웹 기반의 메일 통계 분석 시스템 설계 및 개발

이상도*, 김광혁*, 이상영*, 김태형*, 장범환*, 정태명***
*성균관대학교 전기전자 및 컴퓨터공학과
**성균관대학교 전기전자 및 컴퓨터공학부

e-mail:{sdlee, byraven, sylee, thkim, bhchang, tmchung}@rtlab.skku.ac.kr

Design and Implementation of Web-Based Mail Statistics System

Sang-Do Lee*, Kwang-hyuk Kim*, Sang-Young Lee*, Tae-hyung Kim*, Bum-Hwan Jang*, Tai-Myung Chung**
*Dept of Electrical and Computer Engineering, Sungkyunkwan University
**School of Electrical and Computer Engineering, Sungkyunkwan University

요약

인터넷의 급속한 성장으로 인하여 메일을 이용하는 사용자가 급격히 증가하고있다. 따라서 메일 서버를 운영하는 관리자는 메일 분석에 대한 전문적인 로그 분석 도구들을 점차 요구하게 되었다. 현재 시중에 샌드메일 로그 분석에 관한 여러 구현된 상용 제품들이 있으나 단순한 통계 정보만을 제공하거나 특정 제품에만 맞도록 구현되었다. 이 논문에서는 시스템 로그 파일만을 분석하여 메일에 대한 상세한 통계 정보를 제공하는 웹 기반의 메일 분석 도구의 전체 설계 구조 및 구현 결과에 대해서 살펴보고자 한다.

1. 서론

최근 메일을 이용하는 사용자가 급격히 증가하고, 이에 메일 통계를 내는 분석 도구들 또한 많이 등장하고 있다. 이러한 대부분의 메일 분석도구는 샌드메일 서버의 로그 파일을 분석하여 메일 서버의 이용 현황 및 메일 사용자에 대한 통계치를 챠트를 사용하여 시스템 관리자가 쉽게 이해 할 수 있도록 분석해 준다. 이를 이용하여 운영자는 메일 시스템의 각종 이용현황을 쉽게 파악할 수 있음은 물론, 장애 경험이 많은 이용자나 스팸 메일을 발송하는 이용자 또는 스팸 메일을 이용하여 시스템에 장애를 일으키는 이용자들을 쉽게 파악할 수 있다. 또한 이러한 통계 자료를 바탕으로 앞으로 메일 서버 증설에 대한 근거 자료로도 활용할 수 있다. 논문의 구성은 다음과 같다. 제2,3장에서는 관련 연구 분야에 대한 기술이며 제4장에서는 메일 로그 분석 시스템의 구조 및 설계 부분이다. 제5,6장에서는 메일 로그 분석 시스템의 실행 예와 결론 및 향후 연구 방향에 대한 기술이다.

2. 관련 연구

메일에 관한 통계 정보를 제공하는 메일 로그 분석기는 크게 syslog 로그파일을 이용한 스크립트 프

로그램과 샌드메일이 저장하는 메일 정보를 이용하여 통계 정보를 구하는 방법이 있다.

2.1 syslog-stat.pl

샌드 메일의 syslog 파일을 읽어 메일에 관한 통계 정보를 제공해준다. Paul Vixie이 Perl로 작성하였으며 syslog 파일이 syslog.0, syslog.1,... 이런한 방식으로 요일별로 구성된 경우 모든 syslog 파일들을 읽어 간단한 통계를 내준다[1]. 간단하게 메일에 관한 통계 정보를 얻을 수 있는 장점이 있다.

2.2 anteater

Peter Hanecak 이 C++로 작성 한 로그 분석 도구이다[2]. syslog 파일을 이용하여 메일 통계를 작성한다. 스크립트로 작성된 로그 분석기 보다 좀더 상세한 메일 통계 정보를 제공해 주기 때문에 많이 사용되고 있다.

2.3 mailstats을 이용한 분석

샌드 메일은 송/수신 메일에 대한 통계 정보들을 sendmail.st 파일에 저장한다. 콘솔상에서 mailstats 라는 명령어를 이용하면 sendmail.st에 저장된 통계 정보를 (그림1)과 같이 메일 통계 정보를 얻을 수 있다[3][4].

Statistics from Wed Oct 24 13:51:42 2001

M	msgsfr	bytes_from	msgsto	bytes_to	msgsrej	msgsdis	Mailer
3	52	62K	21	23K	6	0	local
5	0	0K	13	16K	0	0	esmtp
T	52	62K	34	39K	6	0	

(그림1) mailstats을 이용한 메일 통계 정보

- M : Mailer
- msgfr : 받은 메시지 수
- bytes_from : 받은 메시지 크기
- msgsto : 보낸 메시지 수
- byte_to : 보낸 메시지 크기

(그림1)에서 받은 메일은 52통이고 총 크기는 62KByte 임을 알 수 있다. 유닉스 시스템의 cron을 이용하여 mailstats 명령어를 하루 단위로 실행하면 메일에 관한 통계 정보를 하루 단위로 계속적으로 유지할 수 있다.

지금까지 메일 통계 정보를 얻을 수 있는 로그 분석기에 대한 관련 연구 분야를 살펴보았다. 그러나 관리자가 원하는 메일 통계 정보를 제공해 주기에는 메일 통계 정보가 매우 부족하다. 이에 3장에서는 syslog 분석을 통한 세부적인 통계 정보를 관리자에게 제공해 줄 수 있는 방법을 살펴보고자 한다.

3. 로그 정보 분석

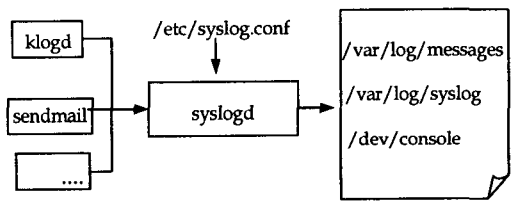
(그림2)는 여러 종류의 로그들이 기록되는 구조를 나타낸 그림이다. syslogd는 시스템의 로그를 관리하는 데몬이다. syslog.conf 설정 파일에 정의되어 있는 설정 값을 참고하여 에러 단계에 따른 로그들을 지정 위치에 기록한다. syslog.conf 파일은 selector와 action으로 나뉘어져있다[5][6].

- Selector : 어떤 종류의 메시지를 기록하는 정의
- Action : 메시지 저장할 곳을 정의

다음과 같이 syslog.conf 파일에 설정되어 있다면,

mail.debug /var/log/syslog

샌드메일이 발생하는 로그 중에서 debug 이상의 중요도를 가지는 것들은 /var/log/syslog 위치에 저장하는 것을 의미한다. 이러한 정보들을 이용하여 syslogd 데몬을 실행한 후 샌드메일을 이용한 메일 송/수신하는 경우 모든 메일 로그 기록들이 로그파일로 기록되는 것을 알 수 있다.



(그림2) 로그 파일 생성 구조

3.1 syslog의 저장 형태

syslog 파일에는 여러 종류의 로그 정보들이 저장된다. 그 중에서 샌드메일과 관련 있는 로그들을 선별하기 위해서 'sendmail' 단어가 포함된 라인만 선택하면 다음과 같은 형태의 로그들을 추출할 수 있다.

• Feb 10 13:48:17 rtlab sendmail[7621]: NAA07621: from=<..>

• Feb 10 13:48:17 rtlab sendmail[7622]: NAA07621: to=<..>
기록된 메일 로그 형태를 분석하기 위해서 샌드 메일의 mail.c를 살펴보면 다음과 같은 행이 존재한다.

openlog("sendmail",LOG_PID,LOG_MAIL);

여기에서, 샌드 메일의 저장 형태를 유추할 수 있다.

1) 로그 파일에 sendmail이라는 이름으로 기록한다.
2) LOG_PID : PID(process identification number)의 정보와 함께 기록한다.
3) sendmail의 로그 타입은 LOG_MAIL로 저장한다.

syslog 파일 중에서 1), 2), 3)의 단어를 포함한 로그 기록들을 추출하면 메일에 관련된 통계 정보를 얻을 수 있다. 보낸 사람/받은 사람의 메일 주소 및 메일 크기에 대한 세부 정보를 얻기 위해서 LOG_PID를 이용하면 된다. 예를 들면 [7621], [7622]에 대한 필드를 통하여 어떤 사람이 누구에게 메일을 보냈는지에 대한 정보를 얻을 수 있다.

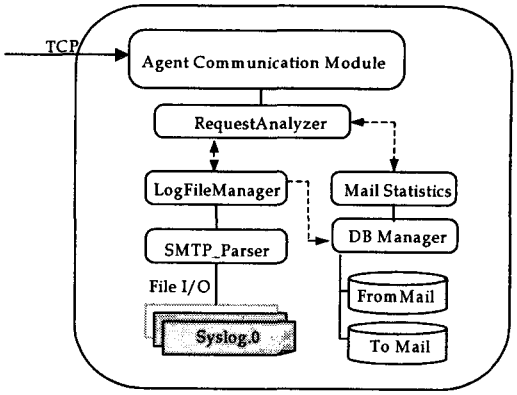
3.2 syslog 설정 파일

syslog를 이용한 로그를 분석하기 위해서 로그 파일의 저장 위치를 알아야 한다. 일반적으로 syslog 설정 파일은 /etc/syslog.conf에 위치한다.

syslog.conf 파일을 살펴보면 샌드메일이 남기는 로그 파일의 위치를 찾아 낼 수 있다. 리눅스 계열은 /var/log/ 아래 저장하는 것이 일반적이며 솔라리스 같은 경우는 /var 밑에 저장하고 있다.(/var/adm, /dev/log 등등)

4. 메일 시스템 전체 구조

웹 기반의 메일 통계 시스템의 전체 구조는 (그림3)과 같다. 웹 클라이언트를 통한 접근으로 시간, 장소의 구애 없이 메일 통계에 대한 정보를 얻을 수 있다.



(그림3) 웹 기반의 메일 통계 시스템 구조

1) Agent Communication Module

사용자의 접속을 TCP 소켓을 이용해서 대기하고 있는 모듈이다. 웹 클라이언트에서 요청이 있을 경우 사용자와 연결설정 과정을 통해 엔진과 통신을 할 수 있도록 통신 기능을 제공하는 모듈이다.

2) Request Analyzer

통신 모듈에서 받은 클라이언트 소켓을 처리하는 모듈이다. 사용자의 요구를 분석하여 해당 모듈로 넘긴 후 일을 처리하여 사용자에게 다시 보내는 역할을 한다.

3) LogFileManager

syslog의 파일에 대한 입출력을 제어하는 모듈이다. SMTP_Parser에서 읽은 로그 파일을 다음과 같은 FromMail 과 ToMail 형태로 변환한 후에 저장한다.

- 받은 메일의 포맷 (FromMail)

AgentIP	FromID	Date	Time	Mail	Size	Relay
(10.51.12.202	12918	20020226	17:09:16	skku@rtlab.skku.ac.kr	740	10.51.12.181)

- 보낸 메일의 포맷(ToMail)

AgentIP	FromID	ToID	Date	Time	Mail	Mailer
(10.51.12.202	12918	12920	20020226	17:09:16	sdlee@rtlab.skku.ac.kr	relay)

위와 같은 형태로 테이블에 저장한다. FromMail, ToMail 테이블의 FromID, ToID 필드를 이용하여 보낸 사람/받는 사람에 해당하는 메일 주소를 얻을 수 있다.

4) Mail Statistics

FromMail, ToMail 테이블에서 데이터를 처리하여 원하는 통계 정보로 변환하는 모듈이다. 다음과 같은 메소드들을 이용하여 통계 정보를 얻을 수 있다.

```
public Class MailStatistics {
 Vector MakeSMTPMailToTal();
 //받은 메일/보내는 메일 대한 합계를 구하는 메소드
 Vector getMailRecvCountByDay();
 //하루 동안 받은 메시지의 수를 구하는 메소드
 Vector getMailSendCoutByDay();
 //하루 동안 보낸 메시지의 수를 구하는 메소드
 Vector getMailRecvWithSize();
 //받은 메일의 크기를 구하는 메소드
 Vector getMailSendWithSize();
 //보낸 메일의 크기를 구하는 메소드
 }
```

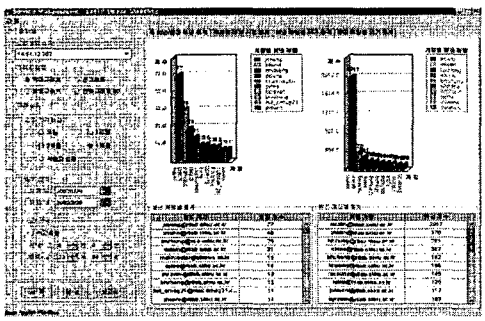
5) DB Manager

데이터 베이스를 관리하는 모듈이다. 로그 파일 분석을 통해서 얻은 받은 메일/보낸 메일에 관한 테이블에 접속 한 후 해당 테이블을 액세스 할 수 있도록 기본적인 Primitive 함수들, 즉 insert, update, delete 등을 제공해 주는 모듈이다. JDBC를 이용하여 쉽게 통계 테이블에 대한 접근을 할 수 있다.

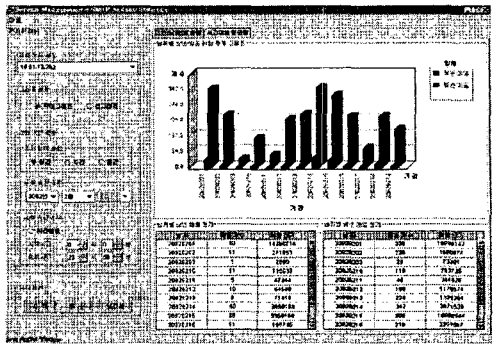
5. 구현 결과 및 특징

5.1 구현 환경

-구현 플랫폼 : SunOS 5.7
-데이터 베이스 : Mysql 버전 3.23.36
-웹 브라우저 : Internet Explorer 5.0
-하드웨어 : Pentium III/Ram 128MB
-구현 환경: java1.2 이상



(그림4) 시간대별 메일 분석



(그림5) 사용자별 통계 정보

5.2 웹 기반의 메일 통계 시스템의 특징

1) 웹 인터페이스를 통한 접근으로 언제 어디서나 메일 통계 정보를 볼 수 있다.
2) 통계 결과를 웹 인터페이스를 통해서 HTML로 저장 가능하다.
3) 다양한 리포트 양식을 지원한다.
4) 다양한 통계 기능을 제공한다.
- 날짜별 송/수신 메일 개수, 메일 사이즈
- 날짜별 송/수신 메일 평균 개수, 최대 개수, 최소개수
- 시간대별 송/수신 메일 개수, 메일 사이즈
- 시간대별 송/수신 메일 개수, 최대 개수, 최소 개수
- 보낸 사람/받은 사람 의 메일 주소

6. 향후 연구 방향

현재 웹 기반의 메일 분석 시스템은 유닉스 기반의 sendmail 8.9.3을 이용하여 구현되었다. 향후에는 여러 운영체제에서 사용 가능하도록 즉, 윈도우즈 계열의 운영체제에서도 메일 통계 정보 분석할 수 있도록 연구를 진행하고 있다.

[참고 문헌]

[1] http://mirror.csociety.org/pub/CPAN/scripts/mailstuff/
[2] http://anteater.drzoom.ch
[3]Sendmail Reference
[4]Sendmail FAQ - http://www.sendmail.org/faq
[5]http://www.sendmail.org/compiling.html
[6]http://www.megaloman.com/~hany/RPM/doors3.0