情報保護學會論文誌
第14券 第6號, 2004. 12

URL 빈도분석을 이용한 스팸메일 차단 방법

백 기 영[a)†], 이 철 수[b)], 류 재 철[a)‡]
충남대학교[a)], 경원대학교[b)], 충남대학교[a)]

A spam mail blocking methcd using URL frequency analysis

Ki-young Baek[a)†], Chul-soo Lee[b)], Jae-cheol Ryou[a)‡]
Chungnam University[a)], Kyungwon University[b)], Chungnam University[a)]

요 약

최근 다양하게 변하는 스팸메일은 단어에 의한 기존의 스팸메일 판별 방법으로는 차단하기 어렵다. 이와 같은 문제를 해결하고자 URL 빈도분석을 이용한 스팸메일 판별 규칙 생성 방법을 제안한다. 제안한 방법은 스팸메일을 수집하고, 수집된 스팸메일에서 특징이 되는 URL을 추출하고, 이를 정규화하여 시간 빈도에 따른 스팸메일 판별 규칙 생성하여 스팸메일을 차단하는 단계로 구성된다. 이는 다양한 스팸메일에 대응할 수 있으며, 변화하는 스팸메일의 형태에 대해서도 대응할 수 있는 구조를 가지고 있다.

ABSTRACT

Recently, it is difficult to block the spam mail that changes variously with past spam distinction method by words. To solve such problem, This paper propose the method of generating spam distinction rule using URL frequency analysis. It is consist of collecting spam, drawing URL that get into characteristic from collected spam mail, URL normalizing, generating spam distinction rule by time frequency, and blocking mail. It can effectively block various types of spam mail and various forms of spam mail that change.

Keywords : *Spam Mail Filtering, Spam Mail Blocking, Frequency Analysis*

Ⅰ. 서 론

스팸메일이 사회문제로 등장한 것은 불과 2-3년 사이의 일이지만, 사회 전체에 미치는 영향력은 그 어떤 정보화 역기능보다 크고도 넓다. 2002년 12월 네티즌 2000명을 대상으로 조사한 바에 따르면, 가장 피해를 입은 정보화 역기능으로 개인정보침해(32.7%)나 바이러스 피해(7.7%)보다 스팸메일(50.6%)을 꼽았다. 물론단일 피해 규모로는 개인정보침해나 바이러스 피해가 스팸메일로 인한 것보다 크겠지만, 스팸메일은 일상적이고 보다 대중적으로 발생하는 피해이기 때문이다.

가장 일반적인 스팸메일 차단 방법으로는 스팸메일의 제목이나 내용에 존재하는 특정 단어나 보낸 사람을 이용하여 차단하는 방법이 있으며, 대부분의 웹 메일이나 메일 클라이언트에서 이를 지원한다. 그러나 이와 같은 방법은 스팸메일이라고 규정지을 수 있는 단어나 보낸 사람에 대한 정의

접수일 : 2004년 11월 2일 ; 채택일 : 2004년 12월 6일
* 이 논문은 정보통신부 대학 IT 연구센터 육성지원사업에 의한 것임.
† 주저자 : cloud@cqcom.com
‡ 교신저자 : jcryou@home.cnu.ac.kr

가 쉽지 않고 계속적으로 변하는 스팸메일에는 적용하기가 힘든 단점이 있다.

이와 같은 문제점을 해결하고자 스팸메일 발송 서버의 IP를 신고하여 이를 이용하여 차단하는 RBL(Real-time Black-hole List)와 같은 방법이 있으나 이는 동적인 IP를 이용하여 스팸메일을 발송하는 스패머들에 의해 선의의 피해자가 발생할 수 있으며, 사용자가 신고한 스팸메일의 내용을 저장, 배포하여 차단하는 SpamNet과 같은 방법과 같은 경우 최근의 스팸메일은 내용에 랜덤한 문자열을 추가하여 발송하기 때문에 차단에 어려움이 있다.[1,2]

따라서 스팸메일의 최근 경향을 분석하고 이를 능동적으로 받아들여 스팸메일을 차단할 수 있는 스팸메일 차단 방법에 대한 필요성이 증가 되었으며, 이에 기존의 스팸메일 차단 방법의 단점을 보완하여 스팸메일을 수집하고 이들 스팸메일에서 특징이 되는 URL을 추출하고 스팸메일의 시간 빈도를 분석하여, 스팸메일 판별 규칙을 만들고 스팸메일을 검사하는 방법을 제안한다.

제안한 방법은 자동으로 스팸메일을 수집하여 이들에서 스팸메일의 특징이라고 할 수 있는 URL 부분을 추출하고 이를 정규화하여 시간 분포에 따른 스팸메일 판별 규칙을 생성하며, 메일 서버에서는 판별 규칙을 이용하여 스팸메일을 차단한다.

2장에서는 최근 스팸메일 동향에 대해 설명하고, 3장에서는 제안한 URL 빈도분석을 통한 스팸메일 판별 룰 생성에 관해 설명하며, 4장에서는 시험환경을 구축하고 시험을 통한 제안한 방법을 우수성을 검증하며, 5장에서 결론을 맺는다.

II. 관련 기술 동향

최근의 스팸메일 형태와 변형 방법, 그리고 스팸메일 차단 방법에 대해 알아본다.

1. 최근 스팸메일 동향

최근의 스팸메일은 스팸메일 차단 방법을 피하기 위해 다양한 형태로 변형되어 발송되며, 다음과 같은 형식을 갖는다.

1.1 제목을 변형하는 스팸메일 구성 유형

최근의 스팸메일은 다양한 형태를 취하고 있으며, 필터링과 같은 일반적인 스팸메일 차단 방법으로는 차단이 어렵다.[3]

1.1.1 일반적인 제목을 가진 스팸메일

사용자의 확인을 유도하기 위해 일반적인 제목을 갖는 메일의 형태로 발송된다.

◉ Re : That movie
◉ 한번 연락드리려니 쑥스럽네요. ^^;;
◉ 수진아 네가 찾던거 같아

또는 쇼핑몰에서 발송한 것과 같은 제목을 가진다.

◉ 요청한신 패스워드 입니다.
◉ 배송 목록입니다.

1.1.2 같은 제목에 랜덤한 문자열 추가

같은 제목에 의한 스팸메일 차단을 방지하기 위해 스팸메일 발송기는 제목이나 내용에 랜덤한 숫자나 알파벳을 추가하여 스팸메일을 작성한다. 이와 같은 경우에 제목이나 내용만 바뀌고 보낸 사람과 스팸메일 발송 서버는 같은 경우가 대부분이다.

◉ 추석선물 고민하지마세요 06007469
◉ 추석선물 고민하지마세요 .....adfd c

이와 같은 경우 MUA(Mail User Agent)에서 보여줄 수 있는 제목의 길이가 얼마 되지 않는다는 것을 이용하여 중간에 공백 또는 "..."을 추가하여 사용자에게 보여질 때는 똑같은 제목으로 보이게 작성한다

1.2 내용을 변형하는 스팸메일 구성 유형

1.2.1 무의미한 HTML 태그 추가

위는 제목을 변형한 스팸메일 작성방법이며, 다음에 설명하는 스팸메일은 내용을 변조하여 일반적인 스팸메일 차단 규칙에 적용되지 않도록 하는

스팸메일 작성방법이다.

현재 사용하는 대부분의 MUA는 HTML 형식의 메일을 볼 수 있는 기능을 지원하며, 스팸메일의 대부분은 HTML 형태를 취한다. HTML 형태의 메일에는 "〈"로 시작해서 "〉"로 끝나는 태그를 지원하며, 이를 이용하여 단어의 중간에 존재하지 않는 HTML 태그를 추가하여, 단어에 의한 스팸메일의 차단을 방지한다.

예를 들어 메일의 본문중에 credit card와 같은 단어가 들어갈 경우 다음과 같은 형태로 변형이 가능하다.

◉ credit 〈abc〉card
◉ c〈abc〉r〈dc〉edit ca〈sg〉rd

1.2.2 URL 변형

스팸메일에 존재하는 URL을 이용하여 스팸메일을 차단할 경우에 베이시안 필터링과 같은 단어의 빈도에 따른 스팸메일 검사 방법에 의해서도 차단될 수 있다.[4] 따라서 스팸메일을 제작하여 발송하는 스패머는 이를 피하기 위해 URL을 변형하여 사용자는 같은 사이트를 방문하거나 이미지를 보게 되더라도 각각의 스팸메일에 다른 형태의 URL이 존재하도록 하는 방법을 사용한다.

URL 변조 방법을 살펴보면 다음과 같다. 예를 들어 "linux.org"와 같은 URL을 변형하도록 한다.[5]

1) 사용자 이름과 패스워드 추가

URL은 다음과 같이 스킴, 사용자 이름, 패스워드, 호스트와 패스로 구성되어 있다.[6]

스킴	사용자이름	패스워드		호스트	패스
http ://	fred	: skdj	@	www.example.com	/ something.html

사용자가 URL을 선택했을 경우 브라우저는 URL에서 사용자 이름과 패스워드를 제거하고 호스트로 접속한다. 호스트로 접속한 후에 사용자 이름과 패스워드를 이용하여 인증을 수행하고 정해진 패스의 파일을 가져온다. 접속한 호스트에서 인증을 지원하지 않을 경우 사용자 이름과 패스워드는 무시되고 정해진 패스의 파일을 가져올 수 있다.

이와 같은 URL 구성을 이용하여 사용자 이름과 패스워드 부분에 의미 없는 긴 문자열을 넣어 호스트 이름 부분이 화면에 표시되지 않도록 한다.

2) 사용자 이름 변형

사용자 이름 부분에는 어떤 것을 넣어도 관계없으므로 특정 사이트의 URL을 넣어 다른 사이트처럼 보이도록 조작한다. 예를 들면 다음과 같이 변형 가능하다.

```
http://www.microsoft.com:windows@www.linux.org
```

위에서 "www.microsoft.com"은 호스트처럼 보이지만 실질적으로 사용자 이름을 나타내는 것으로 뒤의 "windows"는 패스워드를 나타낸다. 위의 URL을 사용자가 선택했을 경우 실질적으로 www.linux.org 사이트에 연결되나 사용자는 www.microsoft.com 사이트로 연결되고 있다고 착각할 수 있다.

3) 호스트 이름 제거

위의 URL에서 원래 사이트인 www.linux.org가 표시됨으로 이를 제거하여 좀 더 완벽하게 사용자들에게 microsoft 사이트처럼 보이도록 할 수 있다.

첫 번째로 www.linux.org의 IP 주소는 "198.182.196.56"이다. 따라서 다음과 같이 변형 가능하다.

```
http://www.microsoft.com:windows@198.182.196.56
```

두번째로 IP 주소를 변형한다. IP 주소의 각각의 숫자를 숫자를 나타내는 ASCII 코드 번호로 변형 가능하다. 예를 들어 1은 ASCII 코드 번호 49에 해당함으로 "&#49"로 변형 가능하며, 9는 ASCII 코드 번호 57에 해당함으로 "&#57"로 변형 가능하다. 따라서 위의 URL을 변형하면 다음과 같다.

http://www.microsoft.com:windows@&#49&#57&#56.182.196.56

이와 같은 식으로 변형하면 사용자는 URL 구성에서 패스부분이 인코딩 되었다고 생각하여 원래의 URL을 알수 없게 되고, microsoft 사이트처럼 착각하게 된다.

4) 호스트 이름 및 패스 변형

위에서 설명한 URL 변형 방법은 URL 자체를 변형하여 여러 가지 변형을 만들고 사용자가 알아보기 힘들게 만드는 것이며, 스팸메일 작성에서는 실질적으로 사용자가 알아보기 힘들게 만드는 것보다 많은 변형을 거치는 것이 스팸메일 차단을 방지할 수 있어 효과가 있다.

이와 같은 방법 중 하나로 같은 내용을 나타내는 여러 개의 URL을 만드는 방법이 있는데, "http:// www.dazzly77.com" URL은 다음과 같은 여러 가지 방법으로 표시할 수 있다.

http://www.dazzly77.com/ghr/
http://www.meaanf.com/v9.gif

"http://www.plarerd.com/v9.gif"와 같은 식의 표시 방법은 URL을 변형하여 표시하는 것이 아니라, 웹 서버에서 지원하는 버추얼호스팅과 같은 방법을 이용하여 같은 내용에 다른 호스트 이름을 적용하거나 같은 내용을 여러 개의 다른 이름으로 만들어 각각의 스팸메일에 다른 URL을 넣을 수 있도록 하는 방법이다.[7] 이는 단순히 URL을 변형하는 방법과는 달리 웹 서버에 이를 위한 설정이 필요하다.

5) 의미 없는 검색 부분 추가

웹 사이트를 나타내는 URL은 다음과 같이 호스트, 포트, 패스, 검색 부분으로 나누어진다.

http://〈호스트〉:〈포트〉/〈패스〉?〈검색부분〉

웹 서버는 실행 가능한 응용 프로그램을 포함할 수 있으며, 사용자는 이를 수행하여 동적으로 결과를 받아 볼 수 있다. 대표적으로 검색 사이트가 이런 응용 프로그램을 사용하고 있으며, 사용자가 원하는 인자를 응용 프로그램에 전달하기 위해 URL에 검색 부분을 이용한다. 스팸메일을 발송하는 스패머는 이를 이용하여 의미 없는 검색 부분을 추가하여 URL을 변형하며, 예를 들면 다음과 같다.

http://www.what-need.biz/img/kgel.gif?CJbgo135
http://www.what-need.biz/img/kgel.gif?CJbgo174

따라서 URL을 이용하여 스팸메일을 차단할 경우 URL을 정규화(normalize)하는 과정이 필요하다.

2. 최근 스팸메일 차단 방법

최근 다양해지는 스팸메일에 대응하기 위한 많은 방법들이 연구되고 있으며, 대표적인 스팸메일 차단 방법은 다음과 같은 것이 있다.

2.1 베이시안 필터링(Bayesian Filtering)

폴 그래험(Paul Graham)이 2002년 9월에 발표한 "A Plan for Spam"이란 글은 베이시안 필터링에 근거해서 개개인마다 다른 스팸판별 규칙을 생성하고 이를 이용하여 스팸을 차단 하는 방법에 대해 소개한다.[4]

베이시안 필터링 기술은 미래 상황의 추측할 수 있게 해 주는 18세기 토마스 베이시스(Thomas Bayes)의 "확률론"에 근거한 것으로, 요점은 다음과 같다.

전자문서 분류에 많이 사용되는 베이시안 분류 방법을 응용한 것으로 문서 내에 단어들을 대상으로 확률적인 방법을 적용하여 분류하기 때문에 특정 패턴에 따르지 않는 스팸메일을 걸러낼 수 있다. 이는 어떤 사람에게는 스팸이 될 수 있는 메시지가 다른 사람에게는 유용한 정보가 될 수 있다는 사실에 대해 '베이시안 스팸 필터'는 각 개인별 스팸 분류 기준에 대한 학습을 한다.

이러한 학습 과정을 통해 '베이시안 스팸 필터'는 시간이 지남에 따라 그 효율성이 배가되고, 일반적으로 99.8퍼센트의 차단률 및 0.05퍼센트의 오탐지율을 보인다.

2.2 SpamAssassin[8]

SpamAssassin은 메일에서 스팸이 가질 수 있는 요소를 분석하여, 각각의 요소에 대해 점수를 주어 모두 합산한 점수가 일정 이상이면 스팸으로 판별하는 방식이다. 점수를 주는 방식에서 스팸에서 제외될 수 있는 요소는 점수를 스팸으로 판별할 수 있는 점수는 + 점수를 주어, 일반 메일을 스팸메일로 판별할 가능성을 줄인다.

메일의 전체 부분에 대해 수행한 결과 값이 5 이상이면 스팸으로 판별하며, 표 1과 같이 스팸메일 판별 과정을 메일 헤더에 넣어서 스팸메일로 판별된 메일에 대한 자세한 정보를 볼 수 있어 잘못된 룰셋 사용을 방지할 수 있다.

또한 SpamAssassin의 룰셋은 로컬 네트워크에서 전송된 메일과 외부 네트워크에서 전송된 메일에 대해 각각 다른 값을 부여하여, 로컬 네트워크에서 전송된 메일에 대해 스팸메일로 판별될 가능성을 줄이고 있다.

SpamAssassin은 메일을 3가지 부분으로 나누어 검사하며 각각의 부분은 다음과 같다.

◉ **헤더 분석** : 메일의 헤더에는 기본적인 보낸 사람, 받는 사람, 제목과 같은 메일의 기본정보에서부터 MUA 고유의 정보를 나타내는 X 어플리케이션 헤더까지 많은 정보를 담고 있다. SpamAssassin에서는 이런 메일이 고유정보에 점수를 부여하고 있다.

예를 들면, 메일이 X-x 헤더를 포함하고 있으면 +4.300, From 형식이 something-offers 형식이면 +4.300, 보낸 사람의 메일 주소가 whitelist에 있으면 -100과 같은 값을 준다.

◉ **본문 분석** : 스팸메일은 독특한 본문 스타일을 가지고 있다. 예를 들면, 텍스트로 구성된 메일을 포함하지 않은 HTML로만 구성된 메일, HTML의 Heading 태그의 잦은 사용, 색이 들어간 글자의 사용과 같은 것들이다.

SpamAssassin에는 이러한 스팸메일의 특징을 룰로 만들어 검사한다.

◉ **블랙리스트**

표 1. SpamAssassin 예제
Table 1. The example of SpamAssassin

```
SPAM: ---------- Start SpamAssassin results ----------
SPAM: This mail is probably spam. The original message has ...
SPAM: so you can recognise or block similar unwanted mail in ...
SPAM: See http://spamassassin.org/tag/ for more details.
SPAM:
SPAM: Content analysis details:  (13.3 hits, 5 required)
SPAM: Hit! (4.3 points) Reply-To: is empty
SPAM: Hit! (0.8 points) Found an X-Em-Version: header
SPAM: Hit! (0.0 points) BODY: Includes a URL link to send ...
SPAM: Hit! (3.2 points) HTML-only mail, with no text version
SPAM: Hit! (1.9 points) Subject is all capitals
SPAM: Hit! (3.1 points) Subject is full of 8-bit characters
SPAM:
SPAM: ---------- End of SpamAssassin results --------
```

Ⅲ. 수집과 URL 추출을 이용한 스팸메일 판별 규칙 생성

최근의 스팸메일은 여러 가지 변형이 가능해 간단한 필터링으로 차단하기 힘들며, 스팸메일 차단 룰을 만들기가 쉽지 않다. 이에 따라 스팸메일의 변형에 대응하기 쉬운 수집과 URL의 빈도분석을 이용한 스팸메일 판별 규칙 생성에 대해 설명한다.

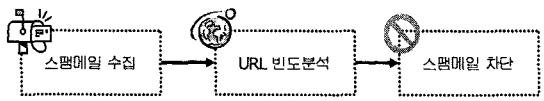


그림 1. 스팸메일 차단 방법
Fig 1. The method of blocking spam mail

그림 1은 스팸메일 차단 방법에 대한 전체적인 설명이며, 차단 단계는 스팸메일을 수집하고 수집한 스팸메일에서 URL을 추출하여 빈도분석 후 스팸메일 판별 규칙을 생성한다. 사용자가 이용하는 메일서버에서는 받은 메일에서 URL을 추출하고 이 URL이 판별 규칙에 존재하나를 검사하여 스팸메일 여부를 판별하여 스팸메일을 차단하게 된다. 다음에서 각각의 단계에 대한 자세한 설명을 한다.

1. 스팸메일 수집

계속적으로 변형되고 진화하는 스팸메일을 차단하기 위해서 본 논문에서는 스팸메일에서 스팸메일의 특징이라고 규정지을 수 있는 URL을 추출하고 이를 받은 메일과 비교하여 스팸메일 여부를 판별하게 된다. 따라서 URL 추출에 필요한 스팸메일이 필요하며, 다음과 같은 과정을 거쳐 스팸메일이 수집된다.

스팸메일의 수집은 실시간으로 이루어져야 하며 자동화되어야만 그 효과를 높일 수 있다. 스팸메일은 특정 시간에 발송되는 것이 아니며, 발송양도 많기 때문에 수작업으로 스팸메일을 수집하기는 어렵다.

스팸메일 수집의 기본 아이디어는 사용하지 않은 메일 주소에 전달된 메일을 살펴보면 대부분이 자신이 받고자 하지 않은 스팸메일이 주를 이룬다. 이는 가입할 때 사용한 메일 주소 또는 인터넷 상의 게시판이나 Usenet과 같은 뉴스 그룹에 포스팅 할 때 사용한 메일 주소를 스패머들이 수집하고 이를 스팸메일 발송시에 이용하기 때문에 원치 않는 스팸메일을 전달 받게 된다.

스팸메일 수집은 이런 스팸메일 발송 도구의 동작 방식을 이용한 것인데, 수집은 다음과 같은 과정을 거친다.

1) 사용하지 않은 가상의 메일 주소를 여러 개 생성한다.
2) 가상의 메일 주소를 이용하여 사람들이 많이 사용하는 게시판에 포스팅 하거나 메일 주소가 노출되는 뉴스 그룹에 포스팅 한다.
3) 메일 주소 수집기는 게시판 또는 뉴스 그룹에서 사용자의 메일 주소를 수집하고, 메일 주소에 스팸메일을 발송한다.[9]
4) 가상의 메일 주소로 도착하는 모든 메일은 스팸메일로 판단하고 저장한다.

스팸메일 수집에 대한 기본적인 아이디어는 메일 받기를 동의하지 않은 메일 주소로 도착한 모든 메일은 스팸메일로 판단하며, 이를 위해 가상의 사용하지 않은 메일 주소를 여러 개 생성하여 메일 주소가 노출되도록 한 후에, 스팸메일 발송을 유도하는 것이다.

2. URL 빈도 분석

전체 스팸메일 차단 시스템의 가장 중요한 단계인 URL 빈도 분석에 대해 알아보며, URL의 스팸메일의 특징을 나타내는 이유와 추출 과정, 빈도 분석 시에 필요한 가중치 값 결정에 대해 설명한다.

2.1 URL을 이용한 스팸메일 판별

스팸메일에서 URL을 스팸메일이 갖는 특징이라고 규정지을 수 있는 이유는 같은 상품에 대한 스팸메일을 분석해 보면, 변하지 않는 부분이 있는데, 이는 광고하고자 하는 상품에 대한 설명 및 상품 설명이 있는 웹 사이트 주소, 수신 거부를 위한 웹 사이트 주소 및 이메일 등이다.

최근의 대부분 스팸메일은 단지 텍스트만을 이용하여 광고하기 보다는 이미지 또는 상품이 설명되어 있는 URL을 첨부한 HTML 형태로 제작한다. HTML 형태의 스팸메일에서 변하지 않는 부분을 찾는 다면 다음과 같은 두가지 종류를 찾을 수 있다.

◉ 상품에 대한 이미지 및 URL
◉ 수신 거부를 위한 URL 및 이메일

상품에 대한 이미지 및 URL은 HTML의 "〈img〉" 태그 또는 "〈a〉" 태그를 사용하며, 수신 거부를 위한 URL 및 이메일은 "〈a〉" 태그와 메일 주소를 나타내는 "mailto:"를 이용하여 나타낸다.

"〈a〉" 태그는 HTML로 표시될 때 〈a href="URL"〉 형태로 구성되며, "〈img〉" 태그는 〈img src="URL"〉 형태로 구성된다. 따라서 이들 태그에서 URL 부분을 상품에 대한 정보로 인식하며, 이를 이용하여 스팸메일 판별 룰을 구성한다. "mailto:" 태그와 같은 경우 "mailto:이메일" 형태로 구성됨으로 "mailto:" 뒷부분의 이메일을

상품에 대한 정보를 인식한다.

2.2 URL 정규화

스팸메일을 발송하는 스팸머들은 변형된 URL을 사용하기 때문에 단순히 수집된 스팸메일에서 URL을 추출한다 하더라도 이를 스팸메일 차단에 이용하기 힘들다. 따라서 추출된 URL은 정규화 과정이 필요하며, URL을 정규화 하는 과정은 다음과 같다.

단계 1:
URL에서 quoted-printable과 같은 문자 인코딩 방법에 의해 인코딩된 문자를 디코딩 한다.
단계 2:
북마크와 사용자 아이디, 패스워드를 제거한다.
단계 3:
URL에서 호스트 이름을 IP로 변환하고 포트번호가 없을 경우 기본 포트 번호를 추가한다.

각각이 단계에 대해 설명하면, 단계 1에서는 URL에 인코딩된 문자를 디코딩한다. URL을 표시하는 데에는 quoted-printable을 비롯한 많은 인코딩 방법에 의해 인코딩 되어 있으며, URL에서 인코딩은 전송 중에 또는 다른 언어체제에서 URL을 표시하기 위한 방법으로 사용되었으나, 스팸메일을 발송하는 스패머들이 URL 변형의 수단으로 사용하기 때문에 디코딩이 필요하다.[9]

단계 2에서는 필요없는 사용자 아이디와, 패스워드 북마크 등을 제거한다. 단계 3에서는 같은 URL을 여러 개의 다른 호스트 이름으로 나타내기 때문에 이를 방지하기 위해서 호스트 이름을 IP 주소로 변환하며, 포트 번호가 없는 경우 기본 포트 번호를 추가한다.

"http://www.microsoft.com:windows@www.linux.org/#top"을 각 단계별로 디코딩 하는 과정은 다음과 같다.

1) http://www.microsoft.com:windows@www.linux.org/#top
2) http://www.linux.org
3) http://198.182.196.56:80

2.3 스팸메일에서 URL 추출

수집된 스팸메일에서 URL을 추출하여 스팸메일 차단 룰을 만드는 과정은 다음과 같이 2가지로 생각해 볼 수 있다.

1) 수집된 스팸메일에서 메일 단위로 URL을 추출하고 이의 해쉬값을 구해 각각의 받은 메일에서의 추출한 URL의 해쉬값과 비교한다.
2) 수집된 스팸메일에 존재하는 URL을 모두 추출하여 각각의 추출된 URL을 받은 메일에서 추출한 URL과 비교한다.

첫번째 방법은 수집된 스팸메일에서 추출한 URL을 메일 단위로 비교하는 것이고, 두 번째 방법은 각각의 URL을 따로 비교하는 것이다. URL은 여러 가지 변형이 가능하기 때문에 메일 단위로 URL을 비교한다면, 스팸메일을 차단할 확률이 적어질 수 있으므로, 스팸메일 차단 확률을 높이기 위해서 URL 단위로 비교하여 스팸메일을 판별한다.

스팸메일 차단에서는 받은 메일 전체 URL 중에 수집된 스팸메일에서 추출한 URL이 얼마나 차지하는 지에 따라 스팸메일로 판별하며, 이는 뒤에서 자세히 설명한다.

2.4 URL 빈도분석의 가중치 값 결정

추출한 URL을 빈도분석하여 판별규칙을 생성하는 과정은 그림 2와 같으며, 간단한 설명은 다음과 같다.

1) URL 추출 : 도착한 스팸메일에서 URL을 추출한다.
2) 중복된 URL 제거 : 하나의 스팸메일에서 중복되는 URL을 제거하여, 하나의 스팸메일이 특정 URL이 하나만 추출될 수 있도록 한다.
3) URL 정규화 : 변형된 URL을 정규화를 통해 일정한 형식으로 만든다.
4) URL 분할 : URL을 호스트와 패스 등으로 분할하여 각기 다른 가중치 값을 적용한다.

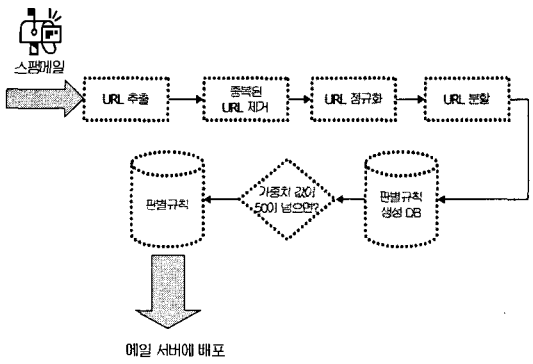


그림 2. URL 판별규칙 생성 과정
Fig 2. Steps of generating URL distinction rule

5) 판별규칙 생성 DB 추가 : 분할된 URL을 가중치 값을 적용하여 판별 규칙 생성 DB에 넣는다.
6) 판별 규칙 추가/삭제 : 판별 규칙 생성 DB의 값(Score)에 따라 판별 규칙에서 추가 또는 삭제한다.
7) 판별 규칙 배포 : 추가 또는 삭제된 판별 규칙을 메일서버에서 스팸메일 차단에 이용할 수 있도록 배포한다.

이와 같은 판별 규칙 생성 과정에서 있어서 각각의 과정이 필요한 이유와 가중치 값 결정의 근거는 다음에서 자세히 설명한다.

스팸메일에서 추출한 URL을 이용하여 스팸메일 판별 규칙을 생성하나, 추출된 URL을 모두 판별 규칙으로 적용할 경우 한번 판별 규칙에 들어간 URL은 계속 존재하게 됨으로, 판별 규칙은 점점 많아지게 되고 스팸메일 검사 시에 많은 시간이 소요된다.

따라서 추출한 URL을 판별 규칙에 추가하고 삭제하는 방법이 필요한데, 본 논문에서는 스팸메일에서 URL이 존재하는 빈도를 이용하여 이를 결정한다.

추출된 URL을 판별 규칙에 추가하기 위한 방법으로 하나의 스팸메일에서 여러 개의 같은 URL이 존재하는 가에 따라 판단하는 방법과 여러 개의 스팸메일에서 같은 URL이 존재하는 가에 따라 판단하는 방법이 있다. 그러나 하나의 스팸메일에서 여러 개의 같은 URL이 존재하는 것은 스팸메일 작성 방법에 따라 다름으로 여러 개의 스팸메일에서 같은 URL이 존재하는 가에 따라 판별 규칙 추가 여부를 결정한다.

일정 시간 내에 같은 URL을 가진 스팸메일이 많이 도착하면 판별 규칙에 추가하며, 도착하지 않으면 판별 규칙에서 삭제한다. 얼마만큼의 시간 내에 얼마만큼의 스팸메일이 도착하면 판별 규칙에 추가 또는 삭제 할 것인지에 대한 가중치 값을 결정하기 위해 SpamArchive에서 제공하는 2만개의 스팸메일을 분석하였다.[11]

SpamArchive는 스팸메일의 연구를 위해 스팸메일을 모아 저장/배포하며, 이는 스팸메일 방지 도구를 이용한 자동화된 방법과 사용자들이 수동으로 검색한 스팸메일을 전송함으로써 이루어지고 있다. 현재 하루에 5천개의 스팸메일이 저장되며, 지금까지 22만개 정도의 스팸메일이 저장되어 있다.

가중치 값을 결정하기 위해 같은 URL을 가진 스팸메일의 도착 시간 분포를 분석해야 되는데, SpamArchive에 저장된 스팸메일의 특성상 같은 URL을 가진 스팸메일을 분류하기 어려움으로, 같은 스팸메일인지를 판별하기 위한 방법으로 보낸 서버 주소를 이용하였다. 따라서 보낸 서버 주소에 의한 시간분포를 분석하여 가중치 값을 정하고 이를 URL을 이용한 판별 규칙 생성의 가중치 값으로 사용한다.

분석한 2만개의 스팸메일에서 보낸 서버에 대

표 2. 보낸 서버 빈도
Table 2. The frequency of sending servers

보낸 서버	빈도
10.2.202.xxx	615
216.65.3.xxx	110
204.200.197.xxx	90
205.158.62.xxx	74
216.65.3.xxx	65
216.65.64.xxx	59
205.158.62.xxx	56
205.158.62.xxx	41
66.14.165.xxx	23
68.171.98.xxx	20
68.19.91.xxx	16
67.15.24.xxx	15
68.218.161.xxx	13

표 3. 보낸 서버의 시간 간격에 따른 분포
Table 3. The distribution of sending servers according to time interval

시간	빈도												
	615	110	90	74	65	59	56	41	23	20	16	15	13
10	495	36	1	8	20	18	6	3	20	8	14	13	10
20	3	1	1	1	1	1	1	1	1	1	1	1	1
30	4	1	1	1	1	2	1	1	1	1	0	0	1
40	2	1	1	2	1	4	1	1	0	1	0	0	0
50	2	1	1	1	1	1	1	1	0	1	0	0	0
1h	3	1	1	1	1	1	1	1	0	1	0	0	0
6h	55	15	45	12	5	5	8	6	0	6	0	0	0
12h	22	9	7	12	6	6	6	6	0	0	0	0	0
18h	7	6	14	6	6	6	8	6	0	0	0	0	0
24h	0	0	0	0	0	0	0	0	0	0	0	0	0
그외	13	33	7	23	11	7	16	9	0	0	0	0	0

한 정보가 부족한 4천개의 스팸메일을 제외한 나머지 스팸메일에서 빈도가 높은 보낸서버 13개를 살펴보면 표 2와 같다.

위와 같은 빈도가 높은 보낸 서버에 대해 각각의 스팸메일을 받은 시간 사이의 간격을 조사하였으며, 간격은 1시간 이내는 10분 단위로 하며, 1시간 이후에는 6시간 간격으로 조사하였다.

표 3의 자료를 이용하여 판별규칙 생성의 가중치 값을 결정하며, 점수(Score)가 50이 넘으면 스팸메일 판별 규칙의 자료로 활용한다. 스팸메일 판별 규칙의 기준이 되는 점수인 50은 임의로 설정하였으며, 각각의 시간 간격 내에 몇 개의 스팸메일을 받을 경우에 판별 규칙으로 활용할 것인가를 이용하여 시간에 따른 가중치 값을 설정하였다.

서버 가중치 값의 설정은 스팸메일에 대한 피해를 줄이는 방향으로 정하였으며, 단기간 내에 많이 발송되는 것들에 대해서는 가중치 값을 높게 두어 빠른 대처를 할 수 있도록 하고, 상대적으로 빈도가 적은 것들에 대해서는 가중치 값을 낮게 두어 정확성을 높게 하였다.

표 3을 분석해 보면 대부분의 스팸메일은 같은 서버에서 10분 내에 발송된 다는 것을 알 수 있으며, 10분 내에서도 2~3분 사이에 발송되는 것들이 대부분이다. 따라서 10분 이내에 발송된 스팸메일일 경우에 각각의 가중치 값을 25로 두고, 2개를 받으면 점수가 50이 되어 바로 스팸메일 판별 규칙의 자료로 활용될 수 있도록 하여, 빠른 차단이 될 수 있도록 하였다.

그 외에는 6시간까지는 대부분 비슷한 분포를 보이나 10분 이내에 비해 반 정도의 빈도를 가지므로, 6시간 이내에 대해서는 5개의 스팸메일일 경우에 스팸메일 판별 규칙의 자료 활용될 수 있도록 하였다.

24시간 이내에는 6시간 이내보다 빈도가 낮으므로 가중치를 낮게 두었으며, 그 외의 자료를 분석해 보면, 대부분이 2~3일 이후에 발송된 것이므로, 2일 동안 업데이트가 없을 경우 판별규칙에서 제외함으로써 일시적으로 스팸메일 발송 서버로 이용된 경우에 피해가 없도록 하였다.

표 4. URL 가중치
Table 4. The weight of URL

마지막 수신 시간	가중치
10분 이내	+25
6시간 이내	+10
24시간 이내	+2
2일간 업데이트 없을 경우	판별규칙에서 제외

스팸메일에서 URL을 추출하여 이를 판별 규칙 생성을 위한 DB에 넣으며, 시간 분포에 따라 표 4와 같은 가중치 값을 부여한다. 하나의 스팸메일에서 중복된 URL 없이 URL에 넣으며, 값이 50이상일 경우 스팸메일 판별 규칙으로 결정하며, 메일서버에 배포한다.

2.5 URL을 이용한 빈도분석의 가중치 값 결정

수집된 스팸메일에서 URL을 추출하는 과정을 살펴보면 다음과 같다.

스팸메일을 발송하는 스패머들은 단어의 빈도에 따른 스팸메일 차단을 방지하기 위해 변형된 URL을 사용한다. 위에서 살펴본 것과 같이 변형된 URL은 URL 자체를 변형하여 정규화 과정을 통해 빈도 분석이 가능한 URL과 버추얼 호스트 및 같은 내용을 다른 URL로 표시함으로써 빈도 분석이 힘든 URL로 구분된다.

이는 해당 URL의 내용을 비교하여 같은 URL인가를 판별할 수 있는데, 이와 같은 경우

내용을 가져와 저장하는 데에 많은 시간이 필요하며, 내용 자체를 계속 변경하는 방법을 많이 사용함으로 URL 자체를 각각의 부분으로 분할하여 빈도분석을 수행한다.

웹 사이트를 나타내는 URL은 다음과 같이 호스트, 포트, 패스, 검색 부분으로 나누어진다.

http://<호스트>:<포트>/<패스>?<검색부분>

검색 부분은 웹 서버에서 수행하는 응용프로그램에 필요한 인자를 전달하는 부분으로 URL에 동적으로 실행 가능한 응용프로그램의 인자를 전달하는 데에 사용된다. 그러나 스패머가 스팸메일을 작성할 때 URL을 변형하기 위해 의미 없는 검색부분을 추가하는 방법을 사용함으로 URL을 이용한 빈도 분석의 정확성을 떨어뜨린다.

따라서 다음과 같이 구분하여 URL 빈도분석 시에 사용한다.

1) 호스트 이름과 패스, 검색 부분이 있는 경우

검색 부분이 있는 경우 다른 스팸메일에서 검색부분까지 일치하는 URL이 존재하면 표 4에서 설정한 가중치 값을 적용하며, 그 이외에는 같은 가중치 값을 적용할 경우에 스팸메일을 작성에 포털 사이트에 내용이나 이미지를 올리고 이를 이용할 경우 일반메일이 스팸메일로 편별될 가능성이 있기 때문에 낮은 가중치 값을 적용한다.

패스까지 일치하는 경우에 일반적으로 같은 URL이라고 판별할 수 있으므로 가중치 값의 2/3 값을 적용하며, 호스트가 일치할 경우 1/2 값을 적용한다.

◉ http://〈호스트〉:〈포트〉/〈패스〉?〈검색부분〉 - 가중치 값
◉ http://〈호스트〉:〈포트〉/〈패스〉 - 가중치 값의 2/3
◉ http://〈호스트〉:〈포트〉 - 가중치값의 1/2

2) 검색 부분이 없고 호스트 이름과 패스만 존재

검색 부분이 없고 호스트 이름과 패스만 존재하는 경우에는 패스까지 일치할 경우 URL이 정확히 일치한다고 판단하여 가중치 값을 적용하며, 패스가 없고 호스트와 포트만 일치하는 경우 위와 마찬가지로 가중치 값의 1/2 값을 적용한다.

◉ http://〈호스트〉:〈포트〉/〈패스〉 - 가중치 값
◉ http://〈호스트〉:〈포트〉 - 가중치 값의 1/2

3) 호스트 이름만 있는 경우

호스트 이름과 포트만 일치하는 경우 원래의 URL임으로 가중치 값을 적용하여, 포트가 다른 경우 다른 서버로 간주하여 가중치 값을 적용하지 않는다.

◉ http://〈호스트〉:〈포트〉 - 가중치 값

3. 스팸메일 차단

스팸메일 수집과 URL 빈도 분석을 수행하면 스팸메일을 판별 할 수 있는 판별 규칙이 생성된다. 일반적으로 수집 및 판별규칙 생성은 스팸메일 센터에서 수행 가능하며 생성된 판별 규칙을 사용자 메일 서버 또는 사용자에게 전달하게 된다.

전달된 판별 규칙을 이용하여 들어오는 메일에 대해 URL을 추출하고 정규화 한 후에 판별 규칙에 존재하는 가를 검사하여 존재하면 스팸메일로 판별하여 차단하게 된다. 차단 방법으로는 여러 가지가 있을 수 있는데, 스팸메일 판별된 메일을 삭제하는 방법과 일정기간 동안 보관하는 방법, 또는 존재하지 않는 사용자로 위장하여 리턴 메일을 보내는 방법들이 존재한다.

Ⅳ. 시험 결과

제안한 아이디어를 검증하기 위한 시험환경을 구축하였으며, 기존의 스팸메일 방지 시스템에 대한 시험 항목 및 결과에 대해 설명하면 다음과 같다.

1. 스팸메일 방지 시스템 시험 항목

PC Magazine에서는 2003년 2월에 몇 개의 제품을 이용하여 스팸메일 판별의 정확성을 측정하는 시험을 하였으며, 시험 항목은 다음과 같다.

◉ 일반메일인데 스팸메일로 처리한 비율(FPP

: False-positive Percentage) : 일반적으로 오탐율이라고 칭한다.

◉ 스팸메일인데 일반메일로 처리한 비율(FNP : False-negative Percentage)

◉ 전체 스팸메일 중에 스팸메일로 처리한 비율(TPP: True-positive Percentage)

◉ 전체 일반메일 중에 일반메일로 처리한 비율(TNP: True-negative Percentage)

2. 시험환경 구축

다음에서는 제안한 스팸메일 차단 방법의 검증을 위한 시험 환경 구축 및 시험 항목에 대해 설명한다.

2.1 시험환경

제안한 스팸메일 차단방법의 효율성을 검증하기 위해 그림 3과 같은 시험환경을 구축하였으며, 시간을 기준으로 수집된 스팸메일과 저장된 일반메일을 이용하여 제안한 스팸메일 차단 방법과 SpamAssassin의 스팸메일 차단방법의 효율성을 비교하였다.

각각의 구성요소에 대한 설명은 다음과 같다.

◉ 수집된 스팸메일

가상의 이메일 주소로 전송된 모든 메일은 스팸메일이며, 이를 전송받은 시간에 따라 데이터베이스에 저장하였다. 저장된 스팸메일은 어떠한 분석이나 가공도 하지 않았으며, 전송된 상태로 저장된다.

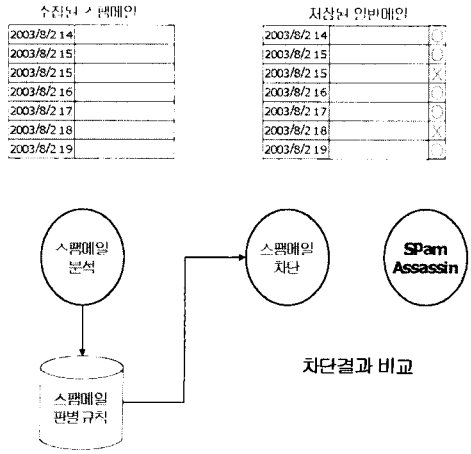


그림 3. 시험환경
Fig 3. The testing environment

◉ 저장된 일반메일

10명의 사용자에게 전송되는 메일을 전송된 시간별로 저장하며, 전송 받은 원본 그대로를 저장한다. 저장된 메일은 사용자가 판별하여 각각의 메일이 스팸메일인지 스팸메일이 아닌지에 관한 정보를 가지고 있으며, 스팸메일 차단 시에 이 자료를 이용하여 에러율을 계산한다.

◉ 스팸메일 분석

시뮬레이션 시간에 따라 주어진 스팸메일을 분석하여 스팸메일 판별규칙을 생성하고 이를 스팸메일 판별규칙 데이터베이스에 저장한다.

◉ 스팸메일 판별규칙

스팸메일 분석 결과로 나온 스팸메일 판별 규칙이며, 시뮬레이션 시간에 따라 계속 변한다.

판별규칙은 SpamAssassin과의 비교를 위해 URL 빈도분석 및 제목, 보낸 사람, 보낸 서버에 대한 빈도분석을 같이 수행한다.

◉ 스팸메일 차단

스팸메일 판별규칙을 이용하여 시뮬레이션 시간에 주어진 일반메일을 검사하여 스팸메일을 차단한다. 스팸메일 차단 결과 및 검사를 수행하는데 소요된 시간을 기록한다.

◉ SpamAssassin

SpamAssassin의 기본적인 설정을 이용하여 스팸메일 차단율을 검사한다.

2.2 시험 결과 비교 항목

스팸메일 검사 후에 각각의 검사 결과와 SpamAssassin의 수행 결과와 비교할 항목은 다음과 같다.

◉ 스팸메일 차단율

스팸메일 차단 검사에 사용된 일반메일은 미리 스팸메일인지 일반메일인지가 표시되어 있으며, 이를 이용하여 스팸메일 차단 후에 스팸메일로 제대로 판단했는지에 관한 정보를 기록한다.

◉ 스팸메일 오탐율(FPP)

일반메일을 스팸메일로 판단하는 확률을 비교

한다. 스팸메일 차단에서 스팸메일로 많이 판단하면 좋은 방법으로 생각할 수 있으나, 일반 메일을 스팸메일로 오판하는 결과도 중요한다. 따라서 일반메일을 스팸메일로 판단하는 확률을 비교한다.

◉ **일반메일 오탐율(FNP)**

스팸메일을 일반메일로 판단하는 확률을 비교한다.

3. 스팸메일 차단 방법의 시험 결과

위의 스팸메일 시험 환경에서 10명의 사용자에 대해 2주간의 메일을 이용하여 시험했으며, 빈도분석 자료는 2주간의 스팸메일 외에 1달전까지의 스팸메일을 추가로 이용하였다.

미리 작성된 스팸메일 판별 자료를 바탕으로 2주간의 메일을 시간에 따라 시뮬레이션 하였으며, 2주간의 메일은 사람이 판독하여 스팸메일과 일반메일을 구분하여 표시하였다.

표 5는 시험한 결과를 나타내며, 실제 일반 메일과 실제 스팸메일은 사람이 판독한 실제일반메일과 실제 스팸메일의 개수를 나타내며, 결과의 비교를 위해 같은 메일에 대해 SpamAssassin이 스팸메일로 차단한 결과를 같이 표시하였다.

표 5의 시험 결과에서 살펴보면 전체 메일의 약 50% 이상이 스팸메일로 판명되었으며, 일반메일을 스팸메일로 판별하는 오탐율은 4% 대로 비교적 낮은 편이며, SpamAssassin과 같은 경우에는 스팸메일 판별율은 높게 나왔지만 일반 메일 오탐율이 높게 나와 정형화된 필터에서 일반메일을 스팸메일로 판별하는 비율이 높은 것으로 나타났다.

V. 결 론

인터넷을 이용한 이메일은 이제 소수의 통신수단이 아닌 일반인이 널리 사용하는 기본적인 통신 수단으로 자리 잡고 있으며, 이에 따른 스팸메일의 피해 규모도 날로 커지고 있다. 현재 다양한 방법의 스팸메일 차단 방법이 제안되고 수행되고 있으나 다양해지는 스팸메일에 대응하기에는 역부족이다.

따라서 이 논문에서는 스팸메일을 수집하고

표 5. 시험 결과
Table 5. The test result

날짜	전체 메일	일반메일					스팸메일				
		실제 일반 메일	제안 방법		SpamAssassin		실제 스팸 메일	제안방법		SpamAssassin	
			판단 개수	오탐지 (FNP)	차단 개수	오탐지 (FNP)		판단 개수	오탐지 (FPP)	차단 개수	오탐지 (FPP)
4/11	313	129	142	7(4%)	118	10(5%)	175	171	3(2%)	187	6(5%)
4/12	484	223	227	8(3%)	204	13(5%)	261	257	4(2%)	280	7(3%)
4/13	377	171	173	5(2%)	154	7(3%)	206	204	3(2%)	223	6(4%)
4/14	365	199	200	4(2%)	178	5(3%)	166	165	3(2%)	187	6(3%)
4/15	496	232	238	10(4%)	213	15(5%)	264	258	4(2%)	283	9(4%)
4/16	436	206	211	8(4%)	190	14(6%)	230	225	3(1%)	246	6(3%)
4/17	357	154	157	6(3%)	141	9(7%)	203	200	3(2%)	216	5(4%)
4/18	105	54	56	2(4%)	50	2(4%)	51	49	0(0%)	55	0(0%)
4/19	366	185	190	8(5%)	171	11(6%)	181	176	3(2%)	195	5(3%)
4/20	518	235	244	12(4%)	219	19(6%)	283	274	3(1%)	299	7(3%)
4/21	466	221	229	11(5%)	205	16(6%)	245	237	3(1%)	261	7(3%)
4/22	273	145	146	3(2%)	129	3(2%)	128	127	2(1%)	144	5(4%)
4/23	169	90	93	5(7%)	93	6(7%)	79	76	2(2%)	86	2(2%)
합계	4725	2253	2306	89	2069	130	2472	2419	36	2662	71
비율		49%	49%	4%	44%	5%	52%	51%	2%	56%	3%

URL을 추출하여 이를 정규화하고 빈도 분석하여 스팸메일 차단에 사용되는 스팸메일 판별 규칙을 생성하는 방법으로 구성된 스팸메일 차단 방법을 제시했다. 스팸메일 수집은 가상 메일 주소를 만들어서 이를 통해 받는 모든 메일을 스팸메일로 규정하고, 이에서 URL을 추출하고 시간에 따른 가중치를 둔 빈도 분석을 통해 스팸메일 판별 자료를 만들어내며, 이를 사용자 메일 서버에 전달하여 들어오는 메일에서 추출한 URL과 비교하여 스팸메일을 판별하고 차단할 수 있도록 한다.

제안한 방법을 스팸메일 차단 솔루션인 SpamAssassin과 비교하여 시험하였으며, 시험결과를 살펴보면 기존의 스팸메일 방지 솔루션 보다 스팸메일 탐지율은 떨어지지만 일반메일 오탐율이 낮아 일반메일을 스팸메일로 차단할 가능성이 적으며, 좀 더 많은 스팸메일 데이터를 수집한다면 스팸메일 탐지율도 기존의 스팸메일 방지 솔루션 보다 높을 것으로 생각된다.

이와 같은 방법은 특정 단어를 인식하는 기존의 방법과는 달리 언어가 다른 스팸메일 시스템에도 적용 가능하며, 스팸메일을 수집하고 빈도 분석하는 센터를 두고 이를 메일 서버에서 스팸메일 차단에 이용함으로써 판별 규칙 생성에 따른 부하를 줄일 수 있다.

참 고 문 헌

[1] MAPS(Mail Abuse Prevention System), http://www.mail-abuse.com

[2] SpamNet, http://www.spamnet.org

[3] 마이크로소프트 아웃룩 익스프레스 필터 설정 방법, http://www.spamcop.or.kr/mbSpam/emai_outexpress.jsp

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

[5] "URL Obfuscation" and how it works, http://www.n3dst4.com/articles/urlobfus

[6] RFC1738, "Uniform Resource Locators (URL)"

[7] "Apache HTTP Server Documentation", http://httpd.apache.org/docs-2.1

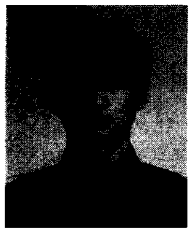
[8] SpamAssassin, http://www.spamassassin.org

[9] Email Spider Easy, http://www.email-tool.com/features.html

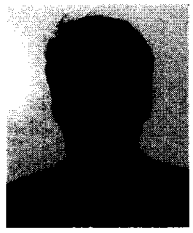
[10] "Coded Character Set--7-bit American Standard Code for Information Interchange", ANSI X3.4-1986.

[10] SpamArchive provides a database of known spam, http://www.spamarchive.org

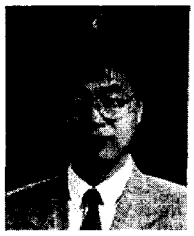
〈著者紹介〉



백 기 영 (Ki-young Baek) 정회원
1996년 2월 : 충남대학교 컴퓨터과학과 졸업
1998년 2월 : 충남대학교 컴퓨터과학과 석사
1998년 3월~현재 : 충남대학교 컴퓨터과학과 박사과정
〈관심분야〉 정보보호, 스팸메일



이 철 수 (Chul-soo Lee) 정회원
1977년 : KAIST 전산과 석사
1981년 : KAIST 전산과 박사
1982년~1993년 : (주) 데이콤
1993년~1998년 : 한국전산원
1999년~2000년 : 한국 정보보호원
2000년~2002년 : 정보통신대학교
2003년~현재 : 경원대학교
〈관심분야〉 정보보호 정책, 공개키기반구조, 침해사고 대응 기술



류 재 철 (Jae-cheol Ryou) 정회원
1985년 2월 : 한양대학교 산업공학과 졸업(학사)
1988년 2월 : Iowa State University 전산학과(석사)
1990년 2월 : Northwestern University 전산학과(박사)
1991년 3월~현재 : 충남대학교 컴퓨터과학과 부교수
〈관심분야〉 컴퓨터 및 통신망 보안, 전자상거래, 분산