스팸메일 방지를 위한 MDA의 필터링방법 개선방안

박은옥*, 김영현, 최은정, 유주영, 김미애, 박유미, 김윤정, 김명주

서울여자대학교 대학원 컴퓨터학과

An improvement of MDA(Mail Delivery Agent) Filtering mothod for prevention of spam mail

Eun-Ok Park*, Young-Hyun Kim, Eun-Jung Choi, Ju-Young Yoo,

Mi-Ae Kim, Yu-Mi Park, Yoon-Jeong Kim, Myuhng-Joo Kim

Department of Computer, Seoul Women's University

요 약

인터넷 이용자가 증가함에 따라 전자메일 사용자도 증가하고 있다. 전자메일 사용으로 통신상의 비용 및 시간이 절약되는 장점이 있지만 소수의 유저들이 상업적 목적으로 많은 유저에게 원하지 않은 메일(스팸메일)을 보냄으로써 물질적, 정신적 피해를 입히고 있다. 따라서 스팸 메일을 방지하기 위한 여러 기법들이 제안 되었다. 본 논문에서는 스팸 메일 문제를 해결하기 위해 먼저 전자메일 시스템에 대한 구조를 살펴보고 MTA, MDA를 이용하는 스팸 메일 필터링 도구들을 비교 분석한 연구결과를 제시한다. 그리고 탐지 성능을 개선할 수 있는 새로운 방안을 제시한다. 제안 방법은 공개 배포용 MDA인 procmail에 기반한 것으로, 규칙(rule)을 매칭(matching)시키는 시간을 줄이는 것이다.

I. 서론

인터넷의 이용자는 해마다 증가추세를 보이고 있다. 2003년 6월 우리나라 6세이상 국민의 64.1%(2861만명)가 인터넷을 이용하고 있으며, 이 중 84.6%가 전자메일을 보유하고 있다[1]. 이러한 인프라의 구축으로 전자메일이 간편하고, 빠르기 때문에 사용자가 급속도로 많아졌다. 그러나 상업적으로 전자메일을 이용하려는 소수의 사용자로 인해 원하지 않은 상업적 메일 즉 스팸 메일을 수신하는 많은 다수의 사용자가 물질적·정신적 피해를 입고 있다.

IT 시장조사 전문기업인 IDC의 조사 자료에 의하면, 전 세계적으로 하루 평균 약 54억 통, 연간으로는 총 1조9천6백억 통의 스팸이 유통되고 있으며, 페리스 리서치가 올해 초 발표한 조사 자료에 따르면, 지난해 스팸 메일로 인한 미국의 피해는 89억 달러(약 10조8400억원), 유럽은 25억 달러(약 3조400억원)로 추산하고 있다. 우리나라의 경우에도 한국정보보호진흥원의 조사에 의하면, 1인당 하루 스팸 메일 수신량이 2001년 약 5통에서 2002년에는 약 35통으로 7배 이상 증가하였고 2003년에는 약 40통으로 계속해서 증가하고 있으며, 이를 1년간 국내에서 유통되는 총 스팸 메일 양으로 환산하면 대략 3천억 통 정도이며, IDC의 자료 대비 전 세계 스팸메일 유통량의 15% 이상을 차지한다[2].

또한 스팸 메일은 청소년의 성교육에도 부정적 영향을 끼치는 것으로 나타났다. 한국통신문화재단이 전국의 초중고교생 1500명과 학부모 500명 등 2000명을 대상으로 '청소년 스팸메일 실태에 관한 조사'를 실시한 결과, 고등학생이 받는 전체 메일 가운데 49.2%, 중학생 35.8%, 초등학생 8.3%가 음란성 광고 메일인 것으로 조사되었다[3].

우리는 스팸 메일을 필터링 하는 많은 도구 중 전자메일의 헤더와 본문을 대상으로 하여 특정한

정보를 찾아낸 후, 미리 정의된 규칙에 따라 조치를 수행하는 메일 프로세서 프로그램인 Procmail을 이용해 좀더 자세한 기능을 살펴보고 필터링 기능 강화를 위한 추가 기술을 제안하고자 한다.

II 장에서는 전자메일 시스템에 대한 구조 및 기존 필터링 도구에 대해서 알아보고 III 장에서는 Procmail의 필터링 기능에 대해서 명세화 하고 IV 장에서는 이를 기반으로 새로운 제안을 하였다. V 장에서는 제안된 기술을 바탕으로 성능 평가를 하였다.

II. 연구배경

1. 전자메일 시스템의 구조

스팸 메일을 필터링 하는 도구들이 전자메일 시스템에서 어떤 역할을 수행하는지 인지하기 위해 전자 메일 시스템의 기본 구조를 알아보고자 한다.

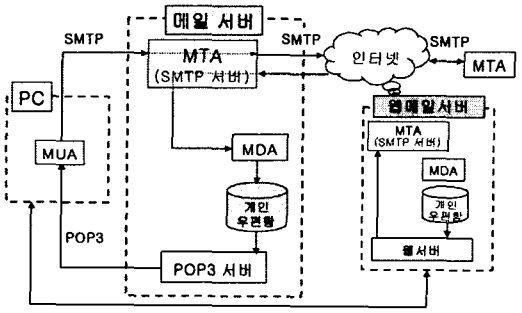

그림 1: 전자메일 시스템 구조

그림 1과 같이 하나의 메일 시스템은 MUA, MTA, MDA 로 구성된다[4]. 각각의 역할은 다음과 같다.

MUA(Mail User Agnet)는 사용자가 메일을 읽거나 작성하거나 보낼 때와 같이 메일을 직접 다룰 때 이용하는 프로그램을 말한다. 이 프로그램의 예로 UNIX Mail 프로그램(/bin/mail)이나 elm, pine, eudora, Netscape Messenger, MS Outlook Express 등이 있다.

MTA(Mail Tranfer Agnet)는 MUA에서 작성되어 전달받은 메일 메시지의 헤더 정보를 읽고 지정된 호스트까지 전송하는 역할을 한다. 예로는 sendmail, MMDF, Zmailer, qmail등이 있다.

MDA(Mail Delivery Agent)는 MTA에 의해 호스트로 전송된 메일을 사용자의 우편함에 배달하는 프로그램을 말한다.

2. 기존 필터링 도구

우리가 제안하는 필터링 방법을 차별화 하기 위해 기존 필터링 도구를 알아보고 각각의 성능을 분석해 보고자 한다.

AMaViS, inflex는 MTA나 MDA가 위치하는 곳 어느 곳에서나 연동할 수 있다. Procmail은 MDA의 역할로 필터링을 한다.

1) AMaViS(A Mail Virus Scan)

AMaViS는 하나 이상의 바이러스 스캐너와 MTA를 결합시키는 스크립트다[5]. AMaViS는 메일이 sendmail 이나 qmail 같은 MTA를 통하여 도착하였을 경우 첨부 파일을 분리하여 바이러스 검출 프로그램으로 검사한다. sendmail(또는 다른 MTA)와 바이러스 스캐닝 유틸리티 사이에서 동작한다.

기본적인 AMaViS의 형태는 sendmail.cf file에서 local delivery agent의 역할을 한다. sendmail은 local에서 메시지를 전달할 때, local delivery agent 대신 AMaViS를 호출한다.

AMaViS에서 지원되는 안티 바이러스 제품

- DrSolomon,
- H+BEDV AntiVir/X
- Sophos Sweep
- Kaspersky Lab AntiViral Toolkit Pro (AVP)
- CyberSoft VFind, Trend Micro FileScanner,
- CAI InoculateIT
- F-Secure Inc. (former DataFellows) F-Secure AV

2) Inflex

Inflex는 메일서버에서 로컬이나 외부로 나가는 전자메일을 검사하여 전자메일에 대한 송수신 정책을 세울 수 있도록 하는 메일 스캐너로, In-Outbound 정책기능을 통해 최근의 바이러스나 인터넷 웜 등의 첨부여부를 조사할 수 있다. Inflex는 sendmail이 sendmail.cf 대신에 inflex.cf 파일을 설정파일로 사용하도록 함으로써 원하는 기능을 제공하게 된다. 결국 /etc/sendmail.cf 설정을 바꾸지 않고도 incoming/outgoing 메일을 검사할 수 있게 된다. 또한 inflex는 백신과 연동하여

사용이 가능하고 관리 및 운영이 용이하다

그러나 첨부파일만 검사할 수 있기 때문에 첨부가 없는 메시지에 대해, Inflex는 다른 활동적인 내용 공격과 쉘 스크립트 공격을 찾을 수 없다.

Inflex는 설치 전에 다음 사항을 반드시 필요로 한다[6][7].

·Sendmail v8.9.x

·F-PROT AntiVirus

·Sophos AntiVirus

·NAI UVScan v4.03

3) Procmail

Procmail은 전자우편물의 헤더와 본문을 대상으로 하여 특정한 정보를 찾아낸 후, 미리 정의된 규칙에 따라 조치를 수행하는 메일 프로세서 프로그램이다.

Procmail은 외부에서 sendmail을 통해 들어오는 메일을 MDA 수준에서 필터링 할 때 주로 사용된다. 수신되는 메일의 헤더정보를 바꿀 수 있고 본문의 내용을 각각의 문자 셋에 따라 코드 변환시킬 수 있는 등 여러 강력한 필터기능을 제공한다.

헤더와 본문 검색에 강력한 필터링 기능을 가지고 있는 procmail은 전자메일을 이용한 공격을 막기 위해 헤더를 검색해서 긴 헤더를 잘라냄으로 버퍼 오버 플로우 공격을 방지하고 공격 가능한 스크립트 태그를 변형시키거나 첨부된 실행파일의 이름을 실행 불가능한 이름으로 변형 시킬 수 있다[8].

procmail은 procmail 자체로 실행되지 않는다. 보통 시스템 관리자가 sendmail.cf에 포함시켜 명시적으로 실행시키거나 사용자가 홈 디렉토리에 .forward 파일을 둠으로써 실행시킨다.

표 1: .forward를 이용한 procmail 사용

"|IFS=' ' && exec /usr/bin/procmail -f- || exit 75 #loginID"

표 1에서 보여주는 것처럼 자신의 계정 홈디렉토리에 .forward 파일을 만들어 procmail을 수행하기 위해 실행파일 위치를 지정해준다.

우리는 위에서 설명한 필터링 도구 중 설치와 운영은 어렵지만 기능과 융통성은 좋은 Procmail을 중점적으로 분석하고 새로운 제안을 하고자 한다.

Ⅲ. Procmail의 필터링 기능 명세

그림 2: 에서는 Promail이 어떻게 동작하고 필터링 하는지에 대한 전체적인 흐름을 보여주고 있다[9].

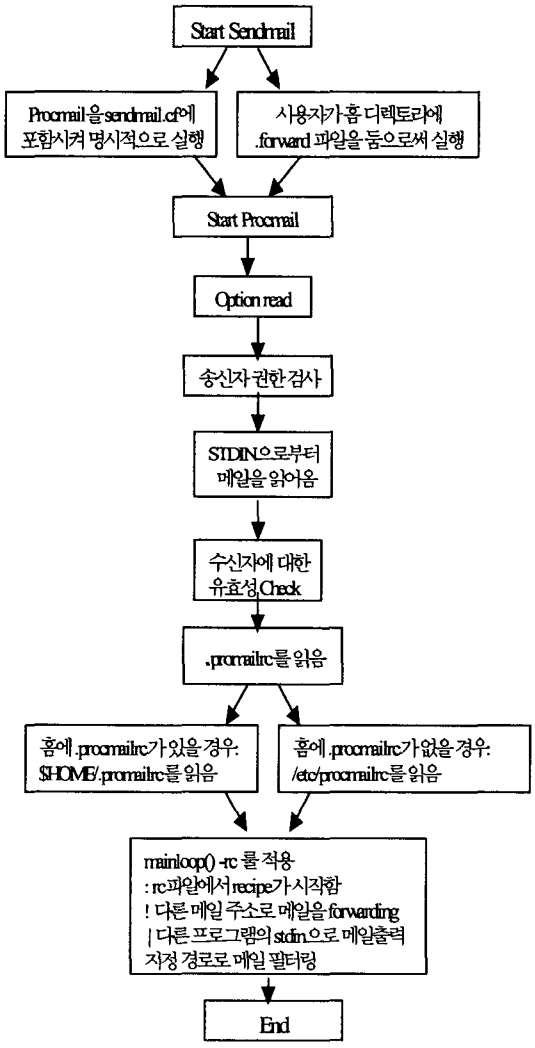

그림 2: procmail 흐름 명세

procmail은 MDA로 sendmail과 같이 연동하며 옵션 사항을 먼저 검사한다. 또한 송신자 및 수신자에 대한 권한 사항을 검사한다. .procmailrc에 있는 규칙을 적용하기 위해 /etc/procmailrc 나 유저의 계정 홈 디렉토리에 있는 procmailrc를 가져와 규칙을 적용 한다. 프락메일의 소스 중

procmail.c에 있는 mainloop()가 procmailrc로부터 recipe를 읽어 들여, 정규 표현식에 해당하는 패턴이 메일에 존재하는지 검사하고, 검사 결과에 따라서, 주어진 action 즉 !, |등을 취한다.

예를 들면 메일을 포워딩하는 !나 들어온 메일을 다른 곳으로 출력할 수 있는 | 등이 동작하도록 수행한다.

Ⅳ. 새로운 제안 방법

procmailrc가 순차적으로 규칙을 매칭 시키므로 마지막에 나온 규칙은 앞의 규칙 수만큼 매칭을 시켜야 하는 부하가 있다. 또한 일반 유저가 procmailrc의 모든 기능을 알고 사용하는 것이 아니기 때문에 일일이 지정하기에는 많은 불편과 시간이 소요된다. 우리는 이러한 점을 착안, 유저가 메일을 확인할 때 스팸 메일로 지정할 것인지에 대한 메시지를 띄우고 유저가 YES라고 하면 해당 메일을 subject, from, 첨부파일 등을 나눠 SPAM_SUBJECT, SPAM_FROM, SPAM_ATTACH 파일에 저장한다. 이 파일들은 스팸메일의 제목과, 수신자, 첨부파일에 대한 목록으로 유저는 procmailrc에 따로 지정하지 않아도 자체적으로 스팸 메일인지 아닌지 분류하고 필터링한다. SPAM_SUBJECT, SPAM_FROM. SPAM_ATTACH에 저장되는 순서는 가장 최근에 유저가 지정하는 것을 우선으로 한다. 또한 DB에 있는(각 파일에 있는) 것이 일정기간 스팸 메일로 매칭 하는 것이 없다면 자동 삭제 시켜 DB의 부하를 줄인다. 그리고 다시 정렬 한다. 이를 그림 3에서 명세화 시켰다.

한 예로 2003년 8월 국내에 유입된 웜인 Win32/Mimail.worm.10784 의 경우 이메일 제목이 don't be late로 메일에 첨부된 READNOW.ZIP 파일을 통해 전파되고 감염되면 대량의 이메일을 발송해 시스템의 과부하를 일으킨다[10]. 이 경우 유저는 처음 이 메일을 받았을 때, 스팸 메일 리스트에 넣을 것인지에 대해 YES 라고 했을 경우, 가장 최근 스팸 메일 리스트 목록에 들어가게 된다. 만약 같은 이런 웜 메일이 스팸 메일처럼 여러 차례 다시 들어 온다면 DB에 있는 최근 목록에서 필터링 함으로써, 네트웍을 통해 빠르게 전파되는 가장 최근의 웜의 피해 줄일 수 있다.

위에서 설명한 것처럼 우리가 제안 하는 것은 기존의 필터링 제품과 다르게 능동적으로 리스트를 만들고, 업데이트 하고 리스트를 정렬함으로써 매칭 시간을 줄인다. 뿐만 아니라 제목, 첨부파일 등을 이용하는 바이러스의 경우는 스팸 메일과 동일하게 적용되어 매칭 될 수 있다.

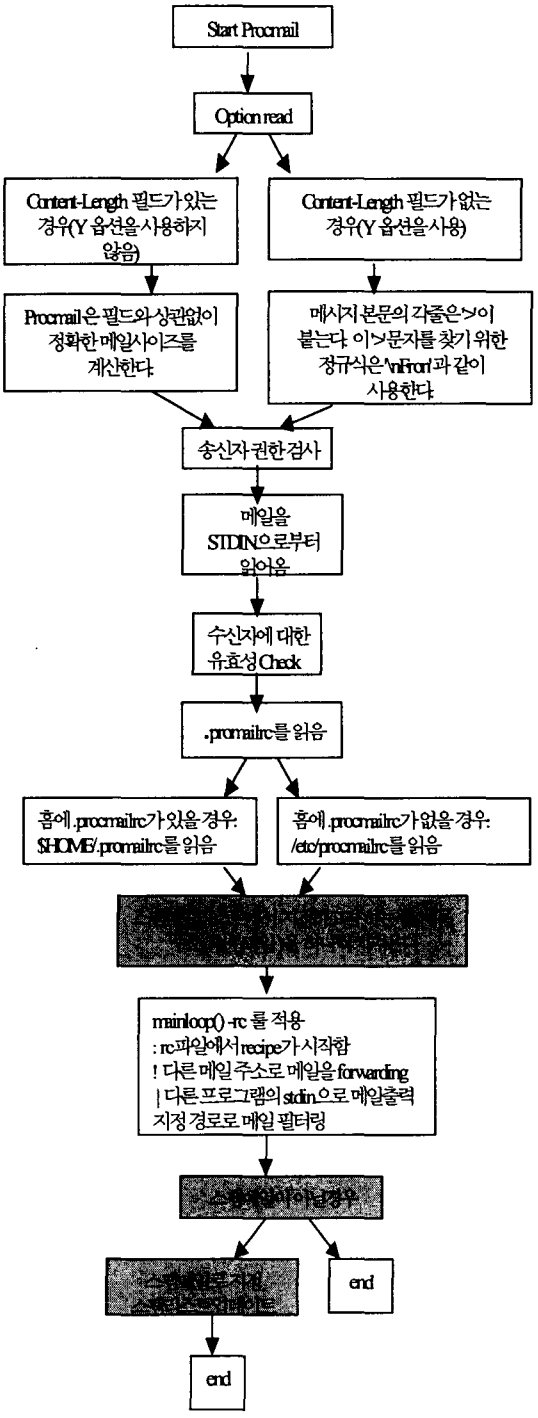

그림 3: 제안하는 procmail 필터링 구조 명세

V. 성능분석

1. 탐지율 분석

1) 실험 방법

실험하고 하는 시스템은 SunOS 5.8 에서 MTA로 sendmail-8.12.9, MDA로 procmail-3.22를 사용했다.

실험 대상은 hotmail을 사용하는 유저 5명으로 했으며, 3일간에 걸쳐 스팸메일 포함 1000통의 메일을 수집했다. 이중 제목이 다음 단어를 포함하는 메일을 필터링 하는 규칙을 세웠다.

*sex, *re, *notice, *pay, *wanted, *money, *service, *adult, *sale

procmailrc 파일은 이 단어들과 제목이 매칭하는지 순차적으로 검색한다. 첫 번째 규칙인, *sex와 매칭했다면 두 번째부터는 검색하지 않는다.

2) 실험 결과

1000 통의 메일 중 규칙을 적용해 267개의 스팸메일을 필터링하였다. 의도한 규칙대로 필터링 된 것이 59개로 나머지 208개인 77.9%가 FP로 나왔다. 그 원인으로는 정해놓은 규칙 중 *re은 [RE] 나 RE:가 독립적인 단어로 이 단어가 포함되는 제목의 메일을 필터링 하는 것이 목적이었지만, read, release, free, correct, great 등 접미사, 접두사 등 re가 들어가는 모든 단어가 필터링 됐다. 반면 FN는 0%였다.

2. 수행시간 분석

수행시간 분석을 위해서는 이론적인 분석을 수행하였다. 만약 1000개의 필터링 규칙을 세웠다면, 가장 마지막 규칙에 매칭 하는 메일일 경우 앞의 999개의 규칙을 다 매칭 시켜야한다. 또한 1000개의 규칙에 적용되지 않는 메일이라도 1000번의 매칭을 해야 하는 결과가 나온다. 이는 규칙이 적을수록, 최신 스팸 메일 관련 규칙을 먼저 매칭 시킬수록 시스템의 부하는 줄어들고, 매칭 시키는 시간도 줄어드는 효과를 가져 오는 것으로 나타났다.

VI. 결론 및 향후 연구과제

본 논문에서는 스팸 메일 필터링을 하기 위한 기존연구 내용 중 MDA로써 Procmail의 필터링 알고리즘을 중점적으로 분석하였다. 이를 통해 보다 효율적인 필터링 성능 향상을 위해 능동적인 매칭이 가능하도록 procmail 기반에서 필터링 기능을 제안하였다. 이에 따른 새로운 방안은 다음과 같다.

필터링 규칙(규칙 목록)을 YES나 NO로 간단히 유저가 지정해 줄 수 있도록 메일 확인 시 메시지를 주며, 유저가 가장 나중에 지정한 것이 가장 최근 스팸 리스트로 자동 업데이트 되어 최근의 규칙부터 매칭 시킨다. 이는 매칭 시간을 줄이고 따로 규칙을 작성해야 하는 번거로움을 제거해 주는 일석이조의 역할을 하고 있다.

이론적 분석결과를 통해 제안된 방법으로 필터링의 성능이 향상된 것을 확인할 수 있다. 실제 수행 시간을 측정하는 기술적 분석을 통한 데이터 획득은 연구·진행 중이며, 지속적으로 또 다른 성능방법에 대한 연구와 확장된 필터링 기능 향상을 위한 연구를 진행할 계획이다.

참고문헌

[1] 한국인터넷정보센터(KRNIC), "정보화실태조사", 2003.6
[2] 한국통신문화재단, "우리집 스팸메일 추방교육운동"
[3] 한국통신문화재단, "청소년 스팸메일 실태에 관한 조사", 2003.06
[4] 윤세안. "악성코드의 유형분석을 통한 메일서버스캐너의 설계 및 구현", 서울여자대학교 석사학위논문, 2002
[5] http://www.amavis.org
[6] http://www.inflex.co.za/mainpage.html
[7] 이현우, "메일 필터링을 통한 e-mail 보안", CERTCC-KR, 2001.03
[8] http://www.procmail.org
[9] 김장복, "A New Architecture the Security the Security of E-mail System", PDPTA, 2003
[10] http://home.ahnlab.com