[인터넷보안] 스팸 메일의 억제를 위한 표준 메커니즘

본 문서는 최근 미국 텍사스 달라스에서 열린 65차 IETF에서 제공된 ‘The Internet Protocol Journal’에서 특집 주제로 다루어진 스팸(Spam) 메일에 관한 내용을 다루고자 한다. 참고로, IETF에서 전자 메일을 다루는 워킹 그룹은 EAI(Email Address Internationalization) WG, LEMONADE(Enhanced to Internet Email to Support Diverse Service Environments) WG, SIEVE(Sieve Mail Filtering Language) WG 등이 있다.

스팸 메일의 문제점과 유형

아마도 최근 들어 인터넷 사용에서 직면한 가장 큰 문제점은 원하지 않는 전자 메일, 즉 스팸 메일의 꾸준한 증가일 것이다. 스팸 메일은 요구하지 않은 벌크메일(Unsolicited Bulk Email), 요구하지 않은 상업메일(Unsolicited Commercial Email)의 형태로 컴퓨터 통신망에서 무차별로 살포되어 이를 원치 않는 사람이 읽거나 처리하는 데 많은 시간과 비용을 낭비하게 된다.

스팸 메일의 유형은 직접 스팸과 중계 스팸 두 가지로 구분하고 있는데, 직접 스팸은 스패머(Spammer:스팸 메일을 보내는 사람)가 직접 자신이 이용하는 ISP의 메일 서버를 통해 불특정 다수의 사용자에게 메일을 직접 보내는 것이다. 중계 스팸은 이보다 한 단계 복잡한 방법으로 스패머가 자신이 이용하는 메일 서버 대신 임의의 다른 ISP나 기업의 메일 서버를 중계 서버로 이용해 마치 중계 서버의 사용자가 불특정 다수에게 광고성 메일을 보내는 것처럼 위장하는 방식이다.

물론, 여러 가지 스팸 억제 방식이 적용되어 스팸을 막고 있지만, 실제 스팸 메일의 양을 줄이지는 못하고 있는 실정이다. 다음에서 스팸 억제를 위한 필터링 방식에 대해서 살펴보기로 한다.

스팸 메일의 억제를 위한 필터링 표준 규격

스팸 메일의 억제를 위한 가장 일반적인 방식은 전자 메일을 조건적으로 허가하는 국소적인 방식인 “필터링(Filtering)” 방식이다 [RFC3028]. 일반적으로 필터링은 수신자 내부에서 사용된다. 그러나, 이러한 필터링은 명백하게 MSA(Message Submission Agent)[RFC2476]를 포함하는 경로를 따라 어느 곳에든 위치할 것이다. 수신측 네트워크 내부에서의 필터링은 스팸 트래픽을 줄일 수 없고, 그 밖에서 줄일 수 있다. 필터링은 의심되는 메시지들을 다루기 위해 다음과 같이 몇 가지 선택적인 방식들을 갖는다.

- 메시지에 특별한 주석을 첨가한다.
- 의심되는 메시지를 특정한 저장물(storage)로 바꾼다.

- Handling Notification [RFC2821] 주소나 클라이언트 SMTP에게 의심되는 메시지를 되돌려 보낸다.
- 간단하게 의심되는 메시지를 지운다.
- SMTP 전송의 속도를 조절하기 위해 "Traffic Shaping"방식을 사용하여 의심되는 메시지를 천천히 수신한다.

"필터링을 적용해야 할 기준은 무엇인가?"는 어려운 질문이다. 아울러, 이에 대한 어려운 대답은 "많다" 이다. 넓고, 변화무쌍하고, 다양한 결정 기준을 지원하기 위한 필요성은 필터링 엔진으로 하여금 스팸 발견을 위한 확장성이 있는 플랫폼으로의 진화를 초래했다. 필터링 알고리즘의 혼용과 복잡성이 더욱더 세련되어 짐에 따라, 수반하는 오버헤드는 실질적으로 더 크게 증가해왔다.

비록 각각은 복잡하지만, 필터링 기술을 세 개의 기본적인 기준으로 나누는 것이 편하다.

- 중복되는 벌크 메일을 발견하기 위한 단어수나 내용 해싱에 대한 Baysian 통계 추적과 같은 **내용 분석** (Content Analysis)
- 허용 (whitelist) 또는 불허 (black list)이냐에 대한 **책임 있는 대리인 평가 (Responsible Agent Assessment)**
- 같은 저자의 주소 혹은 IP 호스트 주소로부터 발생된 메시지의 비율과 같은 **트래픽 분석** (Traffic Analysis)

이 중에서 내용 분석은 항상 부부적으로 성공한다. 단어의 표준을 설정하는 것은 보통 통계적이고 테스트 메시지의 데이타베이스에 의존한다. 하지만, 스패머는 항상 현재의 분석 기술을 회피하기 위한 기술을 개발하고 있다. 더 나아가, 동일한 전자 메일 서비스에 대해 서로 다른 수신자들은 수락 가능한 내용에 대해서 다른 통계적인 양상을 가질 수 있다. 이는 서비스 제공자에 의한 필터링이 완벽할 수 없다는 것을 의미한다. 따라서, 이러한 분석 도구들이 계속 발전되더라도 효과적이지 못하고 장기적인 도구가 될 수 없으며 근본적으로 스팸을 줄일 수 있는 효과를 거의 갖지 못한다. 이 때문에 이러한 분석 도구들은 두 개의 내재적인 결점을 갖는다. 첫번째는 합법적인 메일이 스팸처럼 잘못 분류될 수 있다는 것이다. 예를 들어, 중요한 업무용 메일이 대량 선전 메일로 분류되는 대신에 배달되지 않는 것이다. 아마도 가장 심각한 문제는 스패머(Spammer)가 유명하고 합법적인 사업으로부터 간주되는 메일을 보낼 때이다. 이것을 'phishing(금융기관 등으로부터 개인정보를 불법적으로 알아내 이를 이용하는 사기수법)' 이라 부르며, 이로 인해서 중요한 메일의 합법적인 포스팅이 이루어지지 않는 문제가 발생할 수 있다. 두번째 문제점은 항상 서로간에 자신들의 기술들을 적용해야만 하고, 보다 많은 재원을 소비해야만 하고, 그러나 어느 누구도 절대 승리하지 못하는 스패머(Spammer)와 반스패머(Anti-spammer) 사이의 끝없는 싸움이라는 것이다.

스팸 메일 문제 해결을 위한 표준화의 역할

스팸 메일 문제를 해결하기 위해서 각종 유용한 도구를 사용하는 것도 중요하지만 이와 관련된 보다 다양한 문제 정의 및 솔루션을 제시하는 표준 규격을 만들고 이를 적용하는 노력이 강력히 요구된다. 예를 들어, 스팸 관련 사건의 레포팅, 특별한 형태를 갖는 스팸의 특성에 대한 정의, 스팸 억제 데이터의 전송 등에 대한 문제들이 표준화 이슈 사항이 될 것이다. 이러한 표준 이슈의 한 예로 IETF에서 논의되고 있는 인터넷 문서 “An Extensible Format for Email Feedback Reports”를 통해서 확인할 수 있다. 또한, 스팸 문제를 해결하기 위해서는 서로 다른 언어를 사용하는 네트워크 관리자들간의 상호 운영 서비스를 활용한 국제적인 운영에 대한 협력이 요구된다.

김평수 (한국산업기술대 전자공학과 교수, pskim@kpu.ac.kr)