로 웹 서버에 유입되는 네트워크 트래픽을 수집하며, 분석 트래픽량의 최소화를 위해 웹 서버로 유입되는 인바운드 트래픽 중 HTTP GET과 POST 요청의 헤더를 수집한 뒤 다시 referer 헤더필드를 포함하는 헤더만을 걸러내어 다음 분석단계를 수행한다(〔그림 2〕의 과정 1,2,3).

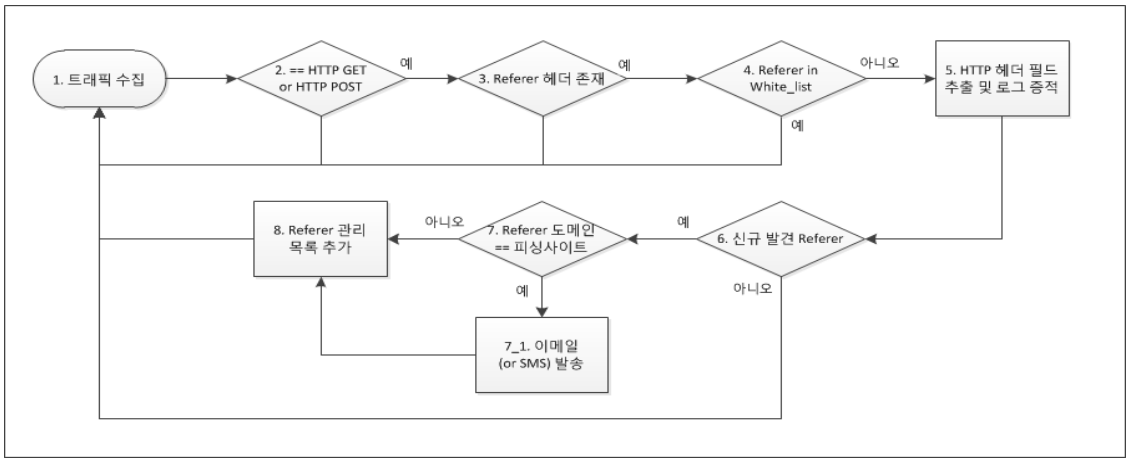



〔그림 3〕 탐지시스템 네트워크 구성

### 4.2 화이트리스트 비교

불필요한 트래픽 분석 및 로그 수집을 최소화하기 위해 수집된 HTTP 헤더에서 referer 헤더필드의 도메인이 화이트리스트에 존재하는지 비교한다. 화이트리스트와 일치하는 경우 해당 HTTP 헤더에 대한 분석을 중단하고 없는 경우에는 로그에 기록할 HTTP 헤더필드의 값들을 추출하여 로그에 기록한다(〔그림 2〕의 과정 4).

도메인 화이트리스트 비교 시 서브도메인을 가진 도메인(예, ime.korea.ac.kr과 cist.korea.ac.kr)은 서로 다른 도메인으로 구분되는 문제점을 해결하




〔그림 2〕 탐지시스템 동작 흐름도

기 위해 도메인의 유효도메인[14](예, blog. naver.com의 경우 naver.com, www.korea .ac.kr의 경우 korea.ac.kr)을 추출하여 비교한다.

화이트리스트에는 일반적으로 검색엔진, 포털사이트, 유관기관 등 웹 사이트의 도메인이 등록되며, 원시사이트 자체 도메인도 등록되는데 이는 원시사이트에 사용자가 접속했을 때 이미지 등 컨텐츠에 대한 요청이 원시사이트의 도메인을 referer로 하여 원시사이트에 전송되기 때문이다.

### 4.3 로그 증적

탐지시스템은 시스템 부하를 줄이기 위해 최소한의 로그만을 수집한다. 피싱사이트 탐지와 사후 피셔의 IP주소 추정 등 사고 분석을 위해 [표 1]의 헤더필드 항목들을 로그에 기록한다([그림 2]의 과정 5).

"time"은 원시사이트 링크에 대한 요청이 발생된 시점으로 특정 사용자가 원시사이트를 링크한 피싱사이트 페이지에 접속한 시점을 의미한다. "source ip"는 원시사이트 링크 요청자의 IP주소로써 피싱사이트에 접속한 사용자의 IP주소를 의미하는데 이 중에는 피싱사이트 피해자 이외에도 테스트를 위해 접속한 피셔의 IP주소가 존재할 수 있기 때문에 사고 분석 시 유용한 정보로 활용될 수 있다. "referer"는 원시사이트 링크가 존재하는 웹 페이지의 URL로써 피싱사이트 도메인을 기록하는 역할을 한다. "request"는 원시사이트로 요청된 URI로써 피싱사이트에서 원시사이트의 어떠한 컨텐츠를 링크하였는지 알려주는 역할을 한다. "accept-language"는 원시사이트 링크 요청자가 사용하는 웹 브라우저의 언어설정에 대한 정보를 기록하여, 피셔들의 국적을 파악하는데 도움을 준다(예, 중국어의 경우 "zh-cn"). "user-agent"는 접속자가 사용하는 웹 브라우저의 종류로써 검색엔진의 봇(bot)에 의한 접속 기록을 식별하거나 사용자를 구분하는데 도움을 준다. "x-forwarded-for"는 HTTP 헤더필드의 "X-Forwarded-For [15] " 값을 기록하여 접속자가 비익명(transparent) 프록시 서버를 경유하여 접속하는 경우 접속자의 원천지 IP주소를 확보하기 위한 용도로 사용한다.

### 4.4 Referer 목록 관리

수집된 referer 정보는 별도의 referer 목록을 두어 관리한다. [그림 2]의 과정 5에서 수집된 데이터의 referer가 referer 목록에 존재하는 경우 분석을 종료하며, 신규인 경우에는 해당 referer를 목록에 추가하고 피싱사이트인지 확인하는 절차를 수행한다([그림 2]의 과정 6, 7).

[표 1] 로그 항목 및 역할

| 항 목 | 역 할 |
|---|---|
| time | 원시사이트 링크에 대한 요청이 발생된 시점 |
| source ip | 원시사이트 링크 요청자의 IP주소 |
| referer | 원시사이트 링크가 존재하는 웹 페이지의 URL |
| request | 원시사이트로 요청된 컨텐츠의 URI |
| accept-language | 원시사이트 링크 요청자가 사용하는 웹 브라우저의 사용 언어 |
| user-agent | 원시사이트 링크 요청자가 사용하는 웹 브라우저의 종류 |
| x-for warded-for | 비익명 프록시를 경유하여 접속한 사용자의 원천지 IP주소 |

[표 2] Referer 목록 기록 항목 및 역할

| 항 목 | 역 할 |
|---|---|
| domain | 원시사이트를 링크한 웹 사이트의 유효도메인 |
| white | 화이트리스트 지정 여부(기본 값: "N(화이트리스트 아님)") |

Referer 목록에는 [표 2]의 항목들을 기록한다. "domain"은 referer의 유효도메인을 추출하여 기록하는데 그 이유는 유효도메인이 아닌 전체 URL을 기록하여 비교하는 경우 서브도메인이 다르거나 하위 경로가 포함된 URL들이 서로 다르게 인식되어 목록이 너무 방대해질 수 있기 때문이다(예, www.naver.com, blog.naver.com, www. naver.com/blog/를 모두 다르게 인식). 따라서 referer 목록의 "domain" 항목은 중복 기록되지 않으며, 기록된 모든 도메인은 원시사이트를 링크한 외부 웹 사이트의 유효도메인을 의미한다.

"white"는 해당 도메인이 화이트리스트 등록 여부를 결정하는 값으로 기본으로 "N(화이트리스트 아님)"이 설정되어 있으나 "Y(화이트리스트 맞음)"로 변경하여 화이트리스트로 설정하는 경우 해당 도메인으로부터 유입되는 트래픽은 더 이상 분석하거나 로그에 기록하지 않는다.

### 4.5 피싱사이트 판별

탐지시스템은 신규 도메인이 발견되면 해당 URL

에 접속하여 소스코드를 분석하는 방법으로 피싱사이트 여부를 판단한다. 본 연구에서는 이전에 발견되었던 시험 대상 기관의 피싱사이트 샘플들의 소스코드 특징을 분석한 결과를 토대로 아래의 탐지기준을 정의하였다.

1) 원시사이트에 대한 링크 수 5개 이상 존재
 - 피싱사이트 접속자가 게시글, 메뉴 등을 클릭했을 때 해당 컨텐츠가 정상 표시되도록 원시사이트를 링크하기 때문에 소스코드에 다수의 원시사이트 링크가 존재하는 경우
2) 웹 사이트 제목에 기관명 포함
 - 원시사이트와 유사하게 보이기 위해 홈페이지 접속 시 웹 브라우저 상단에 표시되는 웹 사이트 제목에 사칭하고자 하는 기관의 이름을 포함하는 경우
3) 특정 문자열 존재
 - 인터넷익스플로러, 크롬 웹 브라우저(파이어폭스, 오페라 제외)의 "다른 이름으로 저장" 기능을 이용하여 원시사이트를 복제하는 경우 HTML 소스코드에 생성되는 주석("〈!-- saved from url=(0034)http://원시사이트 도메인 --〉"), 피셔들이 사용자들이 피싱사이트에 개인정보를 입력하도록 유인하기 위해 사용한 문자열("개인정보침해신고센터") 등 문자열이 피싱사이트 소스코드에 존재하는 경우

신규 도메인에 대한 분석 결과가 위의 조건 중 하나라도 만족하는 경우 피싱사이트 위험성이 높은 것으로 판단하여 관리자에게 이메일, SMS 등을 발송하여 통보한다.

## V. 실험 결과 및 결론

4장에서 제안한 시스템을 검증하기 위해 2011년 10월 3일 21시부터 10월 9일 21시까지 피싱사이트의 주 표적이 되고 있는 국내 모 홈페이지의 DMZ 구간에 시스템을 적용하여 실험하였다.

실험 결과는 〔표 3〕과 같다. 144시간 동안 총 457개의 원시사이트를 링크하고 있는 홈페이지 도메인이 발견되었으며, 이 중 42개 도메인이 4장 5절의 탐지기준에 의해 피싱사이트 도메인으로 판별되었다. 457개 전체 도메인에 접속하여 육안으로 확인한 결과 실제 피싱사이트는 40개였는데, 40개 모두 탐지기준에 의해 피싱사이트로 판별된 42개 도메인에 포함되어 있어 탐지기준을 통한 피싱사이트 판별의 정확도가 100%에 달하는 것으로 나타났다. 나머지 2개 도메인은 정상 사이트이나 원시사이트의 링크를 5개 이상 포함하고 있어 피싱사이트로 오탐된 것으로, 이로 인한 오탐율은 0.48%로 나타났다.

〔표 3〕 피싱사이트 탐지 결과
(TP: True Positive, FP: False Positive)

| 원시사이트를 링크한 사이트 수 (A+B) | 정상 사이트 수 (A) | 피싱 사이트 수 (B) | 휴리스틱 탐지 수 | | 탐지율 | 오탐율 |
|---|---|---|---|---|---|---|
| | | | TP (C) | FP (D) | TP (C/B) | FP (D/A) |
| 457개 | 417개 | 40개 | 42개 | | 100% | 0.48% |
| | | | 40개 | 2개 | | |

탐지시스템 로그를 분석한 결과 34개(85%) 피싱사이트에 접속한 사용자의 IP주소는 모두 중국 IP주소로 나타났으며, 4개(10%)는 한국, 나머지 2개는 피싱사이트 자체(홍콩, 미국)의 IP주소로 나타났다. 분석 결과 한국 IP주소는 4개 모두 국내 웹 서버가 피셔들에 의해 경유지로 사용된 것이며, 피싱사이트의 IP주소 2개는 피셔가 피싱사이트 서버에 원격으로 접속하여 테스트하는 과정에 발생된 것으로 확인되었다. 즉, 탐지된 40개 피싱사이트 모두 피셔가 최초로 접속할 당시 탐지된 것으로 국내 피해자가 접속하기 이전에 성공적으로 탐지되었음을 알 수 있다.

위의 실험 결과가 증명하듯 본 논문에서 제안한 시스템은 피셔가 피싱사이트를 개설한 뒤, 피해자가 접속하기 이전에 실시간으로 높은 정확도로 피싱사이트를 탐지할 수 있음을 확인하였다. 본 논문에서 실험한 것처럼 시스템을 피싱의 표적이 될 수 있는 기관의 관제센터 또는 웹 서버 DMZ 구간에 설치하거나 또는 기존 IDS 장비 등의 확장 모듈로 개발하여 적용한다면 실운영 시스템에 영향을 최소화하면서 효과적으로 피싱사이트를 탐지해 낼 수 있을 것이며, 궁극적으로는 피싱사이트로 인한 개인정보 유출 및 금전적 손실을 방지하는데 기여할 수 있을 것으로 생각된다.

## 참고문헌

〔1〕 R. Manning, "Phishing Activity Trends Report 1st Half 2011," Anti-Phishing Working Group, pp. 3-7, Jul. 2011.
〔2〕 사이버테러대응센터, "공공기관 사칭 피싱사이트

주의보 발령," http://www.police.go.kr/announce/newspdsView.do?idx=97235, 2011년 7월.
〔3〕 한국인터넷진흥원, "웹체크 시스템," http://webcheck.kisa.or.kr/, 2010년.
〔4〕 Google, "Google Safe Browsing for Firefox," http://www.google.com/tools/firefox/safebrowsing/faq.html, 2007.
〔5〕 Microsoft, "Microsoft, Internet Explorer Phishing Filter," http://windows.microsoft.com /en-US/windows-vista/Phishing-Filter-frequently-asked-questions, Aug. 2010.
〔6〕 C. Ludl, S. McAllister, E. Kirda, and C. Kruegel, "On the effectiveness of techniques to detect phishing sites," DIMVA '07, pp. 20-39, Jul. 2007.
〔7〕 S. Garera, N. Provos, M. Chew, and Rubin, "A framework for detection and measurement of phishing attacks," WORM '07, pp. 1-8, Nov. 2007.
〔8〕 Y. Zhang, J. Hong, and L. Cranor, "CANTINA: A content-based approach to detecting phishing web sites," WWW '07, pp. 639-648, May 2007.
〔9〕 M. Jakobsson and S. Myers, Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft, Wiley, pp 48-49, Dec. 2006.
〔10〕 A. Kumar, "Referer Analysis - Mining for a Phisher's Traces," http://phishtrails.blogspot.com/2006/06/referer-analysis-mining-for-phishers.html, Jun. 2006.
〔11〕 R. Rasmussen, "Global Phishing Survey: Trends and Domain Name Use in 1H2011," Anti-Phishing Working Group, pp. 9, Nov. 2011.
〔12〕 사준호, "국내 피싱사이트 주요특징 및 대응방안," 금융보안연구원 이슈리포트, 20, pp 6, 2011년 11월.
〔13〕 R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1," RFC 2616, Network Working Group, Jun. 1999.
〔14〕 M. Still, "Python effective TLD library," http://www.stillhq.com/python/etld/000001.html, Oct. 2009.
〔15〕 Wikipedia, "X-Forwarded-For," http://en.wikipedia.org/wiki/X-Forwarded-For, Mar. 2012.

## 〈著者紹介〉

**사 진**

사 준 호 (Joon ho Sa) 정회원
2007년 7월: 중국 칭화대학교 컴퓨터공학과 졸업
2008년 3월~2012년 1월: 금융보안연구원 연구원
2009년 3월~현재: 고려대학교 정보보호대학원 석사과정
〈관심분야〉 정보보호, 디지털 포렌식

이 상 진 (Sangjin Lee) 종신회원
1987년 2월: 고려대학교 수학과 학사
1989년 2월: 고려대학교 수학과 석사
1994년 2월: 고려대학교 수학과 박사
1989년 2월~1999년 2월: 한국전자통신연구원 선임연구원
1989년 2월~현재: 고려대학교 교수
〈관심분야〉 암호이론, 디지털포렌식