웹 페이지 소스 코드에 기반한 보안 위험도 산정 기법

조상현*, 이민수**, 김영갑***, 이준섭**, 김상록**, 김문정***, 김인호****, 김성훈****
*엔에이치엔(주) 보안분석팀
**KAIST 전자전산학과
***고려대 정보경영공학전문대학원
****한국정보보호진흥원 기술정책팀
e-mail: bungae@nhncorp.com
**{mslee, srkim}@dependable.kaist.ac.kr
***{always, tops}@korea.ac.kr
****{kih, kimsh}@kisa.or.kr

Method for Evaluating the Security Risk based on Webpage Source Code

Sanghyun Cho*, Min-Soo Lee**, Young-Gab Kim***, Junsub Lee**,
Sangrok Kim** , Moon-Jeong Kim***, In Ho Kim****, Sung Hoon Kim****
*NHN Corporation IT Security Analysis Team
**Div. of Computer Science Dept. of EECS,
Korea Advanced Institute of Science and Technology (KAIST)
***Graduate School of Information Management and Security, Center for Information Security Technology (CIST), Korea University
****Korea Information Security Agency (KISA)

요 약

온라인 피싱과 같은 경제적 목적의 정보 수집 행위들이 급격히 증가하고 있는 가운데, 위험 사이트 정보를 관리하는 블랙리스트 기반의 대응과 신뢰할 수 있는 사이트를 기반으로 하는 화이트 리스트 접근 방법이 활용되고 있다. 그러나 블랙리스트 기반 방법은 피싱 사이트의 유효시간을 볼 때 비현실적이며, 화이트 리스트 접근 방법은 인증된 사이트 내에 존재하는 악성 웹 페이지나 악성 사이트로 유도되는 피싱 메일에는 대응하기 어렵다. 이 논문에서는 화이트 리스트 접근 방법을 보완하기 위해 사용자 웹 페이지의 위험도를 정량화 할 수 있는 웹 페이지 위험도 산정 기법을 제안한다.

1. 서론

온라인 피싱과 같은 경제적 목적의 정보 수집 행위들이 최근의 인터넷 오용의 대표 사례로 주목 받고 있는데, 근본적으로 사용자가 공격자가 만든 가짜의 페이지나, 유도 메일에 속지 않으면 안전하다. 그러나, 가짜 페이지를 구별하기 어렵도록 지능화된 형태로 공격자들이 웹 페이지를 만들고 있기 때문에 대응하기가 쉽지 않다.

기존의 피싱 방지 방법은 피싱 사이트의 URL를 블랙리스트로 구성하는 방법을 사용하였는데, 대부분의 피싱 사이트는 수일 내에 사라지는 특성이 있어 현실적이지 못했다 [1]. 다른 방법으로 블랙 리스트가 아닌 신뢰할 수 있는 사이트(도메인)를 모아 화이트 리스트를 구성하는 방법을 이용할 수 있다. 하지만, 이 방법은 대형 포털 사이트등에 존재하는 개인 블로그나, 홈페이지 혹은 신뢰된 사이트내에 존재하는 피싱 페이지등에 대응하기 어려운 제약사항을 가지고 있다.

이에 본 연구에서는 화이트 리스트 기반의 사이트 보안 위험도에 보완적으로 활용할 수 있는 웹 페이지 보안 위험도 정량 모델을 제안한다.

웹 페이지 보안 위험도 모델은 11개의 웹 페이지 위험도 평가 항목으로 구성되어, 3개의 샘플 페이지를 통해 유용성을 보여주었다.

2. 관련 연구

Millersmiles[1]과 APWG (Anti-Phishing Working Group) [2] 은 피싱, 파밍 공격을 방지하기 위하여 여러 기관 및 기업체 들의 데이터를 바탕으로 피싱 이메일 및 피싱사이트의 정보를 제공하고 있다.

대부분의 피싱 대응 연구는 피싱이 이메일을 통해 사용자로 하여금 피싱 사이트로 유도하고 있음에 착안해서 이메일 인증 방법에 비중을 두고 있다.

마이크로소프트가 중심이 되어 제안한 Send-ID [3] 기

법과, Yahoo가 주장한 DomainKey [4] 기법으로 나누어진다. 또한 S/MIME (Security/Multipurpose Internet Mail Extensions) [5] 과 같은 기법이 있다.

반면, 피싱 유도 메일이나 피싱 페이지 소스 분석을 통해 피싱 페이지 유무를 판단할 수 있는 속성들을 찾아내려는 연구도 진행되고 있다 [6].

3. 웹 페이지 보안 위험도 산정 기법

이 장에서는 웹 페이지 보안 수준 평가를 위하여 평가항목을 정의하고 이를 이용하여 웹 페이지 보안 위험도를 정량화하는 평가 단계를 제시한다. 평가항목은 [1]에서 공개하고 있는 피싱 메일과 피싱 유도 사이트의 특징을 종합분석하여 도출하였다.

3.1 웹 페이지 보안 수준 평가 항목

페이지 보안 수준 측정을 위해 페이지 내에 존재하는 링크와 입력 폼(Form) 그리고 포함하고 있는 컨텐츠를 분석한다

3.1.1 페이지내에서 관찰되는 URL 링크

- 링크의 수 : 페이지내에 존재하는 링크의 수는 피싱 위험도와는 큰 상관관계는 없으나, 피싱 유도 메일에서는 대부분 링크가 관찰된다. 그리고, 링크가 많은 페이지의 경우 스팸 혹은 악의적인 목적의 사이트일 가능성이 있다.

- 연결(link)된 외부 도메인 수 : 링크에 명시된 도메인이 원래의 도메인이 아닌 타 도메인으로 연결되는 형태의 피싱 사이트가 다수 관찰되므로, 페이지내에 존재하는 링크 중 외부 도메인으로 연결되는 경우는 고 위험으로 보도록 한다.

- 연결된 외부 사이트의 위험도 : 링크 되어 있는 외부 사이트의 도메인 위험도를 활용한다. 즉, 위험 도메인으로 연결될 수록 피싱 혹은 악성 사이트일 가능성이 높을 것으로 예상된다.

- IP 기반의 링크 : 피싱 유도 메일에서는 IP로 구성된 링크들이 발견되는 경우가 많다. 그러므로 도메인이름이 들어간 URL이 아닌 IP로 구성된 URL의 경우 의심해 볼 필요가 있다.

- 일치되지 않는 URL : 많은 피싱 사이트 혹은 피싱 메일의 경우 출력되는 LINK 정보와 다른 URL로 연결되는 경우가 많다. 따라서, HTML 문서 분석을 통해 <a href> 태그에 나타난 내용이 일치되는지 확인할 필요가 있다.

예) that 과 같은 경우 this와 that 이 일치되지 않는 경우

또한 텍스트로 된 링크 외에 이미지를 사용한 링크에 대한 고려도 필요하다.

- 추상적인 내용의 링크 : click here, link, this와 같이 링크 내용을 정확히 판단할 수 없는 경우 위험이 있을 것으로 예상된다.

- 링크 사이트에서의 도트(.)와 슬러시(/)의 수 : 여러 개의 도트(.)나 슬러시(/)를 이용하여 웹 디렉토리를 이동하여 악성 사이트에 연결되는 사례가 있으므로 이를 확인할 필요가 있다.

3.1.2. 페이지 입력 폼 분석

- Form 태그의 존재 유무 : 입력 폼이 존재하는 경우는 해당 페이지에서 사용자로부터 정보를 입력받을 수 있음을 의미하므로, 그렇지 않는 페이지에 비해서는 좀더 위험성이 높을 것으로 예상된다.

- 중요 정보 Field의 수 : 입력하는 필드의 수도 위험성과 비례할 것으로 예상된다. password나 주민등록번호 등 field 이름이나, manual analysis를 통해 private field를 판단한다. password나 주민등록번호, 카드 번호와 같이 의미상 중요한 요소를 입력받게 할 경우 위험 가능성이 높아진다.

- Hidden Field의 수 : HTML 소스 분석을 통해 숨겨진 필드를 쉽게 발견할 수 있기 때문에 이러한 필드의 수를 평가항목으로 활용한다.

3.1.3. 페이지 소스 분석

- 페이지 내의 스크립트 존재 유무 : 페이지 내에 악성적인 자바스크립트가 방화벽 등을 우회하기 위해 인코딩되어 포함되어 있는 경우가 있다.

3.2 웹 페이지 위험도 수준 측정 단계

본 절에서는 앞서 정의한 위험 수준 평가 항목을 이용하여 웹 페이지의 위험도 측정을 위한 세부 단계에 대해 기술한다. 위험도 측정은 아래와 같이 총 6 단계를 통하여 평가된다.

- 단계1 : 위험측정행렬에 위험요소 정의

본 논문에서는 웹 페이지 위험도를 측정하기 위해 그림 1과 같이 위험측정행렬을 제안한다. 위험측정행렬은 WQL DB에 등록된 항목을 이용할 수 있고, 웹 페이지에 관한 항목을 동시에 이용할 수 있다. 위험측정행렬에서의 위험요소란 위험발생을 야기시키는 요인(조건 또는 상황)을 말하며 앞서 정의된 위험수준 평가항목을 의미한다.

위험점수

위험요소	가중치	4	3	2	1	0	위험지수 (가중치*점수)
…							
…							
…							

(그림 1) 위험측정행렬

- 단계2 : 위험 요소들 사이에 가중치 부여

단계 1에서 정의된 위험 요소들 사이에 가중치(절대적 중요도 또는 상대적 중요도)를 부여한다. 본 논문의 예제에서는 가중치를 5점에서 1점까지 5점 척도를 이용한다.

- 단계3 : 위험 점수 측정

단계 3에서는 각 위험 요소별 위험 점수를 구한다. 위험점수란 위협발생을 야기시키는 요인이 형성될 수 있는 가능성 정도를 나타내는 것으로 0점 ~ 4점까지 5점 척도를 이용하여 측정한다. 이 때, 위험 점수 4는 위험발생을 야기시킬 가능성이 가장 높음을 의미한다. 각 항목 별 점수 부여는 통계 수치를 반영하여 각각 산출하며 4장의 예제에서 좀 더 자세히 볼 수 있다.

- 단계4 : 전체 위험지수 계산

단계 4에서는 웹페이지의 전체 위험지수(Total Security Index, TSI)를 구한다. 각 위험요소에 대하여 위험점수에 가중치를 곱한 값이 위험지수 (Security Index, SI)라 하고, 개별 위험지수들을 합한 값이 전체 위험지수이다.

- 단계5 : 최대위험지수 계산

단계 5에서는 최대 위험지수(Max Security Index, MSI)를 구한다. 최대위험지수란 웹 사이트가 가질 수 있는 이론적 위험 최대값으로써 본 논문에서는 0점에서 4점까지 측정된 값 중 4점으로 측정된 모든 위험요소들의 위험지수 합을 의미한다.

- 단계6 : 웹 사이트 위험도 계산

마지막으로 웹페이지 위험도(Webpage Security Risk Index, WPSRI)를 구한다. 단계 4와 5에서 구한 SRI와 MSI를 이용하여 WPSRI를 다음 식과 같이 정의한다. 보안 위험도 값은 0에서 100 사이의 값을 가지며, 값이 작을수록 안전한 웹사이트를 의미한다.

$$WPSRI = TSI / MSI * 100$$

4. 적용 사례

본 장에서는 앞서 제시한 웹페이지 보안 위험도 측정 기법에 대한 이해를 돕기 위해 적용사례를 제시한다.

- 단계1 : 위험측정행렬에 위험요소 정의

본 예제에서는 3.1절에서 정의한 11개의 위험 요소를 이용한다:

- 단계2 : 위험 요소들 사이에 가중치 부여

정의된 위험 요소들의 가중치는 절대적 중요도 따라 아래와 같은 기준에 의하여 계산된다. 본 예제에서는 3.2절에서 지정하였듯이 4점에서 0점까지 5점 척도로 측정한다.

- 단계3 : 위험 점수 측정

각 위험 요소별 위험 지수를 부여하기 위해 그림 2를 이용한다.

항목	설명	4	3	2	1	0
링크 수	링크 존재유무	링크 있음				링크 없음
링크 외부 도메인	링크에 연결된 외부 도메인 수	1개 이상				없음
링크된 외부 도메인의 위험도	도메인 위험도	80%이상	60%이상	40%이상	20%이상	20%미만
IP 기반 링크	IP로 된 링크사용 여부	사용				미사용
일치되지 않는 URL	페이지에 표시되는 URL과 실제 연결 URL의 일치 여부	불일치		일부 불일치		
링크 불확실성	click here, this와 같이 링크 내용 확인이 어려운 경우	사용				미사용
링크의 특수성	.과 /와 같은 특수 문자의 사용	사용				미사용
입력 FORM	사용자 입력 Form의 존재유무	있음				없음
주요 정보 입력 Field의 수	특정 정보 입력 field의 개수	10개 이상	5~10	5개 이상	1~5	1개이하
Hidden Filed	Hidden Field의 개수	5개 이상		1개 이상		없음
Embedded Script	인코딩된 Java Script의 존재유무	있음				없음

(그림 2) 웹 페이지 위험 지수 항목

그림 2의 측정 기준을 적용하기 위해 사용자가 접속한 웹 페이지는 다음과 같은 특성있다고 가정해 본다.

페이지내 외부 사이트로의 링크가 1개 이상 존재하며, 해당 링크 도메인의 사이트 위험도는 60%이상이다. 그리고 도메인 이름이 아닌 IP어드레스된 링크가 있으며 이 링크는 click here와 같은 이름으로 인터넷 익스플로어에 표현되고 있다. 웹 페이지에는 별도의 입력 창이 없으며, 자바

스크립트를 내장하지 않고 있다.

그림 2 의 측정 기준에 의해 단계 1에서 단계 3까지 진행되었을 때 위험측정행렬은 그림 3 과 같이 작성된다.

위험요소	가중치	4	3	2	1	0	위험지수
링크수	2	V					8
링크외부도메인	3	V					12
링크외부도메인위험도	4		V				12
IP기반링크	4	V					16
URL 불일치	4	V					16
링크 불확실성	1	V					4
링크 특수성	3					V	0
입력 폼 존재	2					V	0
주요 정보 입력 필드수	3				V		3
Hidden Filed	2					V	0
Embedded Script	5					V	0

(그림 3) 웹 페이지 위험 측정 행렬

- 단계4 : 전체 위험지수 계산

제시된 예에서 전체 위험지수(TSI)는 다음과 같다.

TSI = 8+12+12+16+16+4+0+0+3+0+0 = 71 이다.

- 단계5 : 최대위험지수 계산

제시된 예에서 최대위험지수 (MSI)는 다음과 같다.

MSI = 4* (2+3+4+4+4+1+3+2+3+2+5) = 132

- 단계6 : 웹 페이지 위험도 계산

단계 4와 5에서 구한 SRI와 MSI를 이용하여 WPSRI를 다음과 같이 구한다.

WPSRI = TSI / MSI * 100 = 71/132 * 100 = 53.79

위 결과로부터 접속한 웹 페이지는 53.79의 위험을 가지고 있다. 즉, 높은 보안 위험도를 가지고 있어 웹 페이지 접속시 주의할 필요가 있다.

앞서 제시한 보안 위험도 산정 기법을 통해 몇몇 웹 페이지의 보안 위험도를 측정하여 그림 4와 같은 결과를 얻었다. 샘플 A는 국내 금융권의 첫 페이지를, 샘플 B는 국내 유명 포털 사이트의 첫 로그인 페이지, 그리고 샘플 C는 피싱 페이지이다.

항목	가중치	샘플A	위험지수	샘플B	위험지수	샘플C	위험지수
링크수	2	4	8	4	8	4	8
링크외부도메인	3	0	0	4	12	4	12
링크외부도메인위험도	4	0	0	0	0	4	16
IP기반링크	4	0	0	0	0	0	0
URL 불일치	4	0	0	0	0	4	16
링크 불확실성	1	0	0	0	0	0	0
링크 특수성	3	0	0	0	0	4	12
입력 폼 존재	2	4	8	4	8	0	0
주요 정보 입력 필드수	3	0	0	1	3	0	0
Hidden Filed	2	2	4	4	8	0	0
Embedded Script	5	0	0	0	0	0	0
			20		39		64
웹페이지위험지수			15.15		29.55		48.48

(그림 4) 웹 페이지 위험도 사례

결과에서 보듯이, 실제 피싱 사이트 샘플인 C 페이지의 겨우 높은 보안 위험도를 보인다. 따라서 산출한 웹 페이지 보안 위험도가 어느정도 변별력을 가지고 있음을 보여준다.

5. 결론

본 논문에서는 피싱이나 파밍과 같은 공격에 대하여 웹 페이지의 보안 위험도를 산정할 수 있도록 하기 위하여 평가 항목을 정의하고 이를 이용한 웹 페이지 위험도 산정 기법을 제안하였다. 또한 제안 기법의 적용 사례를 통해 , 정량화된 위험 수치의 변별력을 보여주었다.

제안한 보안 위험도 산정 기법을 통하여 기존의 블랙리스트 기반 방법의 비현실적인 측면과, 화이트 리스트 기반 방법의 단점을 극복할 수 있다.

사이트 위험도를 바탕으로 한 화이트 리스트와 웹 페이지 위험도를 보완적으로 활용한다면 온라인 피싱과 파밍에 보다 효과적으로 대응할 수 있을 것으로 기대된다.

참고문헌

[1] http://www.millersmiles.co.uk
[2] Anti-Phishing Working Group. http://www.antiphishing.org
[3] Microsfoft, http://www.microsoft.com/mscorp/safety/technologies/senderid/default.mspx
[4] Yahoo, http://antispam.yahoo.com/domainkeys
[5] W. Stallings, "Cryptography and Network Security", Person Education, 2003
[6] E. Kirda and C. Kruegel, "Protecting Users Against Phishing Attacks", The Computer Journal Vol. 49 No. 5, 2006