개인정보보호를 위한 전자메일 주소 추출 방지 기법

김한섭°, 배수정, 연현정, 황윤철, 이상호
충북대학교 전자계산학과
{camwi°, sujebi, yonnolbu, ychwang, shlee}@cnlab.chungbuk.ac.kr

The prevention of extracting E-mail address for personal information protection

Han-Sop Kim°, Su-Jeong Bae, Hyun-Jeong Yeon, Yoon-Cheol Hwang, Sang-Ho Lee
Dept. of Computer Science, Chungbuk National Univ.

요 약

정보화 시대를 맞이하여 정보 인프라의 확충과 이를 기반으로한 다양한 비즈니스 모델이 출현하면서, 인터넷비지니스는 우리 산업의 중심으로 확산되고 있다. 이에 따라 개인과 개인, 기업과 기업, 기업과 개인간에 새로운 커뮤니케이션 수단으로 부상한 전자메일은 이제 우리 사회를 윤택하게 만드는 필수도구로 자리잡았다. 더군다나 전자메일은 그 파급효과나 비용절감의 측면에서 수많은 인터넷기업의 마케팅 수단으로써 범용적으로 활용되고 있다. 그러나 최근들어 불법 스팸메일과 유해성 메일이 급증하면서 전자메일 환경이 급속도로 악화되고 있어 그 위험수위가 극에 달해 있다. 이러한 불법 스팸메일은 전자메일 마케팅에 대한 신뢰도를 하락시켜 기업 활동을 위축시키고 있으며 유해성 메일에 무방비로 노출되어 있는 청소년 보호와 인터넷상의 개인정보보호 차원에서도 반드시 척결되어야할 "공공의 적"이라는 인식이 확산되고 있다. 하지만 최근 인터넷상의 공개된 전자메일 주소를 수집, 가공하여 스팸메일을 대량으로 발송할 수 있는 자동화된 프로그램의 개발로 인하여 스팸메일을 더욱 증가하고 있는 추세이다. 이 논문에서는 이러한 문제점들을 인식하고 자동화되어있는 전자메일 추출 프로그램으로부터 전자메일 주소를 보호할 수 있는 방안을 제시한다.

1. 서 론

정보통신기술의 발전은 그 개발과 상용화 및 확산과정에서 엄청난 사회적 함의를 내포하고 있다. 특히, 사회, 경제뿐만 아니라 가정, 개인의 삶까지도 영향을 미치는 총체적인 변화를 유도하고 있으며 산업의 각 분야에서 이러한 기술을 통하여 다양한 형태의 정보를 수집, 분석, 저장하고 유통시키려는 노력을 하고 있다. 이러한 사회·경제적인 활동은 기존의 생산활동을 더욱 효율화, 합리화시켜 사회 전반적으로 고도정보사회로의 진입을 촉진시키고 있다.

이러한 정보통신기술의 발전으로 급속히 보급되고 있는 것이 인터넷이라고 할 수 있다. 이용자가 급증하고 있으며, 인터넷 이용의 한 방법인 전자우편의 이용도 활성화되고 있다. 전자우편은 기존의 문서우편에 비해 그 편리성과 신속성으로 인해 인터넷의 활용측면에서 빼놓고 이야기 할 수 없는 중요한 수단이 되고 있다. 전자우편은 기존의 문서우편에 비해 논리적 공간에서 통신망을 이용해 빠른 속도로 다수에게 전달할 수 있으며, 발신자 확인이 용이치 않은 익명성을 띠는 특징을 지니고 있다.

전자우편의 이러한 장점을 살려 인터넷 활용의 극대화를 가져올 수 있는 반면 익명성, 신속성, 경제성, 대량성을 악용해 각종 오·남용사례가 나타나고 있다. 자신의 사업상 이익을 도모하거나 상대방에게 피해를 입히기 위해 수신자가 원치않는 불법적인 전자우편 발송이 남용되고 있는 실정이다. 이러한 원치않는 스팸메일의 수취로 인해 전자메일 사용자의 검색시간증가로 인한 업무 손실, 사용자 프라이버시 침해, 인터넷사업자(ISP)의 비용부담증가 및 전자우편 시스템의 훼손뿐만 아니라 인터넷 등 통신망의 이용질서를 무너뜨리는 사회적인 문제로까지 확산되고 있어 이의 현황파악과 대응책 마련이 시급한 실정이다.

2. 스팸 메일(Spam Mail)

일반적으로 스팸메일(Spam Mail)이라고 하는 것은 상업적 용도의 광고선전을 위해서나 불건전한 정보 등 괴롭힐 목적으로 불특정 다수에게 뿌려지는 전자우편을 이른다.

특히, 그 송신원을 은폐할 목적으로 발신원을 사칭하거나 제3자에게 중계하는 경우가 많다. 또 주소를 수집하는 소프트웨어를 이용하거나 매매되고 있는 주소목록을 사용하는 것 이외에 소프트웨어 툴로 생성한 주소를 이용하여 실존주소인지 아닌지 확인도 되지 않은 상태에서 무작위로 발신하는 경우도 많다. 이런 경우 수신측에도 문제가 발생하지만 발신하는 서버쪽에서도 문제가 발생하게 된다.

스팸메일의 발송을 위해서는 우선 그 대상인 전자우편 주소가 있어야 한다. 이를 위해 여러 가지 방법이 이용되고 있다.

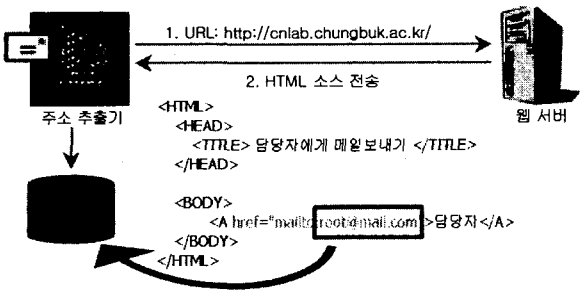
우선 스패머들이 손쉽게 취할 수 있는 방법은 인터넷 상에 공개된 전자우편 주소를 수집하는 것이다. 이러한 경우에는 '전자메일 추출기'라는 프로그램을 이용하여 적은 시간에 다량의 전자메일 주소를 수집할 수 있다.

또한 최근에는 컴퓨터 바이러스나 스니핑 같은 해킹기법을 이용하여 주소를 수집하는 경우가 늘고 있다. 원형태의 바이러스는 감염된 시스템의 중요정보를 수집하여 이를 자체적으로 전송하며, 네트워크의 패킷을 분석하여 전자메일 주소를 수집하는 시니핑기법 또한 증가추세에 있다.

이외에도, 이렇게 주소를 수집하는 경우가 아닌 임의적, 무작위적으로 주소를 생성하는 경우도 있다. 사전(dictionary)의 기본단어와 서버주소를 조합해 임의로 메일아이디를 제작해 발송하는 경우가 있고, 무작위적으로 철자를 조합해 임의의 메일아이디를 만들어 발송하는 경우가 있다. 그리고 많은 경우는 아니지만 유령사이트를 개설하여 마치 정보통신 서비스를 제공하는 사이트인 것처럼 꾸며 놓고 회원가입을 유도한 후 이를 스팸메일 발송이나 개인정보 매매 등에 이용하는 경우가 있다.

3. 전자메일 주소 추출 프로그램

전자메일 추출 프로그램은 일반 웹 브라우저(Web Browser)와 동일한 방법으로 웹서버에 자료를 요청하여 HTML형식의 문서를 전달받는다. 이때 전자메일 추출기는 서버로부터 전달받은 HTML 문서를 해석하여 사용자에게 보여 주는 것이 아니라, [그림 1]에서 처럼 HTML 문서 중에서 "mailto" 또는 "@"을 특정 구분자(Delimiter)로 사용하여 전자메일 주소만을 추출하여 DB에 저장하는 프로그램을 말한다. 이러한 프로그램에는 일반 검색엔진에 특정 검색을 요청 후 나오는 결과의 링크 사이트를 검색하는 방식과 사용자가 특정 홈페이지를 서핑하면 서핑된 Web 페이지들의 HTML 원문들이 PC의 하드공간에 임시 저장되는 특징을 이용하여 사용자 PC의 캐쉬 파일에서 전자메일 주소를 추출하는 방식, 바이러스와 같은 전자메일 주소 수집 에이전트를 이용하여 수집하는 방식이 있다. 현재 여러 종류의 전자메일 주소 추출 프로그램이 개발되어 있지만, 대부분의 경우에는 일반 검색엔진에 특정 검색을 요청 후 나오는 결과의 링크 사이트를 검색하는 방식을 사용하여 이메일 주소를 추출하는 방식을 사용하고 있다. 일예로 eCapture라는 프로그램은 "충북대학교 네트워크연구실"이라는 검색어로 30분 동안 5200여 개의 전자메일 주소를 추출하였다.



[그림 1] 전자메일 주소 추출기법

4. 전자메일 주소 추출 방지 기법

자동화된 전자메일 주소 추출 프로그램을 무력화시키기 위한 방법은 여러 가지가 있을 수 있다. 하지만 이러한 프로그램들은 [표 1]의 조건을 만족하여야 한다.

① 웹 서비스 전체 페이지에 적용가능
② 공개용, 상용게시판에서 방지가능
③ 모든 플렛폼(Platform)에서 동작가능
④ 서버 부하(Overhead) 최소화
⑤ 변환 알고리즘 유출시 추출프로그램 개발 가능성 최소화

[표 1] 전자메일 주소 추출 방지 소프트웨어의 요건

일반적인 모든 전자메일 주소 추출기들은 웹 페이지에서 전자메일 주소를 구분하는 특정 식별자인 "@"나 "mailto"를 검색하여 좌우의 문자를 전자메일 주소로 인식하여 추출한다.

그러므로 전자메일 추출 방지 S/W들은 이것에 착안하여 직접적인 전자메일 주소 스트링을 브라우저에 전달하지 않고 기존의 문장과 동일하게 작동 할 수 있는 스크립트 코드를 전송하는 방식을 사용함으로써 자동화된 전자메일 추출 프로그램을 무력화 할 수 있다. [표 2]는 대표적인 전자메일 주소 추출방지기법들을 비교하여 기술한 것이다.

ASS Mask	javascript를 이용 E-mail 주소를 여러 변수로 변환하여 HTML문장에 "mailto" 나 "@"의 좌우 문자들을 분리 브라우저에 E-mail출력과 링크는 그대로 유지
MaiIMask	javascript 이용 추출을 방지 16진 코드로 변환
MaiITo-Encrypter	10진 코드로 변환
Email Link Encrypter	16진 코드로 변환

[표 2] 전자메일 주소 추출 방지기법의 비교

하지만 이러한 전자메일 주소 추출 방지법은 모두 스크립트 코드의 패턴이 분석될 경우, 전자메일 주소 추출을 방지할 수 없다는 문제점이 있다.

5. 특정 ID값을 사용한 전자메일 주소 추출 방지 기법

지금까지의 전자메일 주소 추출방지기법은 실질적인 전자메일 주소를 웹 브라우저에 전달을 한 후에, Javascript나 16진 코드, 10진 코드로 변환하여 기존의 자동화된 전자메일 추출 프로그램으로부터 정보를 보호한다. 하지만 제안하는 전자메일 주소 추출방지법은 전자메일 주소를 웹 브라우저에 전달하지

않고, 이를 대신할 수 있는 특정ID값을 웹 브라우저에게 전달을 하고 사용자가 이벤트를 발생할 때에 서버에 연결하여 스크립트화된 소스를 전달하여 사용자의 요구를 실행한다.

이 기법은 HTML 소스상에 전자메일 주소의 표기 자체를 금지하여 원천적으로 전자메일 추출 프로그램의 시도를 차단한다. 또한 특정ID값과 Javascript의 함수 이름을 필요에 따라 변경함으로써 만약에 있을지 모르는 패턴분석 공격에 대한 방어를 할 수 있다.

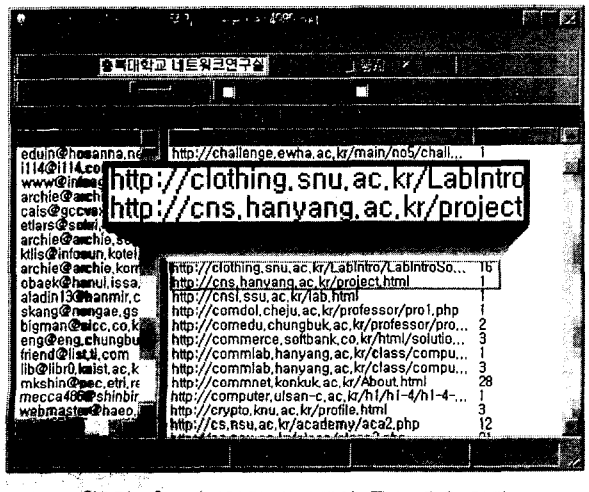
이러한 기법은 [그림 2]에서 보여지는 것처럼, HTML 문서내의 특정 ID값을 스크립트를 통하여 CGI Program에게 전달하고, 그 값이 사용자의 올바른 사용에 의해 발생된 요구인가를 검증한다. 검증이 끝난 후에는 전자메일 주소 DB에 접근하여 전자메일 주소를 검색한 후, 그 주소를 스크립트화하여 사용자에게 전달한다.

[그림 2] 제안하는 전자메일 추출 방지기법의 구조

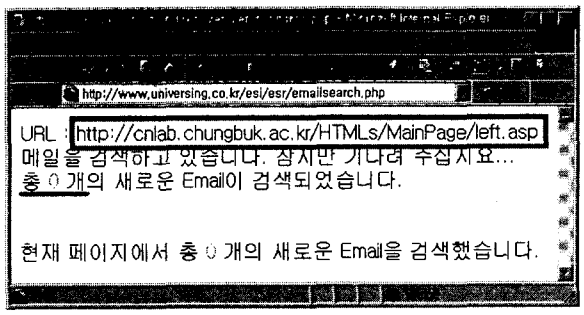
이 기법에서 사용하는 스크립트의 헤더정보에는 사용자의 시스템에 캐쉬파일이 생성되지 않도록 설정하여, 사후에 사용자의 시스템에서 이메일주소를 추출하는 기법에 대하여 대비한다.

또한 이 기법은 CGI Program을 사용하기 위한 스크립트의 이름을 사용자가 임의로 설정할 수 있고, HTML 문서 어디에나 제한없는 하이퍼링크와 Javascript를 비롯한 여러 가지 스크립트 언어를 사용할 수 있으므로, 자동화된 전자메일 주소 추출 프로그램으로부터 안전하다.

[그림 3]과 [그림 4]는 현재 운영중인 충북대학교 컴퓨터네트워크연구실의 홈페이지(http://cnlab.chungbuk.ac.kr/)에 제안하는 전자메일 추출 방지기법을 적용한 후에, 현재 일반적으



[그림 3] 게시판형 전자메일 주소 추출 결과



[그림 4] 페이지형 전자메일 주소 추출 결과

로 사용되고 있는 전자메일 주소 추출프로그램을 실행한 결과이다. 우선 [그림 3]의 경우에는 "충북대학교 네트워크연구실"이라는 검색어로 추출되는 전자메일 개수와 그 전자메일을 추출한 게시판의 주소정보이다. 이 결과에서는 "충북대학교 네트워크연구실"의 홈페이지(http://cnlab.chungbuk.ac.kr/)에서 추출된 전자메일 주소가 없음을 보여주며, [그림 4]도 관리자의 메일이 연결되어 있는 특정페이지에서 전자메일 주소가 추출되지 않음을 확인할 수 있다.

6. 결론

정보통신기술의 발전과 인터넷의 보급확산으로 인류는 종래의 우편제도를 통한 의사소통방식 외에 전자우편에 대한 의존이 증가하고 있다. 하지만 전자메일의 부적절한 사용으로 인하여 전자메일 사용자의 검색시간증가로 인한 업무 손실, 프라이버시 침해 외에 인터넷사업자(ISP)의 비용부담증가 및 전자우편 시스템의 훼손 등으로 인한 심각한 손실 등이 발생되고 있다. 이에 따라 본 논문에서는 기존의 인터넷 사용기법을 분석하여 기존 사용자들이 전자메일 서비스를 사용함에 있어 불편함이 없는 전자메일 주소 추출을 방지 기법을 설계하고 구현하였으며 그 효율성을 여러 가지 전자메일 추출 프로그램을 사용하여 입증하였다. 하지만 HTTP 프로토콜의 특성상 완벽한 전자메일 추출 방지 프로그램을 개발할 수 없었고 100% 웹브라우저의 기능을 구현하는 전자메일 추출 프로그램의 등장시에 방어체계가 깨질 수 있다는 문제점에 대한 향후연구가 필요하다.

7. 참고문헌

[1] 한국정보보호진흥원, 전자우편 사용자 보호에 관한 연구
[2] Paula Kurtzweil Walter, "Chain Emails: Just Another Ploy or the Real McCoy?", Federal Trade Commission
[3] Richard H. Stern, "Essential for sending junk e-mail?", IEEE Micro
[4] Tractebel Energy Engineering, "Spam: The Plague of Junk E-mail"
[5] Deirdre K. Mulligan, "Report to the Federal Trade Commission of the Ad-Hoc Working Group on Unsolicited Commercial Email"
[6] Moore, K "MIME(Multipurpose Internet Mail Extensions", RFC 1522, 1993
[7]Jonathan B Postel, "Simple Mail, Transfer Protocol", RFC 821, 1982