PGP를 이용한 스팸메일 차단시스템의 개발

최홍식[0], 김중환
한국외국어대학교 컴퓨터공학과
hufs_sniper@hotmail.com, jhkim@hufs,ac,kr

Development of A Spam-Mail Blocking System Using PGP

Hong-Sik Choi °, Joong-Hwan Kim
Dept. of Computer Science & Engineering, Hankuk University of Foreign Studies

요 약

전자우편(E-Mail)은 아주 편리한 통신수단이지만, 무분별한 광고성 스팸메일(Spam-Mail)의 침입과 무단으로 타인의 메일을 가로채거나 변조할 수 있기 때문에 메일의 신뢰성이 문제가 되고 있다. 본 연구에서는 이와 같은 문제를 해결하기 위하여 메일의 제목과 내용의 문자열을 분석하여 자동으로 스팸메일을 구분할 뿐만 아니라 보안도구인 PGP(Pretty Good Privacy)를 이용하여 메일을 암호화하고 인증하여 근본적으로 스팸메일을 차단하는 시스템을 개발한다.

1. 서 론

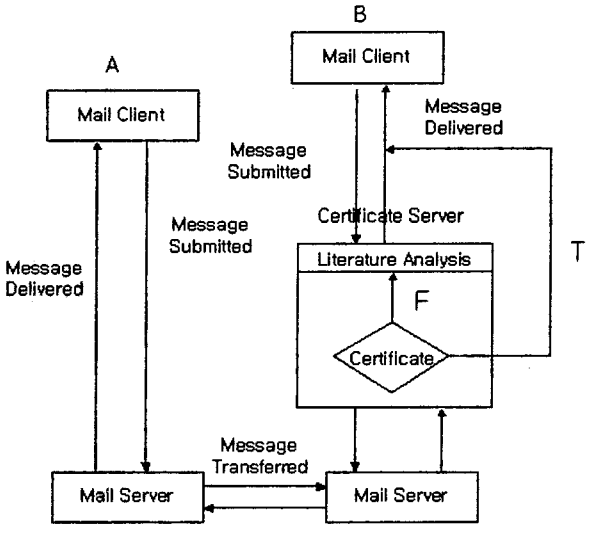
전자우편(E-Mail) 서비스는 매우 편리한 통신수단이지만 무분별한 스팸메일이 침입할 수 있고, 봉투에 넣어져 보내지는 일반적인 편지와는 달리 엽서처럼 내용까지도 그대로 보이는 구조를 지니고 있기 때문에 중간에서 다른 사람이 얼마든지 가로채고 또 변조할 수가 있는 단점이 있다. 따라서 전자우편의 신뢰성을 향상시키기 위해서는 메일이 출발지로부터 인터넷 기반의 네트웍 환경에서 무수히 많은 호스트들을 거쳐 목적지까지 도달하는 동안 보안성이 유지되고, 수신자가 신뢰성이 인증된 메일만 구분하여 수신하는 방안이 필요하다. 그리고 신뢰성이 인증되지 않은 메일은 자동으로 구분하여 별도의 인증 절차에 의해 처리되어야 한다. 지금까지 전자우편의 신뢰성 향상은 주로 보안도구인 PGP (Preety Good Privacy)와 PEM (Privacy Enhanced Mail)을 이용하여 메일의 내용을 암호화하는 방법으로 연구되고 있다. 특히 PGP는 특정한 키가 있어야 메일의 내용을 볼 수 있도록 되어 있기 때문에 기밀성, 인증, 전자서명, 압축 등의 기능을 지원하는 편리한 보안도구이다.[1,2] 또한 스팸메일의 차단은 수신거부와 필터링 기능을 활용하는 방법, 강력한 경고문을 계속 발송하는 방법, 스팸메일 전담용 메일을 만드는 방법 등이 많이 이용되고 있다.

본 연구에서는 기능적인 면과 공개된 기술로서 쉽게 이용 가능하다는 이점이 있는 PGP를 이용한 인증서버를 구축하여 발신자와 수신자 사이에 규정된 방법으로 신뢰성이 인증된 메일만 수신하고, 인증되지 않은 메일은 불량 단어리스트에 의해 메일의 제목 및 내용의 문자열을 분석하여 자동으로 스팸메일을 구분하여 차단한다. 또한 개발된 시스템에서는 메일의 내용을 암호화하기 때문에 메일의 보안성도 향상시킨다.

2. 스팸메일 차단시스템

2.1 스팸메일 차단 처리구조

본 연구에서 개발된 시스템은 스팸메일을 차단하기 위하여 인증 처리 과정과 인증되지 않은 메일의 문자열을 분석하는 과정으로 이루어져 있다. 문자열 분석 시에는 사용자가 직접 필터링에 넣을 단어들을 입력하지 않아도 불량 단어리스트에 의해 필터링을 하고, 인증 과정은 [그림 1]과 같이 기존의 메일 서버와 사용자 각각의 계정인 클라이언트 사이에 또 하나의 인증된 메일만을 수신할 수 있게 하는 서버를 구축하여 처리한다.



[그림 1] 스팸메일 차단 처리구조

인증서버에는 인증 기능과 문자열 분석기능이 있다. [그림 1]에서 송신인 메일 클라이언트 A가 수신인 메일 클라이언트 B에게 메일을 보냈을 경우 수신인 B가 송신인 A에게 첫 번째 메일을 받은 후 그것이 스팸메일이 아니라고 했을 때 인증키를 주어서 다음부터 수신인 B가 송신인 A에게 메일을 보낼 경우는 인증서버에서 문자열 분석의 단계는 거치지 않고 바로 송신인 A에게 갈 수 있게 한다. 그리고 처음 메일을 보낼 경우는 문자열분석을 거친 후 송신인 A에게 전달이 되기 때문에 처음에 오는 메일이라도 스팸메일이 오는 것을 막을 수 있다. 문자열 분석은 메일의 제목 및 내용을 불량 단어리스트에 의해 자동 필터링 하는 방법을 사용한다.

2.2 문자열 분석 과정

스팸메일 차단시스템의 핵심 기능 중 하나인 문자열 분석은 불량 단어리스트에 의한 필터링으로 이루어진다. 본인의 메일 계정에 들어오는 200개의 메일을 분석한 결과 65% 정도인 126개가 스팸메일이었다. 그 중에서 "광고" 라는 문구가 메일 제목에 포함 된 스팸메일은 전체에 63% 정도를 차지하였고, 나머지 37%는 명시하지 않았다. 따라서 개발된 시스템에서는 먼저 메일 제목을 검색해서 "광고"라는 문구가 있는지 검사하고 없다면 메일 내용을 검토해서 미리 작성되어진 불량 단어리스트와 비교하여 스팸메일인가 아닌가를 구분하게 된다. 그러나 전체 메일내용을 일일이 검색하면 처리시간이 길어져 효율성이 떨어지게 된다. 이와 같은 단점을 해결하기 위해 본 연구에서는 스팸메일을 구분하는데 불필요한 조사와 불용어를 제거하고 핵심 단어만을 추출하는 전처리 과정을 두었다. 불필요한 조사는 "은","는","이","가" 등이 있고 불용어는 예를 들면 "이다", "했습니다", "대한", "상관" 등 제거를 해도 문장을 이해하는데 크게 어려움이 없는 것들이다. 이것들을 제거함으로써 핵심 단어들만이 남게 되며, 이 핵심문구를 불량 단어리스트에 있는 단어들과 비교하여 검색한다.

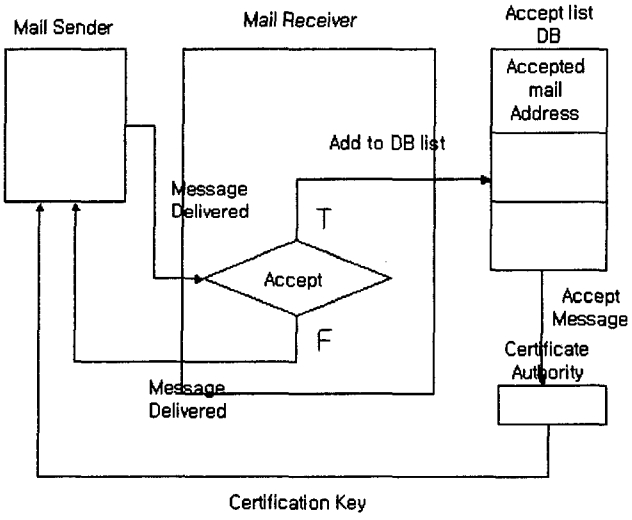
메일 내용의 핵심 단어들을 추출하는 것을 간단한 예로 살펴보면 다음과 같다.

```
 int findWord(char *subject)
{
char *token;
token = strtok("이 상품은 밤에 사용하는 것입니다.",
",₩t₩n"); //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;
}
```

위에 처리 결과를 보면 "상품", "밤", "사용" 만 남게 된다. 그러면 이 결과에서는 불량 단어리스트에 상품, 밤이 들어가 있어 스팸 메일로 구분되고 송신자에게 리턴 메일을 보낸다.

2.3 신뢰성 인증 과정

메일의 신뢰성 인증 과정은 [그림 2]와 같으며 송신자가 처음 메일을 보낸 메일을 받았을 경우에 스팸메일이 아니라면 다음 메일부터 스팸메일이 아님을 인증하는 방식이다. 즉, 수신인은 송신인이 처음 보낸 메일이 스팸메일이 아니고 계속 받고 싶을 때는 송신인의 메일 주소를 Accept list DataBase에 저장하고, 그 결과를 인증기에 통보하고 인증기는 송신인에게 인증키를 주게 된다. 송신인은 인증키를 수신하게 되어 다음부터는 동일한 수신인에게 메일을 보낼 때는 아무 제약 없이 바로 수신인이 신뢰하는 메일을 볼 수 있게 한다.



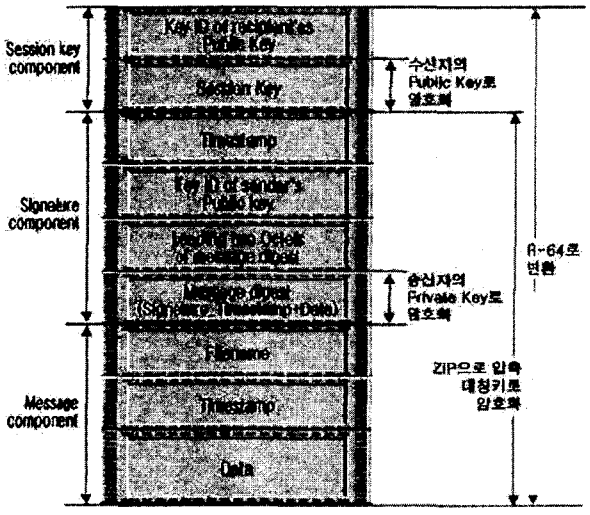
[그림 2] 신뢰성 인증과정

인증기가 생성하는 인증키는 개인키를 이용하여 수행한

다. 파일을 개인키로 인증하여 인증 파일을 넘겨준다. 인증의 확인은 송신자의 인증이라는 것을 모두가 알 수 있는 송신자의 공개키를 적용함으로써 알 수 있다. 송신자의 메일 주소로 송신자의 공개키를 얻어서 송신자의 개인키로 인증된 것을 풀고 인증을 확인한다. 인증이 틀릴 경우는 송신자가 인증된 메일 계정에서 보내지 않은 경우이거나 처음 보내는 송신자일 경우이다. 또한 한번 인증된 메일 계정은 다음 단계인 문자열 분석을 거치지 않고 바로 수신자에게 전달이 된다.

2.4 구현

개발된 시스템은 인증과 메일의 안전을 위해서 PGP를 사용하였다. PGP의 구현에 있어서 자료를 암호화할 때는 비밀키 방식인 IDEA(International Data Encryption Algorithm), Session 키 암호화할 때는 공개키 방식인 RSA(Rivest Shamir Adleman)를 사용한다.[2] 먼저 키 생성을 하게 되는데, 공개키와 개인키로 나누어 생성하고 그 키 값을 저장하게 된다. 이것이 하나의 인증키가 되는 것이다. 암호화는 수신인의 공개키를 이용하게 되는데, 이때 암호화하는 방식은 상대편 메일 주소를 이용해서 공개키 파일 리스트에서 상대편 공개키를 찾은 후 두 번째 인자의 파일을 찾은 상대편 공개키로 암호화한 후 첫 번째 인자에 암호화한 파일명을 넘겨주는 방식이다. 암호화된 파일은 수신자의 개인키를 이용하여 복호화를 하는데, 암호화된 파일을 DES(Deta Encryption Standard) 알고리즘으로 복호화하고, 암호화된 세션키를 개인키로 복호화해서 넘겨준다.[3] [그림 3]은 PGP의 메시지구조를 나타내고 있다.



[그림 3] PGP 메시지 구조

Message-component의 Time-stamp는 Data가 만들어진 시각이고 Signature component의 Time-stamp 는 signature가 생성된 시각을 나타내고 있다. Leading two octets of message digest는 바로 밑의 암호화된 message digest를 모두 비교하면 시간이 오래 걸리므로 앞의 16bit만 비교하기 위해서 필요한 부분이다. 오른쪽에는 암호화되는 부분과 그때 사용되는 암호 알고리즘을 표시해 주고 있다.[4]

3. 결론

본 연구에서는 암호화 방법인 PGP를 이용한 인정서버가 인증된 메일만 수신하고, 인증되지 않은 메일은 문자열 분석을 하여 자동으로 스팸메일을 구분하여 차단하는 시스템을 개발했다. PGP를 사용해서 생성된 파일들은 연속적인 8-bit의 흐름이지만 대부분의 전자우편 시스템은 ASCII 문자만을 인식한다. 따라서 PGP 에서는 Radix-64 변환을 통해서 3개의 8-bit를 4개의 ASCII문자로 변화 시켜서 기존의 전자우편 시스템과 호환성 문제를 해결했다. 또한 효율적인 문자열 분석을 위해 메일의 제목과 내용의 핵심 단어를 추출하여 불량 단어리스트로 검색하는 방법을 사용했다. 향후에 불용어와 조사처리에 대한 부분의 구조를 일부 수정하여 메일의 핵심 단어가 아닌 문장으로 데이터베이스에 저장을 한 후 수신자가 메일의 내용을 알고 싶을 때 열람할 수 있는 기능을 추가하면 더욱 편리한 시스템이 되리라고 본다.

5. 참고 문헌

[1]남궁 재창, "PGP의 개념과 활용" ,PLUS (POSTECH Laboratory for UNIX Security) Security+ for UNIX II, 1997
[2]정윤종, "전자우편 보안 - PGP 활용" http://www.certcc.or.kr/concert/cs9803/present/tf02/index.htm
[3]프로그램의 세계 "RSA 알고리즘", 98년 5월
[4]Dream Security, "전자우편보안", http://maxim.dreamsecurity.co.kr/0111_55PGP.asp
[5]이상엽 , "Visual C++ Programming Bible Ver 6.X" 영진 출판사