홈페이지에 삽입된 악성코드 및 피싱과 파밍 탐지를 위한 웹 로봇의 설계 및 구현

김대유*, 김정태**

Implementation of Web Searching Robot for Detecting of Phishing and Pharming in Homepage

Daeyu Kim, Jung-Tae Kim

요 약

본 논문에서 제안하는 웹 서버 취약점 및 악성코드를 탐지하는 웹 로봇의 기술은 인터넷에서 개인정보보호사고의 원인분석을 통해 도출된 요구기능을 통합 구현하는 기술로 인터넷 이용자의 개인정보 피해 원인을 종합적으로 처리한다는 측면에서 효과가 크다. 인터넷에서 개인정보를 유출하는 홈페이지의 악성 코드 및 피싱과 파밍을 종합적으로 탐지기술을 구현함으로써 개인정보를 유출하기 위하여 사용되는 홈페이지의 악성 코드 및 피싱과 파밍 사이트로 유도되는 웹 사이트를 탐지 할 수 있는 시스템을 구현하였다.

ABSTRACT

Web robot engine for searching web server vulnerability and malicious code is proposed in this paper. The main web robot function is based on searching technology which is derived from analyses of private information threat. We implemented the detecting method for phishing, pharming and malicious code on homepage under vulnerable surroundings. We proposed a novel approachm which is independent of any specific phishing implementation. Our idea is to examine the anomalies in web pages.

키워드

악성코드, 웹보안 , 피싱

1. 서 론

컴퓨팅 환경이 웹으로 전환됨에 따라 다양한 웹 기반 응용 프로그램들이 개발되고 이에 따른 사용자 수 증가로 웹을 통한 트래픽이 증가하고 있다. 다양한 웹 기반 응용 프로그램들이 개발되고 사용자 수가 증가함에 따라 웹을 통한 트래픽이 증가하게 되었다. 웹 기반 서비스(WWW, Web mail, VOD, P2P, IM 등)의 증가는 유해트래픽의 다양한 통로로 사용되게 되고, 내부 정보 유출 등 피해가 확산되고 있다. 웹 기반 서비스(WWW, Web mail, VOD, P2P, IM 등)의 증가는 엄청난 기회를 창조하였지만 이와 더불어 유해 트래픽의 통로로 다양하게 사용됨으로써 불안정한 네트워크 환경을 조성하는 등 어려운 과제를 안겨주었다. 이 같은 유해 트래픽의 확산은 네트워크에 직접적인

* (주)위너다임 정보기술연구소 연구원
** 목원대학교 공과대학 전자공학과 교수

접수일자 2008. 10. 31

피해를 유발할 뿐 아니라, 최근에는 내부 정보 유출로까지 이어지고 있어 심각성이 나날이 증가하고 있다. 웹(World Wide Web)은 대부분이 정보제공이나 서비스 제공을 목적으로 하기 때문에 최근 해킹 사고의 대부분이 웹을 통해 이루어지고 있다. 2004년 말부터 급증하고 있는 웹페이지 변조공격과 같이 웹 어플리케이션의 취약점을 노린 해킹 사건들이 다수를 이루고 있으나 다른 인터넷 서비스들과는 달리 접근제어나 방화벽 등으로 보호가 어렵다. 결국 웹기반 서비스를 효과적으로 보호하기 위해서는 웹기반 서비스에 특화된 웹 어플리케이션 보안기술과 도구들이 필요하다. 응용서비스가 웹으로 변화함에 따라 주로 사용자의 개인정보 유출에 그 목적을 두고 있는 스파이웨어를 전파하는 새로운 기법으로 웹사이트를 이용하는 경향이 있다. 최근 홈페이지 변조사고는 단순 홈페이지 초기화면 변조보다 접속자 수가 많은 웹사이트를 해킹한 후 악성코드를 업로드하여 방문자 PC를 감염시킨다. 감염 후 특정 사이트에 접속하여 로그인할 경우 계정 및 비밀번호 정보가 공격자에게 유출되도록 하는 기법들을 사용한다. 웹 어플리케이션의 보안 솔루션으로 제안되고 있는 웹 방화벽의 경우, 프로그램 개발자의 실수를 룰에 의해 막아주는 기법으로 제안되어, 근본적인 문제해결이 될 수 없다. 웹 서버상의 잘못된 설정이나, 논리적인 프로그램의 오류로 인한 취약점을 점검하고, 룰에 의한 침입을 차단해 주는 수준이다.

2. 웹 유해코드 공격방법과 유형

2.1. Anthi Phising 개념

AntiPhish는 패스워드 같은 개인 사용자 정보와 해당 사이트의 도메인 정보를 유지하고 있다가 비 신뢰적인 웹 페이지의 폼 필드에 개인정보가 입력될 때 사용자에게 경고 조치를 취하여 피싱 사이트로의 접속을 방지 한다. AntiPhish는 모질라 파이어폭스 플러그인 형태로 개발 되어 있다. 개인 정보와 신뢰 도메인 목록을 비교하여 피싱 사이트로의 개인정보 유출을 차단 하는것을 목적으로 한다.

이를 위해 AntiPhish는 개인사용자 정보를 저장하기 위해 웹페이지 폼필드에 입력되는 내용에 대해 마스터 패스워드를 가지고 DES암호화를 수행한다. DES암호화된 개인정보는 해당 웹 사이트의 도메인 정보와 함께 저장된다. 사용자는 저장된 중요 정보와 웹 사이트 도메인 정보를 목록에서 볼 수 있으며 삭제 또한 가능하다. AntiPhish는 개인 정보와 인증된 사이트의 도메인 정보 목록과 유사한 피싱 사이트로의 개인정보 유출 차단 효과를 볼 수 있지만, 신뢰 도메인 내의 피싱 사이트가 존재 한다면 효과를 볼 수 없다는 단점이 있다. 본 기술은 인터넷 사이트에 대하여 사용자 입장에서 웹사이트가 위·변조 되지 않은 안전한 사이트임을 증명하여 인터넷에서의 정보를 신뢰성을 제고하고 전자 거래 시 논리적 해킹 내지 개인정보 사위행위 등을 예방하기 위한 것이다. 상기방법은 가짜 사이트를 식별하기 위해 인증된 사이트의 정보(도메인명, 디렉토리, 웹 페이지명)을 해쉬 알고리즘을 사용하여 해쉬 값을 얻고, 이를 이용하여 바코드 그림파일을 생성하고, 웹페이지에 첨부하여 인증된 사이트임을 표시한다. 위 기술을 사용하면 사용자를 속이기 위한 유사한 URL을 사용하고 원시 사이트의 바코드를 가짜 사이트에 부착한다면 사이트 인증 프로그램이 이를 인지하여 사용자에게 경고를 주어 가짜 사이트를 식별할 수 있다. 하지만 본 기술은 정상적인 사이트에 대해 서버 측에서 업데이트시 매번 바코드를 이용해 인증 사이트로 등록해 주어야 하는 부담이 있다. 또한, 국내외 수많은 사이트들이 존재하는 점을 감안할 때 범용으로 적용하기는 어렵다.

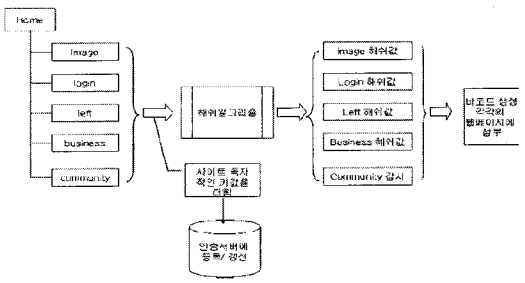

그림 1. Anti Phishing 개념도
Fig. 1 Block diagram of Anti Phishing

2.2. Anth Phising 개념

마이크로 소프트의 피싱 방지 매커니즘이 인터넷 브라우저에 플러그 인 될 것으로 보도되고 있다. 이 방법은 사용자가 피싱 사이트에 접속 시 우선 피싱 사이트에 공통적으로 나타는 문자열을 비교하여 일치하면 "Yellow Warning"을 내린다. 또한 접속 URL을 마이크로소프트 서버에 보내어 전송된 URL과 서버의 피싱 URL을 비교

하여 결과 값을 사용자에게 보내고 일치하면 "Red Warning"을 내리는 방식이다. 하지만 위 방식의 단점들은 다음과 같이 열거될 수 있다.

- MS사이트 접근 네트워크 경로가 단절되면 무용지물이다. 다시 말해 네트워크 접근 경로의 안정성이 확보되어야 한다.
- MS 사이트 응답속도 저하로 가용성이 저해될 수 있다.
- 사용자의 웹 사이트 방문기록 등 개인정보 노출 우려로 인한 서비스 활용성이 저하 될 수 있다.

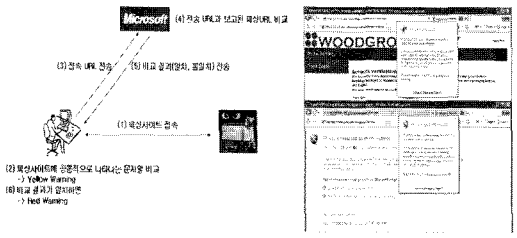

그림 2. 피싱 방지 메카니즘
Fig. 2 Block diagram of Anti Phishing Mechanism

2.3. 웹 유해코드의 공격방법

APWG에 접수된 보고(Phishing Archive)분석을 통해 피싱 사이트의 URL 표현 방법을 크게 세 가지로 분리된다.

<표 1> 악성코드의 형태
Table 1 Form of malicious code

구 분	유 형
Type 1	링크 표현은 정상적인 사이트를 나타내고 있으나 실제 접속 시 브라우저의 주소 창에 표시되는 내용
Type 2	URL이 http://www.usbank.com 인데 악성 사이트의 URL은 http://www.us-bank.com 로 표현
Type 3	사용자가 악성 사이트로 접속 시 자바스크립트를 내려 받게 하여 브라우저의 URL주소를 정상적인 URL것처럼 보이게 함

자료출처 : APWG

<표 2> 웹 유해코드의 특징
Table 2. Characteristics of web malicious code

공통점	1. Type1,2의 경우는 URL형태가 변하는 경우 2. Type3의 경우는 URL의 경우는 같지만 접속하는 IP가 다른 경우 유해코드가 삽입 된 사이트로 유도하여 악성코드가 삽입된 웹 페이지에 접속하게 하여, 사용자의 정보를 전송하는 방식은 모두 일치

3.3. 피싱공격형태

피싱의 공격 방법은 매우 다양한 방법이 존재하고 있으며 그중에 몇 가지 방법과 예제를 간단히 설명 하였다.

가) 유사한 이메일 주소 사용

실제e-mail: service@usbank.com,
위조e-mail: service@usbank-email.com

유사한 도메인 이름 사용

피해자가 접속하기 위한 위조 사이트의 도메인을 실제사이트와 유사하게 만들어 피해자로 하여금 도메인 구분이 어렵도록 하여 피해자를 속이는 방법이다.

실제 사이트 주소 - http://www.usbank.com
위조 사이트 주소 - http://www.us-bank.com

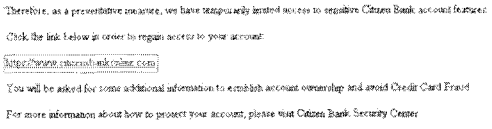
Therefore, as a preventative measure, we have temporarily limited access to sensitive Citizen Bank account features.
Click the link below in order to regain access to your account:
http://www.microsoft-ink.com
You will be asked for some additional information to establish account ownership and avoid Credit Card Fraud.
For more information about how to protect your account, please visit Citizen Bank Security Center

그림 3. 위조 이메일 내용
Fig. 3 Content of forged e-mail

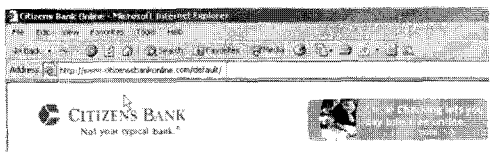

그림 4. 위조 웹사이트
Fig. 4 Forged web site

위의 예에서 공격자는 http://www.citezensbankonline.com을 클릭하였으나 실제 접속된 곳은 http://www.citezenscolorsbankonline.com/default 로 도메인 이름이 교묘하게 다른 것을 알 수 있다.

나) 이메일 주소 스푸핑

공격자가 메일발송 주소를 Spoofing하여 실제 메일발송자와 구분하지 못하게 하는 방법으로 실제 해당 기관의 메일주소와 똑같이 Spoofing 되므로 실제 메일 발송자가 누구인지를 구분하지 못하게 된다. 거의 모든 피싱 메일은 발신자의 이메일주소가 Spoofing 되어 있다.

Subject:eBay Account Verification
Date: Fri, 20 Jun 2003 07:38:39 -0700
From:"eBay"<accounts@ebay.com>
Reply-To:accounts@ebay.com
To:

그림 5. 메일 헤더 정보
Fig. 5 Information of mail head

발신자의 이메일 주소가 account@ebay.com으로 Spoofing 되어 발송된 이메일

다) 하이퍼 링크위조

HTML로 된 메일 본문의 링크에 표현되는 것은 실제 주소와 같으나 본문의 HTML 소스 코드 내에서는 전혀 다른 위조사이트의 주소로 링크 되어있어 실제 접속하게 되는 곳은 위조사이트가 되어 피해자를 속이는 방법이다.

Please follow the link below
and renew your account information.
http://www.fdic.gov/register/cgi-bin/fdic_intsafe/register.jsp

그림 6. 메일 헤더 정보
Fig. 6 Body of mail

그림 7. 연결된 웹브라우저 창
Fig. 7 Linked web browser

위의 그림과 같이 피해자가 링크를 클릭하였을 때, 메일본문에서 클릭한 주소와는 다른 주소가 웹브라우저 주소창에 나타나므로 사용자가 바로 확인이 가능하나 주의 깊게 주소창을 보지 않으면 피해를 볼 수 있다. 공격자가 미리 취약점이 있는 웹서버를 해킹하여 위조된 페이지를 만들어 두고 피해자의 접속을 기다리는 경우가 많다.

라) 스크립트를 이용한 주소창 위조

웹 브라우저의 주소창을 위조하여 피해자가 잘못된 링크에 접속하더라도 자바스크립트를 이용하여 정상적인 페이지에 접속한 것처럼 주소창을 위조하는 방법이다. 해당 스크립트는 먼저 스크립트가 사용 가능한 브라우저(Internet Explorer)인지 확인한 후 적합한 브라우저가 아닐 경우 바로 창을 닫고 적합한 브라우저일 경우 해당 스크립트를 실행하여 주소창에 잘못된 주소를 표시하게 된다.

If you are the rightful holder of the account, click the link bellow, fill the form and then submit as we will verify your identity and register you to CitiSafe free of charge. This way you are fully protected from fraudulent activity on all the accounts that you have with us.

Click to protect yourself from fraudulent activity!

그림 8. 메일 본문의 링크
Fig. 8 Linked mail body

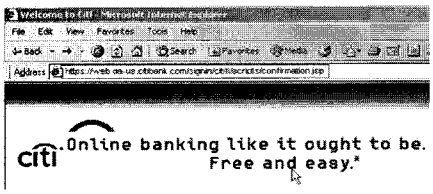

그림 8. 위조된 주소창
Fig. 8 Forged web browser

마) 팝업창 을 이용한 피싱 방법

피싱 메일의 본문 링크를 클릭시 백그라운드로는 정상적인 링크 페이지를 띄우고 Popup 창을 이용하여서는 공격자에 의해 위조된 피싱페이지를 띄우게 된다. 공격자는 메일 본문의 링크를 공격자가 만들어 놓은 페이지로 이동하게 하고 해당 페이지에서는 Popup 창과 정상적인 창을 바로 띄울 수 있게 되어있다. Popup창은 URL 이 보이지 않게 되어있으므로 사용자가 의심하지 않게 된다. 공격자는 Popup창에 중요정보를 입력할 수 있도록 만들어 놓고 피해자는 Popup창에 정보를 입력하게 된다.

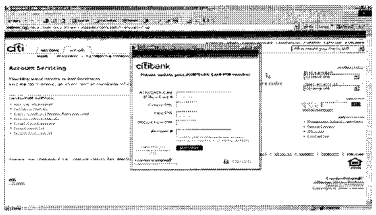

그림 8. 위조된 페이지로 생성된 팝업창
Fig. 8 Web browser by forged page

3. 웹 유해코드 탐지방법

3.1 구현설계

웹 수집 로봇의 수집과정에 웹 페이지의 소스코드를 다운로드하는 과정을 거치게 되는데, 해당 과정에서 HTML 소스코드에서 악성 코드를 검출 하는 기법을 도

입하여 (그림 9)와 같이) 웹 유해코드를 검출 할 수 있다.

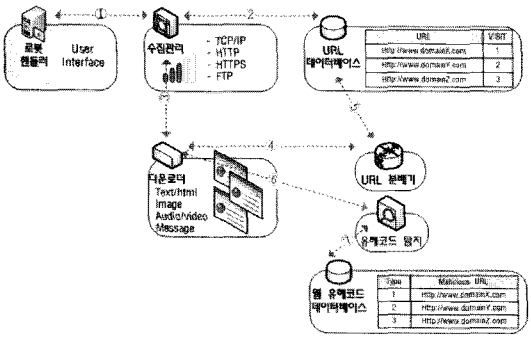

그림 9. 유해코드를 탐지 하는 웹 수집로봇의 구성도
Fig. 9 Configuration of web collecting robot for detecting malicious code

웹 유해코드는 페이지를 악성 사이트로 연결시키는 역할을 하고 있다. 이 연결될 수 있는 방법은 HTML 태그로 사용 되거나 자바스크립트를 이용하여 가능하다.

<표 3> 다른 페이지를 불러올 수 있는 태그 및 스크립트
Table 3 Tag and script for example

<IFRAME SRC= "http://www.link.com" >
<SCRIPT SRC = "http://link.com">
<OBJECT SRC = "http://link.com">

또한, 자바스크립트 함수를 이용할 경우에는 다른 현재 나타내있는 페이지에서 수많은 페이지를 불러오도록 할 수도 있다.

우선적으로 이 논문에서 제안하는 악성코드 탐지 규칙은 다음과 같이 5가지형태로 구분하였다.

1. 레퍼런스 웹 주소와 다른 웹페이지로 연결되는 링크들의 국가정보가 틀린 경우
2. 레퍼런스 웹 주소와 다른 웹페이지로 연결되는 링크들의 IP 대역이 다른 경우
3. 성데이터 베이스에 다른 웹페이지로 연결되는 링크가 포함되어 있는 경우
4. 악성 데이터 베이스에 다른 웹페이지로 연결되는 링크의 IP가 포함되어 있는 경우
5. 악성코드에 사용되는 스크립트 함수가 포함되어 있는 경우

위 5가지 조건 중, 1가지 형태가 검출된다면 악성사이트로 의심하고 보고하는 에이전트 형태로 진행을 계속하였다. 다음 (그림 10) 유해코드를 탐지하는 에이전트의 동작과정의 데이터베이스 점검로그에 해당 기록을 남기도록 한다.

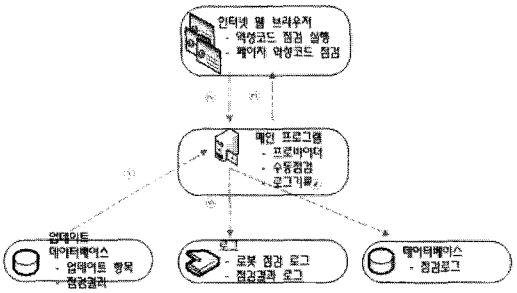

그림 10. 유해코드를 탐지 하는 에이전트의 동작과정
Fig. 10 Process of agent for detecting of malicious code

1. 업데이트 서버에는 (표 4) 악성코드 업데이트 항목을 업데이트 한다.

<표 4> 악성코드 업데이트 항목
Table 4. Update items of malicious code

형태	주소
피싱	http://site.co.kr/page.jsp
파밍	http://site.co.kr/page.asp
악성코드	http://site.co.kr/page.php
성인	http://site.co.kr/page.js
유해	http://site.co.kr/page.html

2. 웹 수집 로봇 핸들러로는 프로바이더(중앙제어) 모듈로 점검 대상사이틀르 전달한다.
3. 메인 프로그램의 프로바이더는 해당 사이트의 점검로그와 점검결과 (표 9)와 같이 나타낸다.

<표 5> 검출결과 내역
Table 5. Detailed contents of detection result

시각	형태	검출내역
08/9/28 2:00	<Iframe>	http://host.co.kr
08/9/28 2:00	<Script>	http://host.com
08/9/28 2:00	<Object>	http://host.co.kr

3.2 구현방안

처리 방법은 매우 간단하지만 위와 같은 방식은 피싱 에이전트 사용자에게 불편함을 줄 수 있다. 왜냐하면 Post Data가 발생할 때 마다 등록 여부 판단 메시지가 출

력되기 때문이다. 이 불편함을 없애려면 에이전트의 성능 부하로 처리가 늦어 질 수 있다. 사용자의 불편함을 조금 덜어주기 위해서 등록 여부 판단 메시지를 나타내기 전에 처리 하는 방안을 2가지로 제시해 보았다.

Type 1. Post Data 발생 Host를 DB 정상URL목록과 유사도 측정한다.

Type 2. Post Data 발생 전에 웹페이지 소스를 DB의 키워드로 검색하여 N-Gram 임계치와 비교 한다.

2가지 형식을 처리하게 되면 여러 가지 문제점이 발생할 수 있다. (유사도 측정 알고리즘과 N-Gram알고리즘)의 처리 속도가 있고, 100% 정확한 피싱 사이트를 판별 할 수 없어지게 되는 문제점이 있다. 이번 논문에서는 URL의 유사도를 측정하는 알고리즘이나 웹 페이지 변조 점검 등에 사용되는 알고리즘은 제외한 부분을 작성하였다.

4. 결론

웹 유해코드를 탐지 하는 기법으로 은닉된 변종 악성코드 검출로 변종 악성코드 확산을 원천적으로 차단하는 기술과 웹 환경의 안전성 확대를 통한 사이트의 신뢰성이 증가하여 해당 사업분야의 관련 기술 발전을 촉진할 것이다. 또한 웹 컨텐츠의 안전성을 제공하므로 다양한 형태의 컨텐츠 기술 발전에 긍정적인 영향을 미칠것으로 예상되며, 본 에이전트의 기술은 PC환경의 활용가능하다. 본 논문에서는 홈페이지에 삽입된 악성코드 연구를 통하여 산업 발전의 기여도 등 국가 경제에 미치는 효과로 보아 앞으로 "홈페이지에 삽입된 악성코드 탐지 기법 분석" 연구가 더욱 필요할 것으로 예상된다.

참고문헌

[1] http://cafe.naver.com/jmkim9064/778

[2] Li Zhuowei, etcs, "Utilizing Statistical Characteristics of N-grams for Intrusion Detection", Proceedings of the 2003 International Conference of Cyberworlds, pp.212-218.

[3] www.itfind.or.kr, "유비쿼터스 사회의 사이버 공격 기술 동향", 권호: 1259 발행일: 2006.08.16

[4] Joon S. Park and Gautam Jayaprakash, "Component Integrity Check and Recovery Against Malicious Codes" Proceedings of the 20th International Conference on Advanced Information Networking and Applications (AINA''06)

[5] Tony Abou-Assaleh, "N-gram-based Detection of New Malicious Code", Proceedings of the 28th Annual International Computer Software and Applications Conference

[6] Frank Adelstein, Matt Stillerman, "Malicious Code Detection for Open Firmware", Proceedings of the 18th Annual Computer Security Applications Conference (ACSAC.02)

[7] A. Murat Fiskiran, "Runtime Execution Monitoring (REM) to Detect and PreventMalicious Code Execution, Proceedings of the IEEE International Conference on Computer Design (ICCD'4)

저자소개

김대유(Daeyu Kim)

2008년 2월 : 목원대학교 대학원 IT공학과 박사과정
2006년 3월~2008년 2월 : 목원대학교 대학과 IT공학과 석사졸업
2006년 3월~현재 : (주)위너다임 정보기술연구소 연구원
※관심 분야: 웹보안, 개인정보보호, 네트워크보안 등

김정태(Jung-Tae Kim)

2001년 8월 : 연세대학교 대학원 전자공학과 박사
1991년 8월~1996년 2월 : 한국전자통신연구원(ETRI) 선임연구원
2002년 10월~현재 : 목원대학교 공과대학 전자공학과 교수
※관심분야: Microwave photonics, Optically fed wireless communication system design, Information security system design, Network Security, ASIC Design.