韓國 컴퓨터情報學會 論文誌
第10卷 第2號, 2005. 5.

2005-10-2-2-3

정책기반의 계층적 스팸메일 제어모델 설계

이 영 진*, 백 승 호**, 박 남 규***, 이 상 호****

Design of A Spammail Control Model Based on Hierarchical Policy

Yong-Zhen Lee *, Seung-Ho Baek **, Nam-Kyu Park ***, Sang-Ho Lee ****

요 약

초고속 인터넷망의 확산에 따른 인터넷 이용과 전자상거래의 급격한 성장과 함께 저비용, 고 효율적 특성의 전자우편 광고가 마케팅 수단으로 각광을 받고 있다. 반면 스팸메일의 빠른 속도의 증가로 인하여 메일서비스업체와 메일사용자에게 정신적 경제적 피해를 주고 있는 것이 현실이다. 이 논문에서는 대학을 중심으로 효율적인 스팸메일 차단을 위해 참여자-사용자, 관리자 그리고 ISP들이 상호 협력하는 계층적 스팸메일 차단정책을 설계하고 그 정책을 기반으로 스팸메일에 효율적으로 대응하는 스팸메일 제어모델을 제안한다. 또한 제안모델에 대한 분석평가를 통하여 모델의 효율성을 보인다.

Abstract

As the internet and E-commerce have been developing, a novel method for marketing is needed. A new advertisement using E-mail is becoming popular, because it has characteristics with low costs and relative efficiency. However, as the spam mails are increasing rapidly, mail service companies and users are deeply damaged in their mind and economically. In this paper, we design a hierarchical spam mail blocking policy through cooperation of all the participants-user, administrator, ISP to cut off the spam mail efficiently and propose an efficient model to block and manage the spam mails based on the policy. Also we prove the efficiencies and effectiveness of the proposed model through evaluation process.

▸ Keyword : 스팸메일(Spam Mail), 계층적 정책(Hierarchical Policy), 필터링 엔진(Filtering Engine)

• 제1저자 : 이영진
• 접수일 : 2005.04.08, 심사완료일 : 2005.05.20
* 충북대학교 전자계산학과 박사과정, ** 충북대학교 전자계산학과 석사과정, *** 충북대학교 전산정보원,
**** 충북대학교 전기전자컴퓨터공학부 교수
※이 논문은 2004년도 한국교육전산망 연구 지원 사업의 연구비 지원에 의하여 연구되었음

Ⅰ. 서 론

정보통신기술의 발전과 인터넷 보급확산으로 인터넷 사용자가 급증하고 있으며, 정보사회에서 전자우편은 유용한 정보의 제공 및 교환을 위한 가장 편리한 통신수단으로 자리잡고 있다. 전자우편은 기존의 문서우편기능을 일부 대체하면서 인터넷 사용자의 편익을 제공하고 있으나, 스팸메일의 대량 수신에 따른 관련 서버의 과부하 및 인적·물적 자원이 낭비되고 있으며 스팸메일에 대한 일반 사용자들의 심리적 부담감 및 이에 따른 정신적 피해가 날로 심각해지는 것이 현실이다.

일반적으로 스팸메일이란 상업적 용도의 광고선전을 위해서나 불건전한 정보 등 불특정 다수에게 뿌려지는 전자우편을 의미한다. 즉, 스팸메일은 수신자가 원하지 않는 전자우편이라 할 수 있다[1].

온라인 리서치 전문 업체인 엠브레인이 2004년 7월 13일부터 16일까지 4일간, 10대 이상 남·여 2000명을 대상으로 조사한 '스팸메일 현황 조사' 자료에 따르면 네티즌 응답자(2000명)의 99.7%가 스팸메일을 받아 본 경험이 있는 것으로 드러났다[1].2003년 5월에 스팸메일 일일 수신건수가 50건에 이르렀으며, 이들 중 63%는 위법 정보를 담고 있는 것으로 나타났다. 이러한 스팸메일로 인한 우리나라의 경제적 손실은 2003년 기준 연간 약 1조 3,810억 원에 달하는 것으로 나타났다[2].

기존 스팸메일 차단 솔루션들은 대부분 단순 필터링기법을 사용하여 스팸메일을 차단하므로 날로 늘어나는 변종/변이/지능적 스팸메일을 차단하는데 근본적 한계를 보인다. 이러한 문제점을 해결하기 위해서 스팸메일 유통과정에 참여하는 업체들과의 유기적 협력 체제 구축 및 운영을 통하여 스팸메일에 보다 능동적으로 대응할 수 있는 새로운 제어 방법을 제안한다.

이 논문의 구성은 다음과 같다. 2장에서는 현존하는 대표적인 스팸메일 차단기법들에 대한 분석 및 스팸메일 차단을 위한 정책적 접근방법 및 규칙변화 분석 결과를 기술하고 3장에서 효율적인 스팸 차단을 위한 정책의 기술과 정책에 기반한 제어모델을 제안한다. 4장에서는 제안모델에 대한 평가 및 활용방안에 대하여 기술하며 5장에서 결론을 맺는다.

Ⅱ. 관련 연구

2.1 스팸메일 차단 기술

현재 많은 스팸메일 차단 제품들이 개발되어 운영되고 있는데 그들이 채택하고 있는 스팸 차단기법은 아래와 같다.

2.1.1 단순 필터링

이 기법은 수신측 메일서버나 수신자 단말에서 특정 발신자, 제목 또는 내용에 특정 단어 등을 설정하여 원하지 않는 메일을 걸려낸다. 이 기법은 스팸메일의 제거에 매우 효과적이나 삭제하지 않아여 할 메일까지 삭제되는 경우가 발생하는 문제점이 있다[1,3,4].

2.1.2 Bayesian Filtering

이 기법은 토마스 베이시스 확율론에 기반하여 2002년9월에 Paul Graham가 제안한 기법으로[5], 각 사용자별로 제시한 스팸분류 기준을 학습시키고 이를 기반으로 스팸메일을 차단하는 방법이다. Bayesian Filtering은 시간이 지남에 따라 그 효율성이 높아지는데

2.1.3 RBL

초기 RBL(Realtime Blackhole List)은 관리자의 판단에 따라 수작업으로 관리되는 Host의 BlackList 였다. 이 리스트는 서버에서 그들의 고객을 위해 사용되는 IP 주소를 가지고 있으며 관리자는 이 리스트에 등록된 IP 주소를 이용하여 직접 SMTP 연결을 거부하도록 설정할 수 있다. 이 방법은 소수 서버 사용자의 스팸메일로 전체 이용자가 피해를 줄 수 있으며, 선의의 사용자의 IP 주소 스푸핑으로 인한 피해가 발생할 수 있다.

2.1.4 SpamAssassin

SpamAssassin[6]은 Justin Mason이 펄(Perl)로 제작한 스팸메일 차단 프로그램으로 전자우편을 분석하는 많은 다른 테스트들(메일의 속성정보)을 통해서 동작한다. 모든 테스트가 실행되고 나면 이메일은 점수를 획득하게 되는데 사용자가 ASCII 설정파일에 명시한 규칙에 의해 각각의 테스트가 얼마나 많은 점수를 획득하게 되는지가 결정된다. 만일 총점이 정해진 규정점수를 넘지 못하면 SpamAssassin은 그 메일이 스팸일 것이라고 판단한다.

SpamAssassin의 가장 큰 잠재적 위험은 정상메일을 스팸메일로 잘못 분류하는 상황이다. 따라서 정기적으로 스팸메일을 확인할 필요가 있다. SpamAssassin의 감도를 낮게 설정 하면 걸러지지 않는 스팸의 수는 늘어나고 적절한 균형을 찾는 것이 어려운 문제로 남아 있다. 때문에 일반 사용자들이 사용하기 어려우며 또한 편리한 GUI환경을 지원하지 못하고 있다.

2.1.5 SpamNet

SpamNet은 클라우드마크회사가 개발한 스팸차단 솔루션이다. SpamNet은 스팸메일이 대량의 같은 내용을 갖는다는 특성을 기반으로, 신고된 스팸수에 따라 스팸을 판정하는 기술을 사용하며, 사용자들의 스팸메일 처리결과를 클라우드마크와 공유하고 또한 스팸메일을 각 사용자의 메일함에 들어가기 전에 분류한다[7].

2.2 스팸메일의 규제 방법

스팸메일에 대한 규제방법은 크게 'Opt-In방식'과 'Opt-Out방식'으로 구분할 수 있다[8].

OPT-out 방식은 수신을 거부하지 않은 모든 수신자에게 전자우편을 보낼 수 있는 방식이다. 즉, 어떤 경로를 통해서든 전자우편 계정을 구해서 수신자에게 전자우편을 보낼 수 있고, 이에 대해 수신자가 거부 의사를 밝히면 더 이상 전자우편을 발송할 수 없다. 그러나 수신자의 거부가 없으면 허락한 것으로 간주해 나중에도 전자우편을 계속해서 발송할 수 있다.

OPT-in 방식은 수신자가 허락한 송신자 이외에는 어떤 송신자로부터도 전자우편을 받지 않도록 하는 방식이다. 즉, 송신자가 광고성 전자우편 등을 보내기 위해서는 수신자의 동의를 미리 구해야 되며 수신자의 리스트에 등록되어 있어야만 메일의 송수신이 가능한 방식이다.

2.3 스팸메일 차단규칙 분석

스팸 차단 솔루션에서 사용되는 규칙의 변화를 파악하기 위하여 C대학에서 사용하고 있는 스팸차단 솔루션의 4월부터 10월 까지의 스팸 차단 규칙을 분석하였다. 이 중 일별 스팸 차단 규칙변화는 규칙의 증가수가 많은 9월과 10월을 선정하여 분석하였으며, 스팸 차단 규칙 일별 변화는 (그림 1)에서와 같이 스팸 차단 규칙이 매일 증가되고 있음을 알 수 있었고, (그림 2)의 스팸 차단 규칙 월별 변화는 월별로 스팸 차단 규칙이 대폭 증가되고 있음을 보여준다.

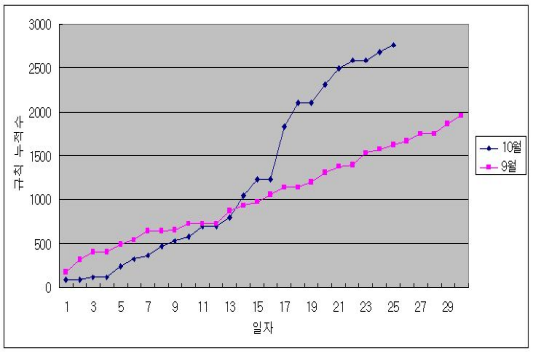

그림 1. 스팸 차단 솔루션에서의 일별 스팸 차단 규칙 패턴 변화(2004년)
Fig 1 Change of Spam Blocking Rule Pattern Per Day in Spam Blocking Solution

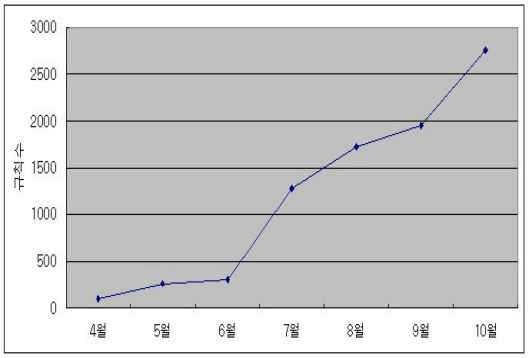

그림 2. 스팸 차단 솔루션에서의 월별 스팸 차단 규칙 패턴 변화(2004년)
Fig 2 Change of Spam Blocking Rule Pattern Per Month in Spam Blocking Solution

(그림 1)의 일일 평균 스팸 규칙 증가를 보면 9월에는 매일 평균 78.2개 증가, 10월에는 매일 평균 138.2개의 규칙이 증가함을 알 수 있다. 따라서 매달 평균 스팸 차단 규칙이 최저 약 2천개가 증가되고 이 추세는 연간 약 2만개 규칙이 증가함을 의미한다. 이는 솔루션에 추가되는 규칙 중 무의미한 스팸 차단 규칙들이 누적되고 있음을 암시한다. 그러므로 어떤 솔루션을 사용하던 스팸 차단 규칙의 지속적인 추가, 삭제 및 수정 등의 관리가 필요하게 되며, 무의미한 스팸 차단 규칙들의 누적은 스팸 차단 솔루션 성능의 양적, 질적 저하를 가져오는 주요 원인이 된다.

III. 정책기반의 계층적 스팸 제어모델

3.1 모델의 개요

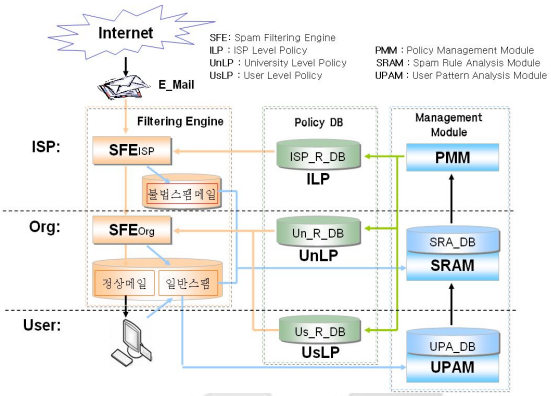

그림 3. 정책기반 계층적 스팸메일 제어 모델
Fig 3. Policy based Hierarchical Spammail Control Model

현재 많은 스팸 차단 솔루션들이 연구 개발되어 대학 등 각급 기관에서 사용되고 있지만 대부분 솔루션들은 기관 단위로 독립적으로 운영되고 있다. 이 논문에서는 (그림 3)과 같이 스팸메일을 방지하기 위한 정책기반 계층적 스팸메일 제어(PHSC : Policy based Hierarchical Spammail Control) 모델을 3계층(ISP, Org., User)의 3기능(스팸 필터링 엔진, 스팸 정책 데이터베이스, 관리-분석 모듈) 구조로 설계하여 각 개체들 간의 스팸 차단 관련 정보의 유기적 공유를 통하여 스팸메일에 효율적으로 대응할 수 있도록 한다.

(그림 3)의 최상위 계층인 ISP 계층은 정보통신사업자 즉 우리나라 대학의 경우 교육전산망에서의 KT나 Dacom을 의미한다, 이 계층의 주요 가능은 인터넷을 통한 스팸메일을 정책 DB를 통해 스팸메일을 필터링하는 역할을 한다. 그리고 중위 계층인 Org 계층은 ISP가 서비스하는 망에 가입된 기관을 의미하며, 최상의 도메인이 교육전산망인 경우 가입된 모든 대학들을 의미하며, 각 기관의 정책 DB를 통해 ISP에서 필터링 되지 않은 스팸메일을 차단한다. 마지막으로 User 계층은 각급 기관의 메일서비스를 이용하고 있는 사용자 대학의 경우 교수, 직원, 학생 등 모든 사용자를 말한다. 이 계층에서는 사용자의 개인 취향에 따라 수신된 메일을 필터링할 수 있도록 설정이 가능하도록 한다.

3.2 스팸메일 차단 정책 설계

PHSC 모델의 3개 계층에 기반하여 각 계층에서 요구되는 스팸차단정책을 설계한다. 이 정책은 스팸메일 차단 및 분류를 위한 목적으로 사용되며, 정책의 설계는 정책에 대한 표현과 그 정책을 적용하기 위한 논리적 연산의 형식 표현이다[15][16]. 정책의 표현은 정책 기본 정보, 정책의 적용 대상 정보 및 정책 규칙 표현 방식 등으로 다음과 같이 구성한다.

첫째, 정책 기본 정보는 <표 1>에서와 같이 정책 번호, 정책 유형, 등록 기관(대학), 등록일(Date), 비고 등의 항목으로 구성한다.

표 1. 정책기본정보
Table 1. Policy Basic Information

항 목	설 명
정책번호	정책을 구분하기 위하여 부여되는 고유 번호
정책유형	1. 일반필터링; 2. 지능필터링; 3. OPT_OUT 4. OPT_IN; 5. 인증방식; 6. 기타
등록대학	등록된 정책을 사용 중인 대학 명
등록일	정책을 등록한 날짜
비고	기타 요구사항이나 정책에 대한 설명

둘째, 정책의 적용대상정보 및 정책 규칙 표현 방식에 대한 3개 계층별 정의는 다음과 같다.

3.2.1 ISP 계층에서의 정책(ILP) 표현

ISP계층에서 스팸메일 차단을 위한 정책 표현 항목은 <표 2>에서와 같이 주로 차단하려는 도메인과 메일서버 주소와 헤더, 제목으로 구분하여 표현하고 그 정보를 실시간으로 이용할 수 있도록 한다. ISP계층의 정책을 이용하는 모든 기관(대학)은 공동 스팸(예를 들어 제목 수준에서의 특정 내용 등) 및 각 기관(대학)별 신고 스팸 (기관에서 의뢰한 스팸메일)을 근거로 스팸 산생 근원지를 차단하여 교육망 내에서의 유해 트래픽 및 스팸메일 피해를 줄일 수 있게 된다. 즉, ISP 차원에서 명백한 불법 스팸메일을 차단함으로써 각 기관(대학)에서는 불범 스팸메일에 대한 처리부담을 가지지 않게 된다.

표 2. ISP계층에서의 스팸 정책 표현
Table 2. Policy Presentation in ISP Hierarchy

항목	설 명
정책번호	사용되고 있는 정책을 구분하기 위하여 부여한 고유번호
차단유형	1. 도메인; 2. 메일서버주소; 3. 헤더; 4. 제목
차단내용	스팸차단을 위해 선전된 단어, 도메인네임, IP 등
적용동작	1. 포함; 2. 일치; 3. 유사;
등 록 자	1. ISP; 2. 대학, 3. 사용자(사용안함)
등록방법	1. 수동; 2. 자동
등 록 일	본 정책이 등록된 일자
비고	기타

3.2.2 기관(대학) 계층에서의 정책(UnLP) 표현

표 3. 대학계층에서의 정책(UnLP) 표현
Table 3. Policy Presentation in University Hierarchy

항목	설 명
정책번호	사용되고 있는 정책을 구분하기 위하여 부여한 고유번호
차단유형	1. 도메인; 2. 메일서버주소; 3. 헤더; 4. 제목; 5. 본문
차단내용	스팸 차단을 위해 선정된 단어, URL, IP 등
적용동작	1. 포함; 2. 일치; 3. 유사;
등 록 자	1. ISP(사용 안함); 2. 대학 3. 사용자
등록방법	1. 수동; 2. 자동
등 록 일	본 정책이 등록된 일자
비고	기타

기관(대학) 계층에서의 스팸 정책은 <표 3>에서와 같이 기관(대학)별 상황에 따라 서로 다르게 정의한다. 예를 들어 어떤 기관(대학)에서는 일반 필터링 방식을 사용하고 또 어떤 기관(대학)에서는 OPT-in/OPT-out이나 인증 방식을 사용한는 경우 이 계층에서는 기관(대학)별 스팸메일 규제 특성을 기반으로 스팸메일 차단 정책을 수립하고 그 정책에 기반하여 스팸 차단 규칙을 설정할 수 있다.

3.2.3 사용자 계층에서의 정책(UsLP) 표현

사용자 계층에서의 스팸 차단 정책은 <표 4>에서와 같이 주로 기관(대학)차원에서 설정된 규칙을 기반으로 분류된 정상메일과 스팸메일의 운용 과정에서 얻어진 정보를 토대로 설정된다. 즉, 기관(대학) 차원에서의 정책에 의하여 운영된 결과는 사용자의 요구에 위배되는 결과를 대상으로 정책이 등록된다. 예를 들어 정상메일로 판정된 메일 속에서 수신을 원하지 않는 메일을 차단하기 위하여 정책 등록을 요청할 수 있으며, 스팸으로 판정된 메일 중에 사용자가 수신을 원하는 메일에 대하여 정상메일로 처리해 주기를 원하는 사용자 요구 기반의 정책들로 이루어진다.

표 4. 사용자 계층에서의 정책(UsLP) 표현
Table 2. Spam Policy in User Hierarchy

항목	설 명
정책번호	사용되고 있는 정책을 구분하기 위하여 부여한 고유번호
차단/허용	1. 차단; 2. 허용
유 형	1. URL; 2. 메일서버주소; 3. 헤더; 4. 제목; 5. 본문
내 용	스팸 차단을 위해 선전된 단어, URL, 메일주소, IP 등
적용동작	1. 포함; 2. 일치; 3. 유사
등 록 자	1. ISP(사용 안함); 2. 대학(사용 안함); 3. 사용자
등록방법	1. 수동; 2. 자동
등 록 일	정책이 등록된 일자
비고	기타

3.3 제어 모듈 설계

이 절에서는 PHSC 모델내의 각 모듈들의 기능과 동작 과정에 대해 기술한다.

3.3.1 불법 스팸 차단 모듈

불법 스팸 차단 모듈이란 ISP 차원에서 동작하는 스팸 차단 시스템으로 주로 불법 스팸 발송 근원지로 인정되는 도메인 및 메일 서버로부터의 모든 메일 및 명백한 불법 스팸메일들을 차단하는 모듈이다.

ISP 계층을 구성하고 있는 스펨메일 필터링 엔진(불법 스팸 차단 모듈)에서는 불량(Black) IP 주소를 가진 메시지를 무조건 차단한다. 그 다음 과정은 ISP 계층의 정책(ILP)을 기반으로 수신된 메일이 스팸메일 여부를 판정하고 지정된 규칙에 의하여 스팸메일로 인정된 불법 스팸메일은 태그를 부여하여 불법 스팸메일 데이터베이스에 저장한다. 스펨메일 규칙 분석 모듈(SRAM)에 의하여 새로운 스팸 규칙을 찾거나 의미 없는 규칙이나 이미 스팸 규칙으로 볼 수 없는 규칙을 수정하거나 삭제하는 작업에 이용된다. ISP에서의 스펨메일 필터링 엔진의 동작 과정은 (그림 4)와 같다.

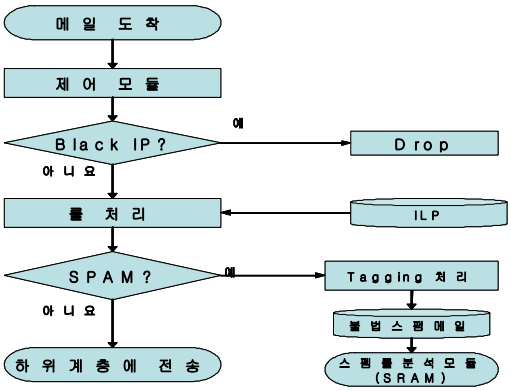

그림 4. ISP에서의 스팸 필터링 엔진의 동작
Fig 4. Spam Filtering Engine Action in ISP Hierarchy

3.3.2 일반 스팸 분류 모듈

각 기관(대학)에서 운영되고 있는 스팸메일 차단 규칙들은 해당 기관의 메일 특성에 따라 운영 정책을 달리하므로 그 내용들이 서로 다르다. 따라서 기관(대학) 계층에서 동작하는 스팸 차단 모듈은 각 대학에서 스팸 차단 정책을 정의하고 그 규칙을 기반으로 차단 정책에 위반되는 메일을 스팸으로 분류하여 서비스하는 것이 바람직하다. 또한 바이러스 메일에 대하여는 주로 백신을 사용하여 메일을 실시간으로 검사하고 차단한다. 도메인 및 주소기반 메일 차단이란 스팸 발송 근원지로 인정된 도메인 및 메일서버에서 오는 메일을 검증하지 않고 차단한다.

기관 레벨에서 동작하는 일반 스팸 분류 모듈은 크게 기관(대학) 스팸 규칙 기반 스팸 분류와 사용자 규칙 기반으로 스팸메일을 분류하며 그 주요 기능은 (그림 5)와 같다. 기관(대학) 공동 스팸 규칙 기반 스팸 분류 모듈은 기관(대학) 차원에서의 메일서비스 가용성을 보장하기 위한 목적으로 설정한 스팸 방지 정책으로서 기관(대학) 내의 모든 사용자가 공동으로 지켜야 할 규정을 의미한다.

사용자 규칙 기반 스팸메일 분류 모듈은 사용자들의 다양한 요구를 충분히 고려한 스팸메일 제어를 위하여 대학 차원에서 정상메일에서 사용자가 원하지 않는 메일을 추출하고 대학 공동 스팸 규칙에 스팸으로 분류된 메일에서 사용자가 수신을 원하는 것들을 추출하여 사용자 기반의 규칙을 업그레이드한다. 또한 사용자가 공동 규칙에 의하여 분류된 스팸에 대한 프로세스 패턴을 분석하기 위해 최종 정상메일과 스팸메일로 분류된 결과를 사용자 패턴 분석 모듈에 전달한다.

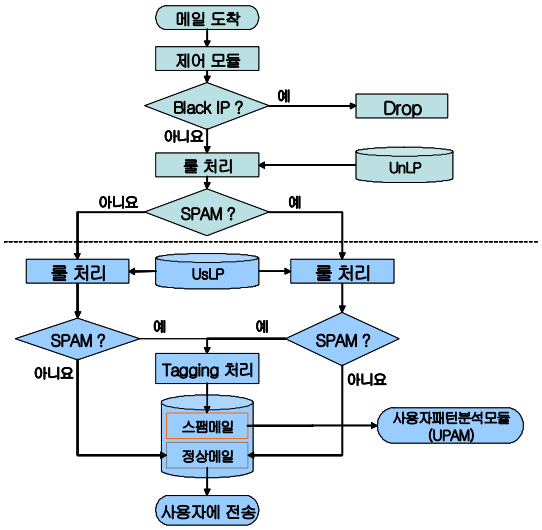

그림 5. 기관(대학)에서의 스팸 필터링 엔진의 동작
Fig 5. Spam Filtering Engine Action in University Hierarchy

3.3.3 정책 관리 모듈

정책 관리 모듈(PMM)은 (그림 6)과 같이 정책 분류 기준에 근거하여 스팸 규칙 분석 결과를 ISP 레벨 정책, 기관(대학) 레벨 정책 및 사용자 레벨 정책으로 분류하여 각 레벨 정책 데이터베이스를 주기적으로 자동 갱신시키는 기능을 수행한다.

정책 분류 기준 등록 관리에서는 스팸메일 관련 정책 관리자가 정책의 기본정보를 등록하고 그에 따른 구체적인 정책 분류 기준을 설정한다. 그 분류기준에 따라 각 레벨의 스팸 차단 정책을 추가, 수정, 삭제 등의 관리 작업을 수행한다. 정책 분류 기준은 스팸 실적 분석 결과와 각 계층에서의 정책 변경 요청에 따라 설정된다.

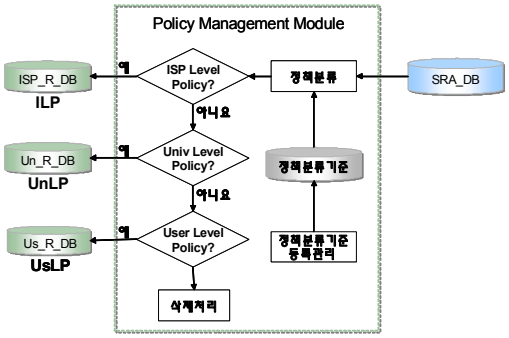

그림 6. 정책 관리 모듈의 동작과정
Fig 6. Action Process of Policy Management Module

3.3.4 스팸 규칙 분석 모듈

스팸 규칙 분석 모듈은 ISP 필터링에서 분류된 3개의 분석결과(스팸메일 결과, 기관(대학) 차원에서 분류된 스팸메일 결과, 사용자 패턴분석 결과)를 통해 각 계층의 스팸 분류 실적을 통계처리하고 분석한다. 이러한 통계분석을 통하여 최적화된 스팸메일 차단 규칙을 설정하여 주기적인 규칙 데이터베이스를 정책을 업그레이드 하는 기능이 이루어진다.

스팸 규칙 분석은 스팸 패턴 분석 결과와 규칙간의 관계 분석이다. 분석 결과는 스팸 규칙 분석 데이터베이스(SRA_DB)로 표현한다. 이 모듈은 사용자의 요청에 따라 스팸으로 분류해야 될 메일에 대한 도메인, 헤더, 제목, 본문 및 대상 메일주소 수신빈도에 대한 분석하고 중심의미의 단어를 추출하여 새로운 규칙을 생성한다. 그리고 그 결과는 ISP 계층이나 기관(대학) 계층의 스팸 차단 규칙에 활용된다.

3.3.5 사용자 패턴 분석 모듈

교육망의 대학들을 대상으로 설문 조사한 결과 스팸메일 대응에 있어서 극소수 사용자만이 클라이언트(Outlook)에서 스팸 차단 규칙을 설정하여 스팸메일을 차단하는 실정이다[14]. 그 주요 원인은 스팸 차단 규칙 설정이 일반 사용자들이 어렵다고 느끼고 있으며 또한 빈번하게 변화하는 스팸메일들에 효율적으로 대응하기가 불가능하다고 여기기 때문으로 파악되었다. 현재 스패머들도 단순하고 제한된 형식의 스팸메일을 만드는 것이 아니라 지능적 도구를 이용하여 다양한 변화를 가지는 스팸메일을 생성하므로 어느 한사람 또는 한 기관 혼자의 힘으로 스팸에 대응하기엔 역 부족이다.

사용자의 스팸메일에 대한 패턴 분석은 두 부분으로 나눌 수 있다. 하나는 정상메일 속에서 수신을 원하지 않는 메일에 대한 처리패턴과 스팸메일로 분류된 메일에서 수신을 원하는 메일에 대한 분석이고, 다른 하나는 분류된 스팸메일에 관한 실적정보를 통계 분석하는 것이다. 통계 분석한 결과를 저장하고 스팸 규칙 분석의 입력으로 사용하여 기관(대학) 차원에서의 효율적인 스팸 차단 규칙을 설정함에 있어 사용자 요구를 기반으로 하는 스팸 차단 규칙의 설정을 지원한다.

Ⅳ. 제안모델의 평가 및 활용방안

4.1 모델 평가

이 절에서는 스팸메일 차단 측면, 대역폭 측면과 사용자 신뢰도 측면에서 제안 PHSC 모델의 효율성, 가용성과 신뢰성 등을 평가한다.

4.1.1 차단 측면

스팸메일 차단을 위한 필터링 방식의 근본적 해결해야 할 문제점은 필터링에 사용되는 스팸 차단 규칙을 실지상황에 맞게 지속적으로 갱신하여 변종/변이/지능적 스팸메일에 대응 하는 것이다[10]. 어떤 방식이든 주기적으로 스팸 차단 규칙의 관리가 필요하며 이 부분이 소홀히 되는 경우 스팸 차단 시스템 성능이 저하된다.

제안 PHSC 모델에서는 스팸 차단 규칙 분석모듈에 의하여 실시간 변화하는 규칙의 변화에 상응하여 새로운 규칙의 추가가 이루어짐으로써 목적하는 스팸메일 차단 성능을 지속적으로 유지할 수 있으며, 의미 없는 규칙의 삭제 또는 무효처리를 통하여 시스템의 처리 성능을 향상시킨다. 또한 발신자, 수신자, 메일 사이즈, 내용 별로 수신된 메일에 대하여 분석을 하고 그 중 대량 메일(정책에서의 기준치를 넘은 발신자 또는 수신자)에 한해서 집중분석을 하고 또한 스팸메일 차단 규칙에 대한 실적 분석결과를 규칙에 지속적으로 반영하여 최신성과 구체적 상황에 맞는 규칙을 기반으로 스팸메일 차단함으로 기존 필터링 방식의 한계를 극복할 수 있다.

4.1.2 대역폭 측면

스팸메일은 유해 트래픽을 발생하는 주요 원인이기도 하다. 제안 모델에서는 스팸메일로 인한 유해 트래픽 발생을 초기에는 ISP, 중기는 대학, 후기는 사용자의 3 계층으로 각각 제어하여 차단하기 때문에 상대적으로 우수한 대역폭의 사용을 보장한다.

4.1.3 사용자 만족도

스팸메일의 최대 피해자는 사용자이다. 스팸메일 차단에 대한 사용자의 만족도는 두 가지 측면으로 볼 수 있다. 하나는 스팸메일을 얼마나 잘 차단하는가에 있고 다른 하나는 스팸메일 차단을 위한 차단 규칙 설정의 편리성 수준이다. 대부분의 현장 실정은 사용자가 수신을 원하지 않는 메일에

대한 수신거부의사가 제대로 관리자에게 전달되어 반영되지 않거나 반영이 되었다 해도 관리자의 수동처리 번거로움으로 인하여 실행하지 못하는 것이 대부분이다. 물론 사용자 클라이언트에서 차단 규칙을 설정할 수 있지만 초보자가 변종/변이/지능적 스팸메일 차단을 위하여 지속적인 차단 리스트를 갱신하는 것은 사실상 불가능하다.

제안 모델에서는 사용자의 의사를 고려하여 사용자가 수신한 정상메일과 스팸메일에 대한 처리 패턴을 분석하고, 사용자 패턴 분석 모듈을 통하여 사용자 스팸 차단 정책에 반영시킨다. 또한 스팸 분류 실적 통계를 분석하여 차단 규칙 갱신에 이용함으로써 사용자의 만족도가 향상된다.

4.1.4 스팸 차단기술 비교

<표 5>에서와 같이 최근에 사용되고 있는 스팸차단 기술들과의 비교분석을 통하여 제안 모델의 우월성을 입증한다.

표 .5. 스팸 차단 기술 비교
Table 5. Comparison of Spam Blocking Technology

기술 / 비교항목	RBL	Bayesian Filtering	SpamNet	Spam Assassin	제안모델
수집방법	수동	자동	수동	수동	자동
스팸정보 공유	가능	가능	불가능	서버별	가능
스팸 정보 분석	수동	자동	수동	자동	자동
스팸 협력대응	X	X	X	X	가능
스팸 초기차단	X	X	X	X	가능
사용자 편리성	낮음	좋음	보통	보통	좋음

4.2 활용 방안

4.2.1 협력적 스팸 대응

현재 기관들의 스팸메일에 대한 대응 방안은 개발업체로부터 스팸 차단 규칙을 다운 받아 갱신함으로써 실제 차단은 스팸메일 관련 솔류션 업체에 의존하고 있다[11].

제안 모델에서는 스팸메일 차단에 있어서 ISP, 기관(대학)과 사용자가 모두 참여하여 계층적으로 스팸메일 차단한다. 또한 기관(대학)별 구체적 상황에 따른 차단규칙과 실적을 서로 공유하여 스팸메일의 피해 상황과 변이/변화/지능화 상황을 파악할 수 있어 최적화되고 효과적인 스팸메일 차단이 가능해진다.

스팸메일을 원천적으로 제거하려면 내부적으로 스팸메일의 발생을 제어하고 외부적으로 스팸메일의 유통을 차단하여야 한다. 이를 위해 메일 전송 과정에 참여하는 모든 개체들의 적극적인 협력이 절실히 필요하다. 따라서 제안한 방안이 협력적 스팸메일 차단의 기초 자료로 활용될 수 있다.

4.2.2 효율적인 규칙 셋 설정

필터링기법을 이용한 스팸 차단 기법에서 효율적인 규칙 셋 설정이 스팸 차단율을 높일 수 있는 중요한 지표로 사용되고 있다[12,13]. 대부분 스팸 차단 솔루션에서의 규칙 셋 설정은 메일서버 관리자가 수동설정이나 솔루션개발 업체에 의존한다. 이런 방식은 메일서버 관리자가 스팸메일 변화 형식 및 피해 현황에 대한 실질적 상황 파악과 분석을 필요로 하여 관리자의 부담을 가중시켜 효율적인 대응이 어려운 실정이다.

제안 모델에서는 대학별 스팸메일 차단 규칙과 그 규칙으로 차단된 실적을 공유하고 또한 스팸규칙분석모듈을 통하여 대학별 맞춤형 스팸 차단 규칙이 자동 설정됨으로 규칙의 최신성과 효율성을 높일 수 있고 메일서버 관리자나 사용자의 추가적의 부담이 줄어든다.

4.2.3 능동적 정책 적용

스팸 차단을 위한 규칙 설정에 있어서 많은 규칙을 사용자들이 수동으로 확인하고 적용하기에는 전문적인 지식이 부족하고 또한 번거로움으로 인하여 대부분 메일서버 관리자에 의존하고 있다.

제안 모델에서는 스팸 차단 정책을 계층적으로 분리하고 스팸 차단 규칙을 정책에 따라 계층적으로 분류하여 적용하고 있다. 기관(대학)별로 서로 다른 스팸 차단 정책을 가지고 있으므로, 이는 정책관리 모듈을 통하여 실적과 함께 관리하므로 기관(대학)의 실정에 맞는 정책 적용이 가능하다. 그리고 정책 적용 효과를 실시간 파악하고 사용자 요구와 스팸메일 차단 실적데이터를 사용자 패턴 분석 모듈, 스팸 규칙 분석 모듈 및 정책 관리 모듈을 통하여 자동 분석하여 각 계층의 정책을 재분류, 정책 이전, 정책 확장, 삭제 관리 등 동적인 정책유지가 가능해 진다.

V. 결론

현재의 대부분의 스팸메일 대응 방안들은 단순 필터링기법을 사용하여 스팸메일을 차단하는 기능만을 수행하고 있어, 날로 늘어나는 변종/변이/지능적 스팸메일에 능동적으로 대응하기가 매우 어려운 실정이다.

이 논문에서는 효율적인 스팸메일 차단 방안을 연구하기 위하여 교육전산망을 대상으로 스팸메일 실태에 대한 설문조사와 관련 솔류션 분석하였고 그리고 인터넷 망을 구성하

고 있는 구성개체(ISP, 대학, 서버 및 개인)들의 협력을 통해 계층적 스팸메일 차단 정책을 수립하여 스팸메일에 보다 효율적으로 운영할 수 있는 스팸 차단 모델을 제시하였다. 제안 모델의 다음과 같은 특성을 갖고 있다. 첫째. 메일 유통 과정의 참여자들을 ISP, 기관(대학), 사용자의 3계층으로 조직화하여 스팸메일에 대해 다단계 협력 대응이 가능하다. 따라서 스팸메일 차단의 효율성 향상과 스팸메일로 인한 유해 트래픽의 조기 제어가 가능하다. 둘째. 스팸메일 차단 정책을 계층적으로 분리하고, 그 정책을 기반으로 스팸 차단 규칙의 공유 및 통합 관리하므로 스팸 차단 규칙의 최신성과 최적화가 가능하다. 셋째. 사용자의 의사를 충분히 반영한 스팸 차단 규칙 설정이 가능하다. 넷째. 스팸메일에 관한 정책, 규칙, 패턴 등 관리 및 분석모듈로 스팸 차단 규칙의 자동화된 업그레이드가 가능하다. 스팸메일 차단 관련한 향후 연구과제는 스팸메일 원천적으로 막을 수 있는 인증기반의 스팸메일 제어기법 연구도 필요하다.

참고문헌

[1] M. Sahami, S. Dumais, D. Heckerman and E. Hoeviz, "A Bayesian approach to filtering junk e-mail", In proceedings of Workshop on Learning for Text Categorization, 1998

[2] 이규태, "[e리서치] 스팸메일 현황 조사," 전자신문, 2004. 8. 18.

[3] J. D. M. Rennie, "ifile: An application of machine learning to e-mail filtering", KDD-2000 Text Mining Workshop Boston, MA USA, 2000

[4] G. lindberg, "Anti-Spam Recommendations for SMTP MTAs", RFC2505, IETF, February.1999

[5] Paul Graham, "A Plan for Spam", http://www.paulgraham.com/spam.html

[6] SpamAssassin home page http://spamassassin.org

[7] SpamNet home page http://www.cloudmark.com

[8] 불법스팸대응센터, "스팸차단 Best Practice 지침서 V 1.0," 2004.년 9월.

[9] 정보통신부, OECD 스팸대응 국제워크샵자료, 2004.년 9월

[10] 김현준, 정재은, 조근식, "가중치가 부여된 베이지안 분류자를 이용한 스팸 메일 필터링 시스템," 한국정보과학회 논문지 B VOL. 31 NO. 08 pp. 1092-1100 2004. 8.

[11] 박정선, 김창민, 김용기, "퍼지관계곱을 이용한 내용기반 정크메일 분류 모델," 정보과학회논문지: 소프트웨어 및 응용 제29권 제10호 2002. 10.

[12] 민도식, 송무회, 손기준, 이상조, "SVM 분류 알고리즘을 이용한 스팸메일 필터링," 정보과학회 2003년 춘계학술대회 VOL. 30 NO. 01 2003. 04.

[13] 정옥란, 조동섭, "개인화된 분류를 위한 웹 메일 필터링 에이전트," 한국정보처리학회 논문지 B VOL. 10 NO. 07 pp. 0853-0862 2003. 12.

[14] 이상호, "정책기반의 계층적 스팸메일 제어모델 설계", 한국교육전산망 연구과제보고서, 2004. 11.

[15] 김강, 전동식 "보안 정책 기반 침입탐지 시스템 모델 설계" 컴퓨터정보학회 논문지 제8권 제4호 2003.12 pp. 81-86.

[16] 손우용, 송정길 "통합보안 관리시스템의 침입탐지 및 대응을 위한 보안 정책 모델" 컴퓨터정보학회 논문지 제9권 제2호 2004.6 pp. 81-87

저 자 소 개

이 영 진

1997년 6월 중국 연변대학교 물리학과 이학석사

2003~현재 충북대학교 전자계산학 박사과정

백 승 호

2003년 2월 한밭대학교 컴퓨터공학과 공학사

2003~현재 충북대학교 전자계산학 석사과정

박 남 규

2004년 2월 충북대학교 전기전산공학과 공학석사

2000년 8월~현재 충북대학교 전산정보원 조교

이 상 호

1989년 2월 숭실대학교 컴퓨터네트워크학과 공학박사

1981년~현재 충북대학교 전기전자컴퓨터공학부 교수