제 19 회 한국정보처리학회 춘계학술발표대회 논문집 제 10 권 제 1 호 (2003. 5)

PGP 기반의 성인광고메일 차단용 메일시스템

°김성식*, 안양재*, 김중환*, 김상철*
*한국외국어대학교 컴퓨터공학과
e-mail : kss1213@hotmail.com

A PGP-based Mail System for Blocking Adult Mails.

Seong-Shig Kim*, Yang-Jae Ahn**, Joong-Hwan Kim**, Sang-Chul Kim**
*Dept. of Computer Science Engineering, Han-Kook University of Foreign Studies

요 약

정보화 사회가 도래하고 우편이나 통신체계도 물리적인 공간에서 인터넷(Internet)이라는 가상 공간으로 점차 옮겨져 광범위하게 사용되고 있다. 전자우편의 중요성이 대두되면서 스팸 메일(Spam-Mail)로 인한 여러 가지 피해들이 속출하고 있다. 특히, 많은 사람들이 성인광고광고 스팸 메일로 인해 상당히 곤혹스러운 경우를 경험을 하고 있다. 본 논문에서는 PGP(Pretty Good Privacy)의 개념을 이용해서 사용자 인증(User Authentication) 기능을 수행하고, 인증되지 않는 사용자가 보낸 메일에 대해서는 텍스트 분석 뿐만이 아니라 이미지와 동영상을 처리해서 성인광고 메일 여부를 판단하는 메일 시스템을 제안한다. 우리의 조사에 의하면, 성인광고메일을 차단하는 메일시스템에 관한 연구는 거의 발표되지 않고 있다.

키워드 : PGP, Spam Mail, 성인광고메일

1. 서론

인터넷을 통해 움직이는 전체 데이터 통신량 중의 많은 부분이 전자 우편이 차지하고 있다. 일부 ISP 업체에서는 자신들의 메일서버 저장공간의 10~45 퍼센트가 스팸 메일로 채워지고 있으며, 최근 개개인의 경험으로도 전송되는 전자우편 중에서 상당히 많은 양의 메일들이 스팸 메일이며 그 중 성인광고 메일들이 대부분을 차지하고 있다는 것을 알 수가 있다. 이러한 상황에서 우리는 성인광고 메일이 청소년들에게 전달된다거나 성인광고메일로 인해 여러 가지로 당황스러운 상황에 처하는 경우가 많다.

메일시스템에서의 보안성을 위한 기술로 PGP 나 PEM 을 제안하고 있다. 현재의 대다수의 메일시스템에서 안전성을 보장 받기 위해서 PGP 을 주로 사용하고 있는 실정이다.[1]

본 논문에서는 PGP 이용한 메일시스템에서의 성인광고성 메일을 차단하는 방안을 제안한다. 전자우편은 일반적으로 텍스트뿐만 아니라 이미지, 동영상으로 구성된다. 본 논문에서 제안하는 메일시스템에서는 일차적으로 송신자의 메일주소와 소속 등을 판단하여 분석하고, 다음으로 문자열과 이미지을 분석하여 성인광고 메일인 여부를 판단한다. 만약 성인광고 메일로 판단되면 이미지 삭제, 거부, 경고메시지와 같은 다양한 수준의 차단을 제공함으로써 사용자의 요구에 맞게 선택할 수 있게 하였다.

기존에는 스팸 메일 차단에 대한 연구는 많이 수행되어 왔지만, 우리의 조사에 의하면, 성인광고 메일에 대한 연구는 거의 찾을 수가 없었다. 관련연구로서 유해사이트 차단을 들 수가 있는데, 이것은 유해사이트 데이터베이스를 사용하는 단순한 방법으로 제한되어 있는 것으로 판단된다.

본 논문의 구성은 다음과 같다. 2 장에서는 보안 메일시스템을 위한 PGP 와 스팸 메일 방지기술에 관한 기종 연구를 소개하고, 3 장에서는 본 논문에서 제시한 PGP 를 이용한 성인광고메일 시스템을 기술했으며, 4 장에서는 실험 및 분석 그리고 마지막 5 장으로 결론 및 향후 과제를 기술한다.

2. 관련 연구

2.1 PGP

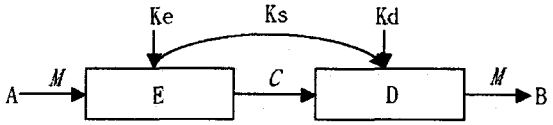
PGP 에서의 보안요소에는 몇 가지 기능들이 있다. 기밀성(Confidentiality), 메시지 인증(Message Integrity),

사용자인증(User Integrity), 송신부인방지(Nonrepudiation of origin), 수신부인방지(Non-deniability of receipt), 메시지 재현 방지(Message replay prevention)가 있다. 현재 PGP 에서는 위쪽의 앞 네 가지 기능은 지원하고 있는데 기능과 알고리즘을 보면 아래 표 1 과 같다.

<표 1> PGP의 기능 & 알고리즘

기 능	알 고 리 즘
기 밀 성	IDEA, RSA
전자서명(무결성, 사용자인증, 송신부인방지)	RSA, MD5
압 축	ZIP
전자우편과의 호환	Radix-64 Conversion

또한, 어떤 비밀정보를 전달하려는 송신측 A 는 평문 *M* 을 암호화 알고리즘 E 와 암호화키 Ke 를 이용해서 암호화 한 뒤 암호문 *C* 로 생성한 뒤 전달하게 된다. 전달된 메시지는 수신측 B 에 다시 복호화 알고리즘 D 와 복호화키 Kd 를 이용하여 수신측 B 에서 알 수 있도록 평문 *M* 으로 만들어 낸다.



A = 송신자 B = 수신자
E = 암호화 알고리즘 D = 복호화 알고리즘
M = 평문 C = 암호문
Ke = 암호화키 Kd = 복호화키

(그림 1) 기본 메카니즘

PGP 에는 관용암호방식과 공개키 암호화방식이 있다. 관용 암호방식에서는 암호화 키 Ke 와 복호화키 Kd 가 서로 같은 경우로써, 송신자 A 와 수신자 B 가 세션키(Session key)Ks 를 공유해야 한다. 여기에서 송, 수신자 A, B 는 공유하고 있는 세션키 Ks 를 이용하여 암호문을 만들며 복호화 역시 공동으로 가지고 있는 세션키로 복호화를 한다. 이 방식은 문서의 암호화, 복호화 과정은 빠르나 세션키 공유방법이 또 다른 문제로 떠오르고 있으며, 세션키의 공유방법이 안전하지 못하면 정보의 보호를 보장할 수 없게 된다. 그리고 관용암호화방식과는 달리 이 방식의 단점을 해결한 방법인 공개키 암호방식에서는 두개의 키를 사용하고 있다. 암호화키 Ke 와 복호화키 Kd 가 서로 다른 경우인데, 모두에게 공개된 공개키(Public Key)와 자신만이 가지고 있는 비밀키(Secret Key or Private Key)로 송신자 A 는 공개키로 암호화된 문장을 송신자 A 의 비밀키를 사용해야만이 복호화 할 수 있다. 공개키 암호화방식은 암호화, 복호화에서의 수행능력이 좋지 못해서 직접적으로 사용하지는 않고, 관용암호방식에서 사용했던 세션키 Ks 의 암호화, 복호화에 사용되며, 대표적인 것으로는 RSA 가 있다.[2][3][4][5]

2.2 스팸 방지 기술

우리 조사에 의하면, 스팸 메일 방지 기술에서 성인광고 메일차단에 관한 연구는 발견할 수가 없었으며, 관련연구로서 웹 브라우져에서 유해사이트를 차단해서 여러 가지 피해를 줄이려고 기술들을 개발하고 현재 상용화되고 있는 실정이다. 그러나, 웹 브라우져상에서의 유해사이트차단은 일반적으로 URL 데이터베이스를 이용해서 차단하는 단순한 구조를 가지고 있다.

스팸 메일 차단에 관한 연구는 여러 가지로 제안을 하고 있는데 아래와 같다. 스팸 메일을 방지할 수 있는 접근법은 세가지 관점에서 접근할 수가 있는데 수신자 접근법, 서비스 제공업체 접근법, 입법적인 접근법이 있다.

일반적으로 수신자 접근법은 수신자가 들어오는 메일을 필터링할 수 있고 메일의 소재를 파악할 수 있으며, 스패머(스팸 메일을 보낸 사람)에 대해 복수할 수 있는 형식을 취할 수도 있다. 그리고 서비스제공업체 접근법은 서비스제공업체는 확인된 스패머에 대해서는 서비스를 거부할 수 있고, 자신들의 메일 서버를 남용하지 못하도록 기술적인 절차를 취할 수 있다. 세 번째 입법적인 접근법은 정부기관은 스패머를 처벌하거나 혹은 서비스 제공업체와 개인들의 행동이 보다 효과를 발휘할 수 있도록 하는 행동 양식에 따라 스패머들이 행동하도록 규제할 수 있는 입법을 추진할 수 있다.[6][7]

마지막 세 번째는 사실상 인터넷이 지구상 모든 국가에 걸쳐 있고 사용되기 때문에 개별 국가의 법률이 실질적으로 적용되기는 불가능한 현실이다.

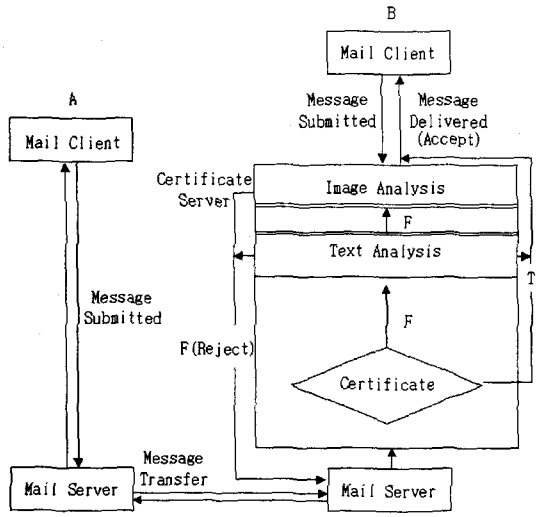
이 논문에서는 첫번째와 두 번째 방식의 기능을 채택하였다.

3. 구현 모델

3.1 메일분석시스템

이 시스템에서는 기존의 메일서버와 수신자의 클라이언트 사이에 또 하나의 인증된 메일만을 수신할 수 있게 하는 인증서버를 구축하여 처리한다. 인증서버에서는 인증기능, 텍스트 분석 그리고 이미지를 차단할 수 있는 이미지 분석이 있다. 우선적으로 인증기능에는 수신자측 메일서버에서 들어온 메일이 인증된 메일인지 여부를 판단하고, 다음 텍스트의 문자열 분석에 있어서는 뒤에 자세히 다루겠지만, 불량단어리스트를 가지고 비교 분석하며, 이미지분석에서는 이미지나 동영상에 대한 유무의 여부를 판단하여 사용자가 볼 것인지 안볼 것인지를 결정할 수 있게 하였다.

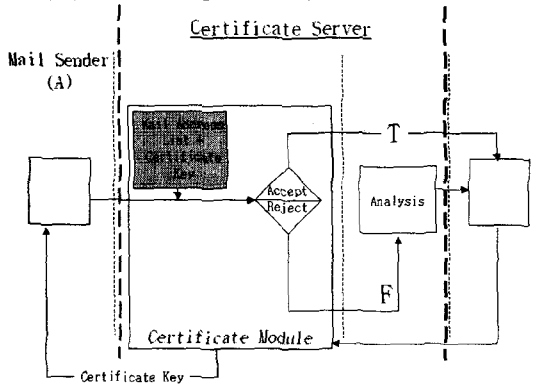
이 시스템에서는 우선적으로 인증결과에 의해서 인증된 메일은 수락(Accept)되어 수신자로 하여금 보여지게 되며 인증되지 않은 메일일 경우 일단 문자열을 분석하는 과정을 거치게 된다. 그 과정을 거치더라도, 100%의 문자열 분석에서의 성인광고 메일 차단을 보장 받을 수 없기 때문에 문자열 분석에서도 검출되지 않을 가능성도 있기 때문에 이미지 분석 단계를 거치게 하였다.



(그림 2) 메일 분석 시스템

3.2 신뢰성 인증에 대한 절차

전자우편은 일차적으로 Mail Receiver(수신자 B)로부터 인증을 받아야 신뢰성 있는 메일전송을 전달 받을 수 있다. 이 시스템에서의 신뢰성 인증 절차의 작동원리는 아래 그림 3 과 같은 구조이다.



(그림 3) 인증 절차

Certificate 모듈은 전자메일의 전자우편 주소가 Accept list 에 존재하는지 여부와 수신된 메일이 인증키(Certificate Key)를 사용해서 전송된 메일인지 여부를 검사한다. 이들 두 가지 검사 중 하나 이상이 만족되면 True 가 되어 메일은 Mail Receiver(수신자 B)로 전송되고, False 가 되면 문자열 분석과 이미지 분석을 한다. 나중에 성인광고 메일이 아닌 경우에는 Mail Receiver(수신자 B)가 그 메일의 송신자에게 인증키를 줄 것인 여부를 판단해서 인증키를 할당할 수 있도록 했다. 이런 경우 그런 송신자가 보내는 메일들은 복잡한 분석 과정을 거치지 않고도 성인광고 메일이 아닌 것으로 처리되어 수신자에게 전달되어서 메일 시스템의 효율성을 높이도록 했다.

3.3 텍스트 분석

일반적으로 스팸 메일 뿐만이 아니라 웹상에서 유해사이트를 차단하려면 텍스트의 문자열 분석이 가장 중요하다. 이 시스템에서는 먼저 성인광고 메일에 주로 쓰여지는 단어를 비교하여 성인광고 메일인가를 분석하게 되었다. 텍스트의 내용을 일일이 다 검색하는 것은 전체적인 시스템의 효율성이 떨어짐에 따라서 핵심적인 단어들의 전처리 과정을 거쳐 불필요한 조사들 "은", "는", "이", "가", 등은 제거함으로써 문장의 핵심 단어들만을 가지고 분석하게 하고, 먼저 앞에서 조사된 것과 같이 메일제목에 "(XX)"라는 단어를 검사하여 없다면 메일 내용을 분석하게 된다.

메일 텍스트의 문자열 분석에는 핵심 단어들에 대해서만 기존 불량단어 데이터베이스와 비교하여 검색하게 된다. 메일 텍스트의 핵심 단어를 추출하는 알고리즘을 보면 아래와 같은 구조이다.

```
int findWord(char *subject)
{
char *token;
token = strtok("xxx는 xxx입니다.", ",WtWn"); //메일의 내용
while(token != NULL){
if token=rmSuf(token); //조사 제거
stlen=stlen(token);
else {
((rmUnuse(token)) == FALSE) //불용어 제거
token= strtok(NULL,seps);
if(stlen<4)
token=strtok(NULL,seps);
else
{
work[i]=token;
token = strtok(NULL,seps)
i++;
                )
            }
        }
return i;
}
```

위의 처리 결과를 보면 "xxx", "xxx"만 남게 된다. 이 결과를 가지고서 불량단어 데이타베이스와 비교, 분석하여 성인광고 메일임이 판단되면 걸러지게 했다.

3.4 이미지 분석

스팸 메일에서의 성인광고 메일에서 텍스트뿐만 아니라 여러 가지 파일형태가 포함되어 전송되는데 특히 이미지파일이 첨부되어 전송되는 경우가 가장 많고 그 다음으로는 동영상파일로 전송되는 경우가 많다. 여기서 우리는 사운드에 대해서도 고려할 수 있는데 실질적인 성인광고 메일 중에서 사운드가 같이 첨부되는 경우는 드물기 때문에 생략하기로 했다.

우리의 노력에도 불구하고, 이미지 분석을 위한 좋은 알고리즘을 고안하기는 힘들었다. 이미지 분석을 위해서는 일반적으로 칼라(color), 모양(shape), 텍스처(texture)을 사용된다. 하지만 우리의 실험에 의하면, 이들 방법은 성인광고 메일의 여부를 판단하는 도구

로 사용하는 데는 한계가 있었다. 성인물 이미지에 주로 사람의 피부색이 많이 등장하는 점을 감안하면 칼라 히스토그램을 사용할 수 있을 것이다. 하지만 등장인물의 인종, 나이 및 장면에 따라서 피부색이 무엇인지를 미리 정하기는 힘들다. 그런 이유로, 본 시스템에서는 사용자가 성인물 여부를 판단하는 기준이 되는 피부색 칼라(들)와 퍼센티지(전체 영상에서 차지하는 비중)을 정할 수 있도록 했다.

본 시스템은 URL 을 가지고서 분석하는 기존 방법도 제공한다. 예를 들면, URL 에서 jpg 와 같이 첨부된 주소를 분석하여 http://www.xxx.xxx/©©©©©©/adult.jpg 와 같은 단어와 함께 구성된 URL 을 가지면 성인광고 메일이라고 판단하고 이미지나 동영상을 차단하여 수신자의 의사에 의해서 결정할 수 있게 하였다.

참고로, 동영상에서도 위와 유사한 방법을 사용한다.

3.5 차단 제어

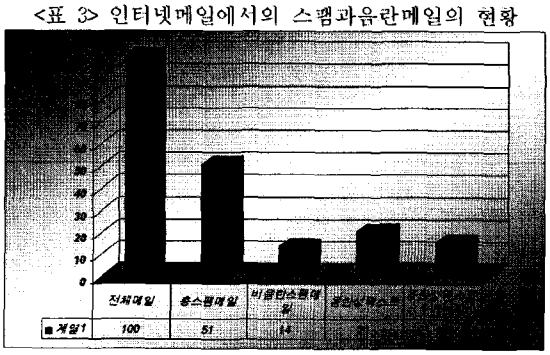
메일 시스템에서의 차단 제어는 세 가지로 구분된다. 낮음, 중간, 높음으로 구분하는데 낮음에서는 텍스트 분석과 수행하고, 중간에서는 텍스트 분석과 이미지 분석만 수행하고, 높음에서 송신자의 메일주소와 인증키 사용여부를 검사한다. 특히 낮음에서는 성인광고 메일 여부를 판단하는 기준이 되는 불량단어의 최소 수 (예를 들면, 3 개)을 사용자가 선택할 수 있도록 한다.

만약 성인광고 메일로 판단되면, 그런 메일은 거부되어서 송신자에게 되돌아 간다. 또한 이때 경고성 메시지를 보낼 것인지 여부를 사용자는 선택할 수 있다. 또한, 성인광고 메일로 최종적으로 판단되지 않더라도 사용자는 이미지를 보는 방법을 선택할 수 있다. 예를 들면, 사용자는 이미지를 축소해서 보거나 이미지의 내용을 감추도록 선택할 수 있다.

4. 실험 및 결과

본인의 메일 계정으로 온 200 개의 메일에 대해서 텍스트의 문자열 분석만 수행한 결과, 51%가 스팸 메일이었으며 그 중에서 "광고"라는 문구가 메일제목에서 포함된 것이 63%정도 차지하였다. 전체 스팸메일 중에서 성인광고 메일이 50%이상을 차지하고 있었다.

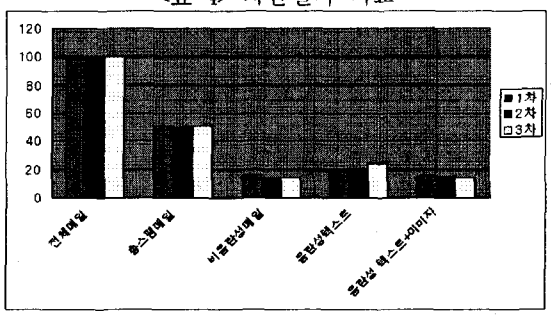
<표 3> 인터넷메일에서의 스팸과음란메일의 현황



이 실험은 본인의 인터넷메일에 전송된 것을 기준으로 해서 실험하게 되었다. 분포를 보면 <표 3>과 같다. 총 100 개가 전송된 메일 중에서 51 개가 스팸 메일이고 여기에서 일반적인 비 성인성 스팸 메일이 14 개, 성인광고 메일이 총 37 개이고 여기서 텍스트만 가진 성인광고 메일은 21 개이고 텍스트와 이미지가 같이 두 종류 이상으로 구성된 메일이 16 개였다.

<표 4>는 이 시스템에서의 스팸 메일 차단과 그 중에서 비 성인광고 메일과 성인광고 메일 그리고 성인광고 메일에서의 텍스트만 가진 성인광고 메일과 텍스트와 이미지와 같이 두 가지 이상의 포맷으로 전송된 메일을 3 차에 걸쳐서 분석하였고, 95%이상의 성공률을 볼 수 있었다.

<표 4> 차단결과 비교



5. 결론 및 향후 과제

본 연구에서는 강력한 메일 보안을 위한 PGP 의 기능을 이용해서 구현하였다. 성인광고메일에 대한 수신으로 가정이나 공공장소에서 당황스런 상황을 만들지 않게 하는 것이 목적이었다. 이 시스템으로 인해 좀더 성인광고 메일을 분석에서 거부할 수 있도록 하였다.

향후 우리는 문자열분석에서의 메일 제목과 내용의 핵심단어 추출에 있어서, 신뢰성을 높일 수 있도록 자연어처리를 통한 분석을 수행하고 또한 여러 형태의 이미지 포맷을 분석할 수 있는 추가하고자 한다.

참고문헌

[1] 박창섭, "암호이론과 보안" 대영사, 1999 [2] 박현동, 류재철, 임채호, 변옥환, "전자우편 보안 - PGP" 학회지, 제 5 제 4 ,1995

[2] 박동욱, 박재희, 김진상, 김일민, "PGP 방식을 이용한 웹 기반 전자우편 보안 시스템" 한국정보처리학회 논문지, 제 8-C 권 제 1 호, 2001

[3] H.X.Mel & Doris Baker, "Cryptography Decrypted" Addison Wesley

[4] http://www.superuser.co.kr/open_lecture/e_mail

[5] 정윤종, "전자우편 보안 - PGP 활용" http://www.certcc.or.kr/concert/cs9803/present/tf02/index.htm

[6] 임채호, "전자우편 보안 - 프라이버시 대책" http://www.privacy.or.kr

[7] David Wood, "Programming Internet Email" O'Reilly & Association, Inc.