유사 사이트명을 가진 피싱 사이트의 접근 제어 구현 기술 분석

김대유, 김정태
목원대학교

Analyses of Detection Techniques of Phishing in the Web Site

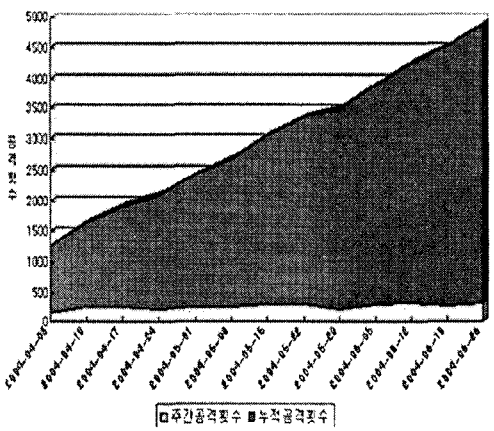
Daeyu Kim, Jung-Tae Kim
Mokwon University
E-mail : jtkim3050@mokwon.ac.kr

Abstract

피싱(Phishing)은 불특정 다수의 이메일 사용자에게 신용카드나 은행계좌정보에 문제가 발생해 수정이 필요하다는 거짓 이메일을 발송하여 관련 금융 기관의 신용카드 정보나 계좌정보를 등을 빼내는 해킹 기법으로써, 개인정보(Private data)와 낚시(Fishing)의 합성어로 낚시하듯이 개인정보를 몰래 빼내는 것을 말한다. 이 논문에서는 개인정보를 훔쳐가는 피싱의 유형과 방법을 분석하고 피싱(Phishing) 웹사이트를 탐지하는 방법을 제시 할 것이다.

1. 서론

Phishing은 1996년 American Online(AOL)을 사용하던 10대들이 일반 사용자들에게 가짜 이메일을 보내는 해킹 기법에서 유래됐으며 이들은 당시, 자신의 이메일을 AOL에서 보낸 이메일이라고 속이는 방법을 통해 일반 사용자들의 계정정보를 훔쳤다. Phishing은 온라인상에서 가짜 미끼를 걸어, 고객의 개인정보(Private Data)를 낚시질(fishing)하는 것을 의미하는 것으로 여기서 Private Data(개인정보)와 Fishing(낚시)의 단어가 합쳐져 Phishing이라는 단어가 탄생되었다. 이 단어는 당시 alt.2600이라는 해커가 주로 이용하던 뉴스그룹에서 처음 언급되었다. Phishing에 주로 사용되는 방법은 수신자가 원치 않는 이메일 또는 스팸 등을 발송하여 인터넷 사용자들을 Phisher들이 운영하는 웹사이트로 이동시키는데, 그 웹사이트들은 합법적인 전자 상거래 사이트처럼 위조되어 있다. 여기서 Phisher는 가짜 이메일을 보내어 사용자들을 속이고 정보를 빼내는 사람으로 Phishing 공격을 수행하는 사람을 의미한다. 또한 이러한 사이트에서는 사용자에게 계좌 정보를 갱신한다는 명목 하에 패스워드, 주민등록번호, 은행 계좌 혹은 신용카드 번호를 제공하도록 유도한다. Phisher들은 이렇게 획득한 정보를 다양한 용도로 사용하게 되며 Phishing 공격의 피해자는 해당 사실을 어느 날 자신이 사용하지 않은 내역이 다수 포함된 고지서 등과 같은 정보를 통해서 알게 될 것이다.



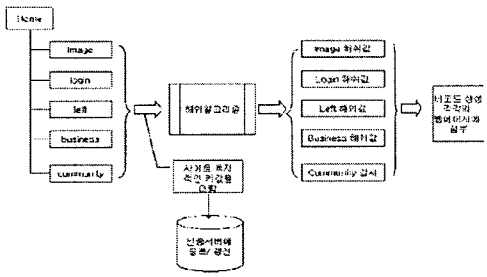
이러한 Phishing 범죄의 최근 경향을 다룬 Anti-Phishing Working Group(APWG)의 보고에 의하면 공격이 급증하고 있으며 올 6월에만 1,422개의 새로운 공격이 APWG에 보고되었으며, 이는 5월에 비해 19% 증가한 수치로 2004년에는 이러한 공격 보고가 매달 52%씩 증가하였다고 한다.

2. 관련 연구

2.1 AntiPhish

AntiPhish는 패스워드 같은 개인 사용자 정보와 해당 사이트의 도메인 정보를 유지하고 있다가 비신뢰적인 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 개인정보와 인증된 사이트의 도메인 정보 목록과 유사한 피싱 사이트로의 개인정보 유출 차단 효과를 볼 수 있지만, 신뢰 도메인 내의 피싱 사이트가 존재한다면 효과를 볼 수 없다는 단점이 있다.



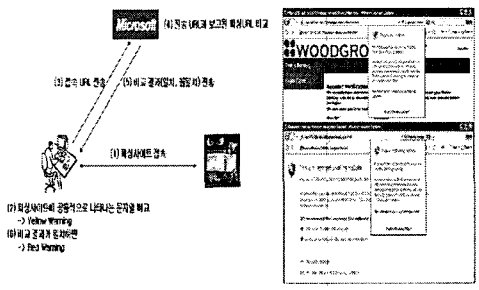
본 기술은 인터넷 사이트에 대하여 사용자 입장에서 웹사이트가 위·변조 되지 않은 안전한 사이트임을 증명하여 인터넷에서의 정보를 신뢰성을 제고하고 전자 거래 시 논리적 해킹 내지 개인정보 사위행위 등을 예방하기 위한 것이다.

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.

2.2 MS피싱 방지

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 비교하여 결과값을 사용자에게 보내고 일치하면 "Red Warning"을 내리는 방식이다. 하지만 위 방식의 단점들은 다음과 같이 열거될 수 있다.

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



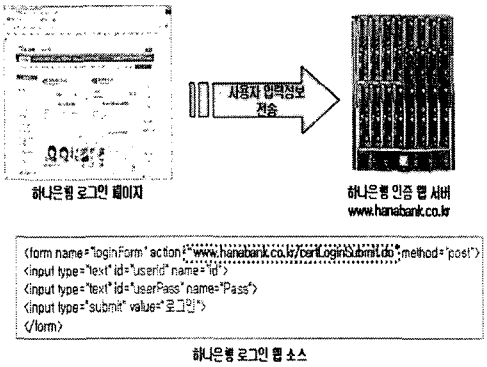
3. 피싱 사이트 탐지방법 제시

3.1 피싱 사이트의 공통점

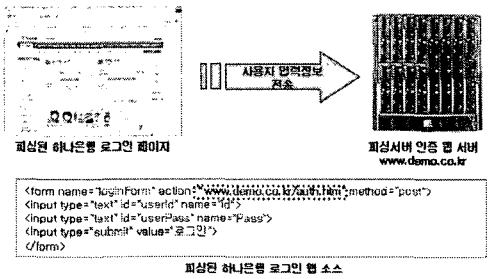
피싱 사이트의 형태는 크게 3가지로 분리 하였는데 Type1, 2의 경우는 URL의 형태가 변하는 형태였고 Type3의 경우는 URL의 형태가 같지만 다른 IP로 접속하는 형태였다. 위 3가지의 형식은 피싱 된 사이트로 유도하여 피싱된 사이트에 접속하게 하여, 피싱된 사이트에서 사용자의 정보를 전송하는 방식은 모두 일치한다. 위조 웹사이트에서 사

용자의 입력 값을 전달받은 방법이 일치한다는 것이다.

<form name=[폼 이름] method=[전송 방식] target=[타겟] action=[전송될 문서] enctype=[데이터 타입] autocomplete=[자동 완성 사용 여부]>



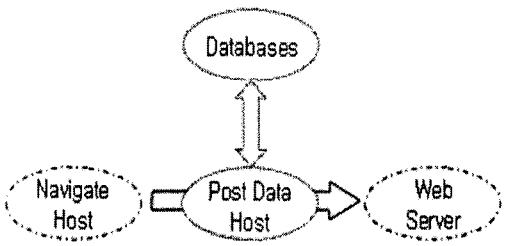
피싱의 공통적인 공격방법은 피싱 대상의 웹 사이트를 똑같이 복제 하여 form태그내의 Action 속성부분을 수정하여 해커의 웹 서버로 사용자가 입력하는 정보를 전달하게 하는 것 이다.



3.2 피싱 사이트의 탐지방법 제시

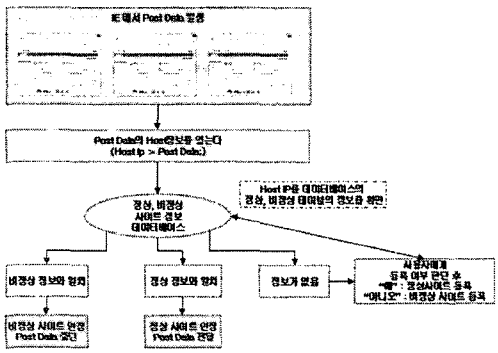
피싱 사이트는 사용자가 입력한 정보를 다른 서버로 전달한다. 사용자의 입력 값(Post Data)가 다른 서버로 전달되어 생기는 문제이다. 그래서 본 논문에서는 이 문제를 해결하기 위하여 사용자가 입력한 정보(Post Data)가 웹 브라우져를 통하여 전달 할 때의 시점에서 방안을 제시 한다.

본 피싱 탐지 에이전트에서는 데이터베이스에 정상사이트 목록과 피싱 사이트 목록이 존재해야 한다.



사용자의 입력정보(Post Data)가 웹 브라우져(Internet Explorer)를 통하여 전달되는 이벤트를 피싱 에이전트에서 가로채서 Post Data의 호스트를 IP로 변경하여 데이터베이스에 등록되어진 IP임을 확인하고, 등록되지 않은 사이트의 경우 에이전트에서 사용자에게 알린다. 이때 "정상사이트로 등록 하시겠습니까?" 라는 문구가 나타날 때 "예" 버튼을 누르면 정상 사이트 데이터베이스 항목에 들어가고 "아니오"를 선택하면 피싱 사이트 데이터베이스 항목에 추가 되며 다음부터는 해당 사이트에 전송을 하지 않게 된다.

데이터베이스 정보 (정상, 비정상)동일	
Host	IP
www.kbstar.co.kr	210.107.128.31
www.hanabank.co.kr	211.61.121.18
www.bestez.co.kr	210.92.0.200



3.3 구현방법

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

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 작성 하였다.

4. 결론

피싱 사이트의 탐지 방법은 매우 다양하지만 3가지 형식으로 분리 되고 있으며, 기존에 연구된 피싱 탐지 방법은 Post Data에 관한 부분을 언급하지 않고 있다. Post Data가 웹 서버로 전달 될 때 Host IP주소를 정상, 비정상 목록과 비교하여 처리하는 방안을 제안해 보았다. 하지만 정상, 비정상 목록을 전부 데이터베이스화 시키는데 문제가 있어 사용자에게 판단여부 메시지를 주게 되는데 이렇게 되면 사용자에게 불편함을 줄 수 있게 된다. 그 부분은 URL유사도를 측정하는 알고리즘이 추가 되어야 할 것이다. (문자열 유사도 측정) 또한 URL이 완전히 다른 피싱 사이트가 있을 수 있음으로 웹페이지의 소스를 N-Gram과 유사한 알고리즘으로 처리하는 부분이 더 개선해 나가야 할 것이다.

참고문헌

[1] Jong-Hyuk Roh, et, al, "Privacy Authorization for Internet Identity Management System", Journal of KICS, Vol.30, No.10, 2005

[2] G. Karjoth, et, al. "Platform for Enterprise Privacy Practices: Privacyenabled Management of Customer Data", LNCS 2482, 2002

[3] L. Cranor, et, al. "The Platform for Privacy Preferences 1.0 Specification", W3C, 2003

[4] E. Kirda and C. Kruegel, *""Protecting users against phishing attacks with antiphish,""* in *Proc. 29th Ann. Int. Computer Software and Applications Conf. - COMPSAC 2005* (Edinburgh, Scotland), July 26-28 2005, vol. 1, pp. 517-524.