Journal of Digital Contents Society Vol. 8 No. 4 Dec. 2007(pp. 451-458)

링크 유알엘 접속을 통한 스팸메일 자동 차단 방법에 관한 연구

정남철*

요 약

본 연구는 링크 유알엘 접속을 통해 스팸메일을 자동으로 차단하는 방법에 관한 것이다. 본 연구의 링크 유알엘 접속을 통한 스팸메일 자동 차단 방법은 다음과 같다. 1. 인터넷을 통해 연결되어 이루어지는 전자메일 시스템(서버)에서 수신되는 전자메일의 메시지 원본에 존재하는 링크 유알엘 정보를 추출하고, 2. 추출된 링크 유알엘 정보에 링크된 웹페이지에 접속을 수행하며, 3. 웹페이지의 컨텐츠 중에 미리 규정된 스팸 키워드가 존재하는 경우에 수신된 전자메일을 스팸메일로 분류하여 차단한다.

A Method to Block Spam Mail Automatically Through the Connection to Link URL

Nam-Cheol Jung*

Abstract

In this paper, I developed a method whereby spam mail is automatically blocked through the connection to link URL. The blocking system works as follows. First, the system extracts information of URL linked to electronic mail which was delivered from any server on the internet. Next, the system lets itself be connected to the web pages through this URL. Last, the system blocks the electronic mail if those web pages contain any key word which was defined as a clue to spam mail.

Keywords : 전자메일(e-mail), 스팸메일(spam mail), 키워드(keyword), Blocking System, 유알엘(URL)

1. 서론

인터넷을 이용한 전자우편은 의사소통의 중요한 수단이 되고 있다. 그러나 휴대폰과 이메일로 날아드는 귀찮은 스팸들은 끊이지 않고 있어 이들을 정리하는데 많은 시간과 노력이 필요하다. 스팸방지에 많은 관심을 갖는 이유는 스팸메시지들이 성인광고나 대출광고 등의 불법광고 외에 넷치기(사이버상의 사기행위 등)를 목적으로 하고 있어 이에 대한 피해가 심각하기 때문이다. 또 웜이나 신종 바이러스 등을 몰고 돌아다니는 매개체로서의 역할도 하고 있다. 스팸을 퇴치해야 하는 이유가 바로 여기에 있다. 그럼에도 불구하고 스팸메일은 인터넷을 통해 '암세포'처럼 무섭게 빠른 속도로 퍼져 가면서 전 세계 메일의 3분의 2를 차지하고 있다[1].

잘 알려진 바와 같이, 전자메일 시스템은 인터넷을 통하여 특정한 사용자와 데이터나 메시지를 교환하는 시스템을 말하고, 스팸메일은 수취하는 쪽의 의향을 무시하고 대량으로 퍼뜨리는 광고, 선전, 권유 등의 상업적인 목적을 갖는 전자 메일을 말한다. 보다 광범위하게는, 명시적인 수신거부의사에 반하는 영리목적의 광고성 정보메일, 시스템 성능을 저하시킬 목적으로 발송하는 대량메일, 프로그램저작권을 침해하는 불법 소프트웨어 등의 판매 광고메일, 공서양속을 해치는 성인용품 및 불건전 사이트 등의 광고메일 및 공포심이나 불안감을 유발하는 메일 등이 스팸메일로 간주될 수 있다.

한편, 이러한 스팸메일이 수신자에게 주는 폐해를 최소화하기 위해 정부에서는 정보통신망 이용촉진 및 정보보호 등에 관한 법률 및 그 하위 법에서 다양하고 구체적인 스팸메일 제한 규

※ 제일저자(First Author) : 정남철
접수일자:2007년09월14일, 심사완료:2007년09월28일
* 동남보건대학 웹컨텐츠과
ncjung@dongnam.ac.kr
▣ 본 연구는 2007년도 동남보건대학 연구비 지원에 의하여 수행된 것임.

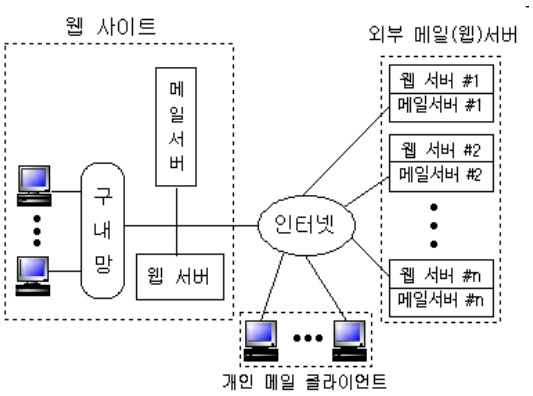
정을 마련해 놓고 이를 위반하는 경우에는 소정의 벌을 부과할 수 있도록 조치하고 있다. 한 예로 스팸메일 제목의 서두에 '(광고)'나 '(성인광고)'라는 키워드를 반드시 표시하도록 규정하고 있는데, 수신자가 스팸메일 여과 프로그램을 설치한 후에 이러한 키워드를 제목으로 갖는 전자메일을 스팸메일로 분류하도록 설정하면 이후 이러한 제목을 갖는 전자메일이 자동적으로 스팸메일로 분류되게 된다. 더욱 현재에는 메일 클라이언트가 스팸메일 여과를 위한 임의의 키워드를 설정할 수도 있어서 제목 또는 본문내용 중에 상기 설정 키워드를 포함하고 있는 전자메일이 자동으로 스팸메일로 분류되도록 되어 있다. 이러한 종래의 스팸메일 차단 방식을 '자체 정규 방식'이라 한다. 한편, 스팸메일로 걸러지는 것을 회피하기 위한 노력도 한층 강화되고 있는데, 특히 성인 사이트 광고메일의 경우에는 평범한 메일 제목 하에 본문에는 이미지만을 포함시킨 상태에서 성인사이트의 홈페이지 URL을 링크하는 방식이 많이 사용되고 있다. 그러나 전술한 종래의 기술에 따르면 수신 전자메일의 제목 또는 본문 중에 메일 클라이언트가 설정한 스팸 키워드가 존재하는지를 검사할 뿐 본문 내용에 링크된 URL로 접속되는 웹페이지의 컨텐츠까지는 검사하지 않기 때문에 앞서와 같이 변형된 스팸메일을 전혀 걸러 낼 수 없다는 문제점이 있다.

이러한 사회적, 기술적인 문제점을 조금이나마 해결하기 위하여 본 연구에서는 링크 유알엘 접속을 통한 스팸메일 자동 차단 방법을 통해 수신된 전자메일에 링크되어 있는 유알엘 정보를 추출하여 해당 웹페이지에 접속한 후 미리 규정된 소정의 스팸 키워드가 있는 경우 이를 스팸메일로 걸러 불법의 스팸메일에 대하여 사전에 자동 차단 방법을 제공하고자 한다.

2. 관련 연구

2.1 전자메일 시스템

(그림 1)은 일반적인 전자메일 시스템의 네트워크 구성도로서, 일반적인 전자메일의 송/수신은 인터넷에 연결된 전자메일 클라이언트(이하, '메일 클라이언트'라 한다)가 자체 웹 사이트에 구비된 메일서버(이하, '자체 메일서버'라 한다) 또는 자체 웹 사이트 외부에서 전자메일 서비스를 제공하는 각종 웹 사이트, 예를 들어 크고 작은 포털사이트나 ISP(Internet Service Provider)의 웹사이트에 구비된 메일서버(이하, '외부 메일서버'라 한다)부터 전자메일 계정을 부여받아 수행될 수 있다.



(그림 1) 일반적인 전자메일 시스템의 구성도

자체 웹사이트는 웹 문서를 처리하는 웹서버와 메일의 송/수신을 담당하는 메일서버 및 메일 클라이언트가 구내망, 예를 들어 랜이나 인트라넷을 통해 연결되어 이루어진다. 물론, 각각의 외부 웹사이트에도 웹서버와 메일서버가 구비되어 있다. (그림 1)에서 개인 메일 클라이언트는 자체의 웹사이트를 구비하지 않은 채로 주로 가정이나 개인 사무실 등에서 전화 회선이나 케이블, 위성, ADSL, ISDN 등을 통해 인터넷에 접속하며, 이러한 개인 메일 클라이언트는 주로 외부 메일서버에 가입하여 메일 계정을 무료로 할당받은 후에 전자메일 서비스를 이용한다. 물론 자체 웹사이트에 속한 메일 클라이언트들도 외부 메일서버에 가입하여 전자메일 서비스를 이용할 수가 있다.

한편, 전술한 전자메일 시스템에 있어서, 임의의 발신 측 메일 클라이언트가 내부 메일서버 또는 외부 메일서버를 통해 보낸 전자메일은 SMTP(Simple Mail Transfer Protocol: 단순 메일 전송 프로토콜)에 의해 수신 측 메일 클라이언트의 전자메일 주소로 대표되는 내부 또는 외부 메일서버의 메일박스로 전송된 후에 보관된다. 이후, 수신측 메일 클라이언트는 인터넷을

통해 자신의 메일박스에 접속하여 거기에 보존되어 있는 전자메일을 읽을 수가 있다. 수신측 메일 클라이언트가 자기 수신의 전자메일을 읽어 오는 방식으로는 메일서버의 메일 박스에 보존되어 있는 전자메일을 메일 클라이언트 컴퓨터의 하드디스크에 복사하는 방식인 POP3(Post Office Protocol #3: 메일서버로부터 전자메일을 받아 오기 위한 우체국 프로토콜) 방식 및 자기 수신의 전자메일을 메일 클라이언트 컴퓨터에 복사할 필요 없이 메일박스에 저장시킨 상태로 이용할 수 있게 하는 IMAP(Internet Message Access Protocol) 방식 등이 있는 바, 이러한 POP3나 IMAP은 통상적으로 내부 메일서버 시스템에서 많이 채택하고 있다.

최근에는 IMAP에 더하여 인터넷에 수신메일을 저장해 두고 어디에서나 접근할 수 있도록 하는 웹메일 서비스가 등장하였는데, 이러한 웹메일은 메일서버와 웹 서버가 합체된 서비스라고 할 수 있다. 즉, 메일 클라이언트 측에서 본 웹 메일서버는 일반 웹 서버와 아주 똑같은 형태를 갖고 있다. 따라서 메일 클라이언트가 웹 메일 서버에 접근하기 위해서는 웹 브라우저만 있으면 되고 전자메일 소프트웨어는 필요가 없다. 이러한 웹 메일 기능은 각종 포탈 사이트나 ISP 등에서 서비스 차원에서 무상으로 제공하고 있다.

2.2 스팸 메일 차단 방법

2.2.1 베이시안 필터링

베이시안 필터링(Bayesian Filtering) 기술은 미래 상황을 추측할 수 있게 해 주는 18세기 Thomas Bayes의 확률론에 근거로 한다. 이 기술은 전자문서 분류에 사용되고 있는 베이시안 분류 방법을 응용한 것으로 문서 내에 단어들을 대상으로 확률적인 방법을 적용하여 분류하기 때문에 특정 패턴에 따르지 않는 스팸메일을 걸러 낼 수 있다. 이는 어떤 사람에게는 스팸이 될 수 있는 메시지가 다른 사람에게는 유용한 정보가 될 수 있다는 사실에 대해 베이시안 스팸 필터는 각 개인별 스팸 분류 기준에 대해 학습한다. 이러한 학습 과정을 통해 베이시안 스팸 필터는 시간이 지남에 따라 그 효율성이 배가 되며, 일반적으로 99.8 퍼센트의 차단율과 0.05 퍼센트의 오탐지율을 보인다[2].

2.2.2 URL 빈도 분석

URL 빈도 분석을 이용한 스팸메일 차단 방법은 스팸메일을 수집하고, URL을 추출하여 이를 정규화하고 빈도 분석하여, 스팸 메일 차단에 사용되는 스팸 메일 판별 규칙을 생성하는 방법으로 구성된 스팸메일 차단 방법이다[3]. 스팸메일 수집은 가상 메일 주소를 만들어서 이를 통해 받는 모든 메일을 스팸메일로 규정하고, 이어서 URL을 추출하고 시간에 따른 가중치를 둔 빈도 분석을 통해 스팸메일 판별 자료를 만들어 내며, 이를 사용자 메일 서버에 전달하여 들어오는 메일에서 추출한 URL과 비교하여 스팸메일을 판별하고 차단할 수 있도록 한다.

이 방법은 기존의 스팸메일 솔루션보다 스팸메일 탐지율은 떨어지지만, 일반메일 오탐율이 낮아 일반메일을 스팸메일로 차단할 가능성이 적은 것이 특징이다.

2.2.3 SpamAssassin

SpamAssassin은 스팸을 인지하는 메일 필터이다. 이것은 허락받지 않고 적하된 메일, 즉 보통 스팸으로 더 알려진 메일을 인지하여 다양한 범위로 테스트할 수 있는 지능형의 메일 필터이다. 이들 테스트는 발전된 통계적인 기법을 사용하여 메일을 분류하여 헤드와 내용에 적용된다. SpamAssassin은 스팸에 대하여 빠르게 영향을 미치게 되는 다른 기술을 허용하는 모듈 구조를 가지고 있으며, 가상의 어떤 시스템으로 수월하게 통합하도록 설계되었다[4].

SpamAssassin은 메일에서 스팸이 가질 수 있는 요소를 분석하여, 각각의 요소에 대해 점수를 부여하고 이들을 모두 합하고 그 점수가 일정 이상이면 스팸으로 판정하는 방식이다. 메일 전체 부분에 대해 수행한 결과가 5이상이면 스팸으로 판별하며, 스팸 메일 판정 과정을 메일 헤더에 삽입하여 스팸메일로 판별된 메일에 대한 자세한 정보를 볼 수 있도록 하여 잘못된 룰셋 사용을 방지할 수 있다. 또한 SpamAssassin의 룰셋은 로컬네트워크에서 전송된 메일과 외부 네트워크에서 전송된 메일에 대해 각각 다른 값을 부여하여, 로컬네트워크에서 전송된 메일에 대해 스팸메일로 판별될 가능성을 줄이고 있다.

2.2.4 IP 기반

IP 기반의 스팸메일 차단 방법[5]는 메일 서비스 이용자가 메일 서버로부터 POP3로 메일을 가져올 때, ISP 사업자의 인터넷 접속 서버에서 인터넷 접속을 위한 인증 및 IP 부여 단계에서 인터넷 사용자에게 부여되는 IP 정보를 획득하고, ISP 사업자의 가정용 메일 차단 서버가 고객에게 할당된 IP를 기반으로 해당 IP로 유입되는 메일을 네트워크단에서 IP를 기반으로 스팸을 차단함으로써, 즉 고객단이나 서비스단에서 스팸을 제거하는 대신에 네트워크단에서 IP를 기반으로 스팸을 차단함으로써, 스팸성 또는 음란성 메일인 경우 이를 필터링한 후 고객단으로 전달하는 기능을 갖는다. 그러므로 스팸 대응에 미숙한 가정용 인터넷의 주 사용 층이나 불법 스팸에 노출 빈도가 많은 가정 구성원들에게 필터링 장비를 구비하여 사용하지 않고도 이들 스팸으로부터 보호하여 건전한 네트워크 환경을 제공한다.

2.2.5 이메일의 링크 구조분석

이 스팸메일 차단 시스템은 이메일 안에 존재하는 하이퍼링크가 가리키는 웹문서가 다른 임의의 웹문서에 의해 링크(인용)된 횟수를 측정한 후, 이메일 안에 존재하는 하이퍼링크의 개수와 하이퍼링크를 인용하는 다른 웹문서의 개수를 자질로 하여 기계학습을 한 후 스팸메일로 분류한다. 추출된 링크가 가르키는 웹문서를 링크하는 웹문서의 개수를 측정하기 위해서 페이지랭크 알고리즘을 사용하는 검색엔진 구글(Google)의 구글 웹 API를 사용한다. 또한 이메일의 하이퍼링크에서 호스트 URL 부분만을 자질로 하여 기계학습을 한 후 스팸메일로 분류한다. 더불어 이러한 이메일의 하이퍼링크 분석과 함께 내용기반 분석으로 기계학습을 한 후 이메일 내용 분석 결과를 통합하여 스팸메일 필터링 효율을 더 높인다[6].

2.2.6 패킷 모니터링

패킷 모니터링을 통하여 스팸의 송수신을 차단하는 방법[7]은 네트워크를 통하여 전송되는 메일 패킷을 캡처하여 메일 패킷으로부터 발신지 주소정보 및 수신자 주소 정보 등의 접속정보를 추출하여 메일 패킷이 스팸에 속하는지 판단하고, 스팸에 속하는 메일 패킷으로 판단될 경우에 수신 메일 서버와 발신 메일 서버에 메일의 송수신을 정지시키는 차단 패킷을 생성하여 전송한다.

이 방법은 임의의 도메인으로 수신되는 모든 메일을 캡처하여 차단할 수 있으며, 스팸의 수신뿐만 아니라 발신 자체도 방지할 수 있어서 메일 서버와 네트워크의 부하를 방지할 수 있고, 또한 스팸 차단 모듈이 정지되어 있어도 메일 발신측과 수신측 또는 네트워크의 흐름에 전혀 영향을 주지 않는 장점이 있다.

3. Link URL을 통한 스팸메일 자동 차단 방법

3.1 프로그램의 구성

서론에서 기술한 연구 목적을 달성하기 위하여, 다수의 메일서버가 인터넷을 통해 연결되어 이루어지는 전자메일 시스템에서 수행되는 링크 유알엘 접속을 통한 스팸메일 자동 차단 방법은 3 단계로 구분하여 구성할 수 있다. 첫째로 수신되는 전자메일의 메시지 원본에 존재하는 링크 유알엘 정보를 추출하는 단계, 둘째로 추출된 링크 유알엘 정보에 링크된 웹페이지에 접속을 수행하는 단계, 셋째로 웹페이지의 컨텐츠 중에 미리 규정된 스팸 키워드가 존재하는 경우에 수신 전자메일을 스팸메일로 분류하는 단계로 구성한다.

3.2 전자메일의 메시지 원본

통상의 전자메일 메시지 원본은 Header부와 Body부로 구분된다. 예를 들어 (그림 2)에서, Header부에는 전자메일을 바르게 배송하기 위해 필요한 전자메일 발신처(sda35s5g32-131@yahoo.com) 및 메시지 제목 등이 기재되어 있고, Body부에는 본문 내용(몸매도 예쁜게.....)과 광고하고자 하는 성인 사이트의 홈페이지의 URL(a href='http://www.ohmybuin.com/index.asp?ID=3956623') 및 링크되어 있는 복수개의 이미지 정보가 포함되어 있는 이미지 URL(IMG alt=OhMyBuin.Com height=85 src='http://www.ohmybuin.com/banners/title.gif' width=504)정보가 기재되어 있다.

```
Received: from yahoo.com ([211.106.168.154]) by mail.daraelaw.co.kr
with Microsoft SMTPSVC(5.0.2195.3779);
Mon, 27 Jan 2003 19:59:00 +0900
Reply-To: sda35s5g32-131@yahoo.com
Return-Path: sda35s5g32-131@yahoo.com
From: 바람난여자에요. <sda35s5g32-131@yahoo.com>
To: 0 <admin@daraelaw.co.kr>
Subject: 몸매가 참예쁘네요.....
Sender: 바람난여자에요.. <sda35s5g32-131@yahoo.com>
Mime-Version: 1.0
Content-Type: text/html; charset="ks_c_5601-1987"
Date: Mon, 27 Jan 2003 19:32:37 +0900
Message-ID: <MAILE8NpqhNPPzQT456000002e1@mail.daraelaw.co.kr>
X-OriginalArrivalTime: 27 Jan 2003 10:59:01.0089 (UTC)
FILETIME=[193C1D10:01C2C5F3]
<html>
<head>
<title> 몸매도 예쁜게.....</title>
<meta name="generator" content="Namo WebEditor v4.0">
</head>
<body bgcolor="white" text="black" link="blue" vlink="purple"
alink="red">
<a href="http://www.ohmybuin.com/index.asp?ID=3956623">
<IMG alt=OhMyBuin.Com height=85
src="http://www.ohmybuin.com/banners/title.gif"
width=504><IMG alt=OhMyBuin.Com height=85
src="http://www.ohmybuin.com/banners/title.gif"
width=504><IMG alt=OhMyBuin.Com height=85
src="http://www.ohmybuin.com/banners/title.gif"
width=504><IMG alt=OhMyBuin.Com height=85
src="http://www.ohmybuin.com/banners/title.gif" width=504></a>
<p> 몸매도 예쁜게...
<p></p>
</body>
</html>
```

(그림 2) 전자메시지 원본(예)

(그림 3)은 수신된 메시지 원본에 대한 수신된 전자메일의 팝업(Pop-up) 창을 예로 보여준다.

3.3 프로그램 설치 및 스팸 키워드

본 연구의 스팸메일 자동 차단 방법에 따른 프로그램(이하, '스팸차단 프로그램'라 한다.)은 내부 또는 외부 메일서버에 통합되거나 별개로 설치될 수 있고, 개인 메일 클라이언트의 경우에는 클라이언트 컴퓨터의 전자메일 소프트웨어에 통합되거나 별개로 설치될 수가 있음을 밝혀 둔다. 또한, '스팸 키워드(Keyword)'라 함은 스팸메일에서 자주 사용되는 두 개 이상의 글자의 집합을 말한다. 예컨대, 스팸 키워드는 공서양속

(그림 3) (그림 2)에 대한 전자메일(예)

을 해치는 성인용품 및 불건전 사이트 등의 광고메일에 자주 사용되는 '섹스(sex)', '오르가즘', '누드', '포르노', '야동', '변태' 또는 '자위' 등이 포함될 수 있다.

(그림 4)는 웹페이지(http://www.ohmybuin.com/index.asp?ID=3956623)의 컨텐츠 소스 중 일부를 발췌하여 나타낸 HTML 문서이다. 본 연구에서 제안하는 스팸차단 프로그램은 '변색', '오르가즘', '성인유머', '변태', '성인소설', '애널섹스', '성기구' 또는 '자위' 등의 키워드 내용이 있는 전자메일이 스팸메일로 용이하게 걸러질 수 있을 것이다.

```
<tr><td height="18"><img src="../img/square3.gif" wid
th="7" height="11" align="absmiddle">
<a href="../communi/board.asp?code=1">우리 변색 해
요</a><td>
<td width="26%><img src="../img/square4.gif" width=
"7" height="11" align="absmiddle">
<a href="../communi/board.asp?code=11">오르가즘 체
험단</a><td>
<td width="26%><img src="../img/square5.gif" width=
"7" height="11" align="absmiddle">
<a href="../communi/board.asp?code=21">우리사진 교
환해요</a><td>
<td width="22%><img src="../img/square7.gif" width=
"7" height="11" align="absmiddle">
<a href="../communi/board.asp?code=31">성인유머</
a><td></tr>
```

(그림 4) 스팸 키워드가 있는 HTML문서(예)

3.4 스팸차단 알고리즘

3.4.1 링크 유알엘 정보 추출

(그림 5)에 도시한 바와 같이, 먼저 임의의 전

자메일이 수신된 경우에 스팸차단 프로그램은 수신된 전자메일의 메시지 원본에서 링크되어 있는 웹페이지(이하, '링크 URL'이라 한다.) 및 전자메일의 발신처 이메일 주소(이하, '전자메일 발신처'라 한다.)를 추출한다.

다음, 상기 전자메일 발신처가 차단 발신처 리스트(데이터베이스)에 등재된 주소들인지의 여부를 조회하는데, 이를 위해 스팸차단 프로그램은 차단 발신처 리스트 데이터베이스를 운용한다.

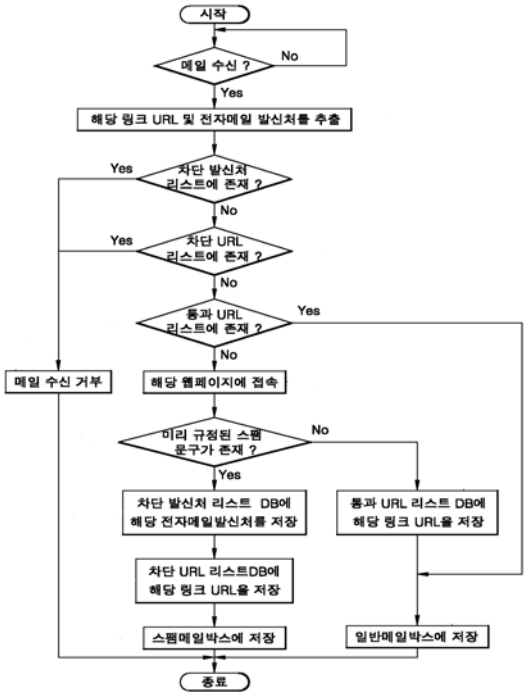
차단 발신처 리스트 데이터베이스에 등재된 주소인지 판단한 결과, 전자메일의 발신처가 차단 발신처 리스트에 등재되어 있는 경우에는 당해 전자메일의 수신을 거부, 예를 들어 본문의 내용을 읽지 않은 채로 바로 삭제 처리를 하게 된다. 이에 따라 메일서버(또는 클라이언트)가 당해 전자메일이 스팸메일인지의 여부를 확인하는데 따르는 부하를 경감시킬 수가 있다.

한편, 차단 발신처 리스트 데이터베이스에 등재된 주소인지 판단한 결과, 상기 전자메일의 발신처가 차단 발신처 리스트 데이터베이스에 등재되어 있지 않은 경우에는 링크 URL이 차단 URL 리스트 데이터베이스에 등재된 웹페이지 주소들인지의 여부를 조회하는데, 이를 위해 스팸차단 프로그램은 차단 URL 리스트 데이터베이스를 운용한다.

차단 URL 리스트 데이터베이스에 등재된 웹페이지 주소들인지의 여부를 조회하여 판단한 결과, 상기 링크 URL이 차단 URL 리스트에 등재되어 있는 경우에는 당해 전자메일의 수신을 거부하게 된다.

한편, 차단 URL 리스트 데이터베이스에 등재된 웹페이지 주소들인지의 여부를 조회하여 판단한 결과, 링크 URL이 차단 URL 리스트에 등재되어 있지 않은 경우에는 링크 URL이 통과 URL 리스트 데이터베이스에 등재된 웹페이지 주소들인지의 여부를 조회하는데, 이를 위해 스팸차단 프로그램은 통과 URL 리스트 데이터베이스를 운용한다.

URL 리스트 데이터베이스에 등재된 웹페이지 주소들인지의 여부를 조회하여 판단한 결과, 링크 URL이 통과 URL 리스트 데이터베이스에 등재되어 있는 경우에는 당해 전자메일을 일반 메일박스에 저장하게 된다.



(그림 5) 스팸차단 알고리즘

3.4.2 해당 웹페이지에 접속

반면에, 링크 URL이 통과 URL 리스트 데이터베이스에 등재되어 있지 않은 경우에는 링크 URL을 이용하여 당해 웹페이지에 접속을 수행하게 된다.

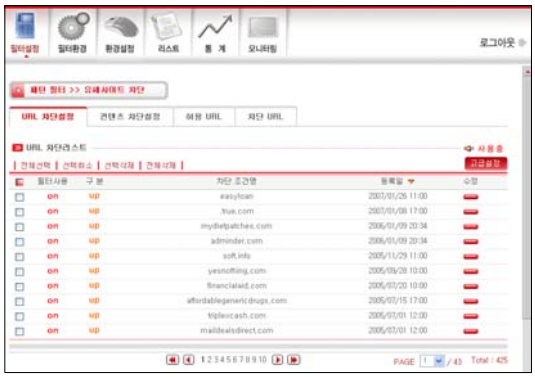
3.4.3 스팸메일 분류

접속된 웹페이지의 컨텐츠를 검색하여 자체적으로 운용하는 스팸 키워드 데이터베이스에 미리 규정된 대로의 스팸 키워드가 존재하는지의 여부를 판단한다. 판단 결과, 당해 접속된 웹페이지의 컨텐츠에 규정된 스팸 키워드가 사용되지 않은 경우에는 통과 URL 리스트 데이터베이스에 해당 전자메일의 링크 URL을 저장하게 된다. 한편, 당해 접속된 웹페이지의 컨텐츠에 규정된 스팸 키워드가 그대로 존재되고 있는 경우에는 당해 전자메일을 스팸메일로 추정하여 상기 차단 발신처 리스트 데이터베이스에 해당 전자메일의 발신처를 저장함과 동시에 차단 URL 리스트 데이터베이스에 해당 전자메일의 링크

URL을 저장하고, 당해 전자메일을 스팸메일박스에 저장한 다음 종료하게 된다.

3.5 프로그램 구현

3.4 절에서 설계한 알고리즘에 따라 C++언어로 개발하였으며 데이터베이스는 해쉬 데이터베이스를 사용하여 차단리스트나 스팸 키워드 등을 기록하였다. 또한 스팸차단 자동 프로그램은 윈도우즈(Windows), 솔라리스(Solaris), 리눅스(Linux) 등에서 수행되도록 구현하였다. (그림 6)은 필터설정 메뉴에서 패턴 필터를 유해사이트로 하고 URL 차단을 설정하는 화면이다.



(그림 6) 유해사이트 차단을 위한 설정

(그림 6)에서 필터설정, 필터환경, 환경설정, 리스트, 통계, 모니터링으로 구분하였다. 특히 필터설정 메뉴에서는 URL 차단설정, 컨텐츠 차단설정, 허용URL, 차단 URL로 나누어 각각 설정할 수 있다.

또한 통계 메뉴에서는 전체 통계, 기간별 통계 등을 볼 수 있다. (그림 7)은 실제로 C회사의 서버에서 2007년 5월 23일부터 2007년 6월 22일(1개월)간 모니터링한 결과로, 총 수신 차단 메일 2,276,229 건 중 링크 유알엘을 이용한 스팸메일 자동 차단 프로그램에 의해 115,251건이 차단되어 5.06 퍼센트가 필터링되는 것으로 나타났다. 그 밖의 다른 스팸메일 차단 프로그램들과 함께 필터링을 한다면 좋은 결과를 얻을 수 있을 것으로 기대된다.

구분	수신 차단 메일		송신 차단 메일		총 차단 메일	
	메일 수	비율	메일 수	비율	메일 수	비율
스팸처리 합계	2276229	99.999	22	0.001	2276251	100
헤더정보	1159	0.051	0	0	1159	0.051
용량제한	83	0.004	0	0	83	0.004
첨부파일	48905	2.148	0	0	48905	2.148
동일발신	0	0	0	0	0	0
패턴필터	1629905	71.609	22	0.001	1629927	71.609
유해사이트	115251	5.063	0	0	115251	5.063
바이러스	4917	0.216	22	0.001	4939	0.217
DNScan	1509817	66.329	0	0	1509817	66.329
메시지필터	536097	26.188	0	0	536097	26.188

(그림 7) 서버에서의 모니터링 결과(C 社)

4. 결론 및 향후 과제

이상에서 논의한 바와 같이 본 연구의 링크 유알엘 접속을 통한 스팸메일 자동 차단 방법에 의하면, 수신된 전자메일의 URL정보를 추출하여 해당 웹사이트에 접속한 후 키워드 검색을 통해 미리 규정된 소정의 스팸 키워드가 있는 경우 이를 스팸메일로 분류하여 자동으로 차단함으로써, 스팸메일의 차단 효율을 보다 증진시킬 수 있을 뿐만 아니라 기존 시스템에서 관리자가 수신차단목록 및 수신허용목록을 찾아내어 등록하는 과정을 거치지 않고 자동으로 처리 및 관리하여 효율적이고 신속할 수 있는 이점이 있다.

그러나 웹페이지의 컨텐츠에 규정된 스팸 키워드가 아닌 다른 키워드로 기망한 경우에는 차단하지 못하는 단점을 가지고 있어, 차후에는 웹 컨텐츠 분석을 이용한 네트워크 기반의 스팸메일 차단 장치에 관한 연구가 필요하다.

참 고 문 헌

[1] 이홍섭, 사이버 암세포 스팸과의 전쟁, 디지털타임스, 2005. 11. 29

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

[3] 백기영, 이철수, 류재철, "URL 빈도분석을 이용한 스팸메일 차단 방법", 정보보호학회 논문지, 제 14 권, 제 6 호, 2004

[4] Justin Mason, "SpamAssassin", http://wiki.apache.org/spamassassin/SpamAssassin

[5] 임영숙, 김영헌, “아이피-기반 스팸메일 차단시스템 및 그 방법”, 대한민국특허청 공개특허정보, 공개번호 10-2006-0026666, 2006.3.24.
[6] 이신영, “이메일의 링크구조분석을 통한 스팸메일 차단 시스템”, 대한민국특허청 공개특허정보, 공개번호 10-2005-0111566, 2005.11.25.
[7] 신미연, “패킷 모니터링을 통한 스팸메일 차단 방법 및시스템”, 대한민국특허청 공개특허정보, 등록번호 10-0525758, 2005.10.26.

정 남 철

1977년 : 경인교육대학교 졸업
1983년 : 광운대학교 전자계산학과
1987년 : 성균관대학교 정보처리학 (석사)
1996년 : 성균관대학교 대학원 전산통계학전공(박사)
1983년～1989년 : 수협중앙회 전자계산소
1989년～1992년 : 세계일보 전산정보국 과장
1992년～현 재 : 동남보건대학 웹컨텐츠과 부교수
관심분야 : 교육용콘텐츠, 인터넷방송, 멀티미디어 응용, 전자상거래시스템, 통계학습시스템