정크메일 차단을 위한 FQDN 확인 시스템의 구현 및 평가

김 성 찬[†] · 이 상 훈[††] · 전 문 석[†††]

요 약

인터넷 사용의 급격한 증가로 전자우편은 모든 분야에서 가장 보편적인 통신 수단이 되었다. 하지만 전자우편의 사용 급증으로 사용자들의 전자우편 주소가 인터넷상에 노출되고 그 부작용으로 정크 메일, 스팸 메일이라 불리는 수신을 원하지 않는 메일의 수신빈도와 그로 인한 피해가 갈수록 높아져 그 문제가 심각한 수준에 이르게 되었다. 더구나 근래의 스팸, 정크 메일은 단순히 광고성 메시지를 전달하기 보다는 시스템을 공격하기 위한 바이러스나 해킹 도구를 전파하는 수단으로 이용 되어 컴퓨터 침해 사고의 심각한 원인으로 지적되고 있다. 따라서 본 논문에서는 이러한 스팸, 정크 메일을 FQDN 확인을 통해 차단할 수 있는 모델을 구현해서 사용해 보고 그 결과를 평가하여 개선 방향을 제시 하였다.

키워드 : 정크메일, 차단시스템, 스팸메일

An Implementation and Evaluation of FQDN Check System to Filter Junk Mail

Sung-Chan Kim[†] · Sang-Hun Lee[††] · Moon-Seog Jun[†††]

ABSTRACT

Internet mail has become a common communication method around the world because of tremendous Internet service usage increment. In other respect, Most Internet users' mail addresses are exposed to spammer, and the damage of Junk mail is growing bigger and bigger. These days, Junk mail delivery problem is becoming more serious, because this is used for an attack or propagation scheme of malicious code. It's a most dangerous dominant cause for computer system accident. This paper shows the Junk mail filtering model and implementation which is based on FQDN (Fully Qualified Domain Name) check and evaluates it for proposing advanced scheme against Junk mail.

Key Words : FQDN, Junk Mail, Spam Mail, Filter System

1. 서 론

오늘날 전자 우편은 가장 보편적인 통신 수단으로 이용되고 있다. 모든 비즈니스에서 일상적인 통신수단으로 빠르고 정확한 전자우편의 효용성에 대한 인식이 급격하게 확산되고 있고, 전 세계적으로 퍼져있는 업무 파트너나 고객과의 접촉에서 전자우편이 제공하는 저렴하고 편한 방법에 그 사용 또한 급격하게 늘어나고 있지만, 이로 인한 부작용 또한 심각한 문제가 되고 있다. 그 문제 중의 하나로 우리가 "스팸" 혹은 "정크" 메일이라 부르는 것이다. 인터넷 메일 사용자들에게는 수신을 원하지 않는 메일을 다량으로 받게 함으로써 사용자들의 시간을 낭비하고 혼란을 주며, 다량의 메일 메시지로 인한 네트워크의 대역폭 손실도 문제가 되고 있다. 근래의 연구 보고에 따르면 네트워크에 유입되는 수신 메일의 데이터 크기가 전체 네트워크 통신 트래픽의 10%정도를 차지한다고 한다[1]. "정크" 메일의 수신 문제는 단순히 원하지 않는 메일을 수신한다는 문제의 차원을 넘어서, 사용자들이 인터넷 메일 사용의 효용성을 크게 떨어뜨리고, 컴퓨터 시스템에 문제를 발생시키는 악성코드의 전달 매개체로서 그 문제가 심각해지고 있다. 본 논문에서는 현재의 "스팸" 혹은 "정크" 메일의 패턴을 연구하고, 이에 대응할 수 있는 시스템을 구현 한다. 그리고 일정기간 실제 사용을 통한 결과를 분석하여 평가하고 문제가 되는 부분에 대한 향상된 대응방안을 제시하고자 한다. 본 논문의 구성은 다음과 같다. 2장에서는 본 연구의 기초가 되는 관련연구에 대하여 기술하고, 3장에서는 본 논문에서 분석한 정크 메일 패턴과 구현한 정크 메일 차단 방법에 대하여 기술한다. 4장에서는 구현된 시스템에 대한 사

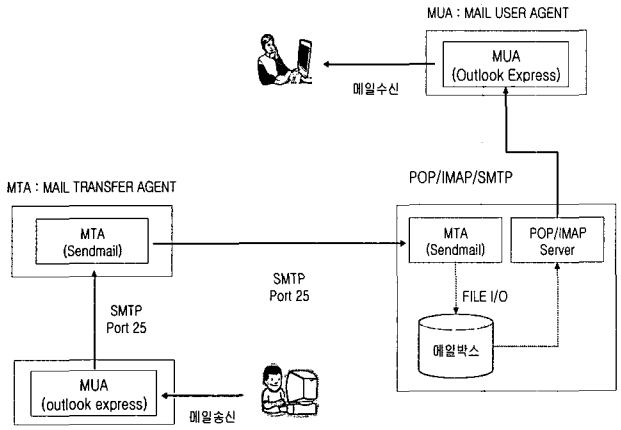
※본 연구는 숭실대학교 교내 연구 지원비에 의해 수행되었음.
† 정 회 원 : (주)유코레일 정보시스템부 과장
†† 준 회 원 : 숭실대학교 컴퓨터학과 박사과정
††† 종신회원 : 숭실대학교 컴퓨터학과 교수
논문접수 : 2004년 11월 26일, 심사완료 : 2005년 3월 21일

용결과를 분석 평가한 후, 문제가 된 부분에 대한 향상된 방안을 제시하고 5장에서 결론을 맺는다.

2. 관련 연구

2.1 전자우편 송수신 프로세스

수신을 원하지 않는 메일을 효과적으로 차단하는 기술을 개발하기 위해서는 메일을 주고받는 절차에 관한 기본적인 이해가 선행되어야 한다. 일반적으로 '전자우편을 주고받는다."라고 하는 것은 크게 (그림 1)과 같은 내부적인 프로세스에 의해 이루어진다.



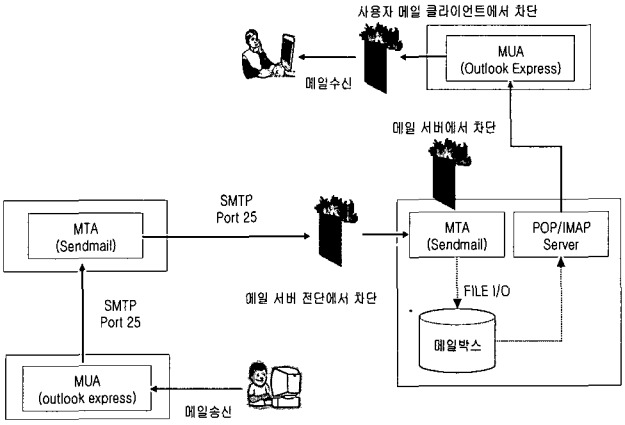
(그림 1) 전자우편 전달 절차

MUA(Mail User Agent)는 사용자가 전자 우편을 송수신할 때 사용하는 클라이언트 프로그램으로 일반적으로 사용자용 전자우편 프로그램을 말한다. MTA(Mail Transfer Agent)는 MUA로부터 전달 받은 메일을 목적지까지 전달하는 서버를 말한다[2]. 전자우편 발송 측 MUA는 사용자가 작성한 전자우편을 이용해서 전자우편 데이터를 생성하고 MTA의 25번 포트에 접속해서 전자우편 수신 측 정보 및 전자우편 데이터를 전송한다. 이때 MUA에서 MTA로 수신 측 정보 및 전자우편 데이터를 보낼 때 사용하는 프로토콜이 SMTP(Simple Mail Transfer Protocol)이다. MTA는 MUA로부터 받은 수신 측 정보를 이용해서 DNS로부터 수신 측 메일 서버의 IP를 얻어 전자우편 데이터를 수신 측 MTA에게 송신한다. 수신 측 MTA는 송신 측 MTA로부터 수신자의 정보 및 전자우편 데이터를 받은 후에 사용자 전자우편함에 전자우편을 저장하게 되는 것이다[3].

전자우편함에 저장된 사용자 메일은 MUA에서 POP3, IMAP와 같은 프로토콜에 의해 사용자 컴퓨터까지 메일을 전달하게 된다. 따라서 이와 같은 전자우편 전달 프로세스에서 가장 중요한 것은 전달자 상호간의 메일서버가 어떤 것인지 판단하는 과정이 있게 된다. 이때 DNS를 이용하는데, DNS는 호스트 명에 해당하는 IP를 알려주는 역할도 하지만 특정 도메인에 해당하는 메일 서버를 알려주는 역할도 한다[4].

2.2 기존의 정크 메일 차단 방법

정크 메일을 완전히 차단하는 방법을 찾아내는 것은 불가능하다. 왜냐하면 정크 메일의 정의 자체가 "수신자가 받기를 원하지 않고 요청하지 않았음에도 전송되는 메일"[5]이라는 메일 사용자들의 "정서적인 정의"이기 때문에 정량적인 기준으로 구현된 시스템에서 수신자가 받기를 원하는 메일과 받기를 원하지 않는 메일을 판단하는 것이 불가능하기 때문이다. 정크 메일 차단 방법은 정크 메일에 대한 대응 시점에 따라 받는 메일 서버 전단에서 차단하거나 받는 메일 서버에서 차단하는 방법, 사용자 메일 클라이언트에서 차단하는 방법이 있다[6].



(그림 2) 정크 메일 대응시점에 따른 차단방법

정크 메일의 문제가 심각해지자 전자 메일 서비스 제공업체 및 메일 클라이언트 프로그램에서 정크 메일 차단 기법을 제공하고 있는데, 정크 메일을 차단하는 방법은 <표 1>과 같이 요약할 수 있다[7]. 정크 메일 차단 방법은 크게 두 가지로 구분 할 수 있는데, 첫 번째는 등록된 룰에 기초한 필터링 방법이고, 두 번째는 룰에 등록된 주소를 수신 거부함으로서 차단하는 방법이다.

〈표 1〉 정크 메일 차단 방법

분 류	차 단 기 술
주소 거부	정크 메일을 보내는 서버 혹은 클라이언트의 IP주소 차단 알려진 도메인에 대한 메일 송신 서버 확인
필터링	제목 또는 내용에 특정단어가 있으면 정크 메일로 처리 정크 메일 규칙을 만들어 규칙에 부합한 메일은 정크 메일로 처리
기타 방법	정크 메일이 가질 수 있는 특성에 점수를 주어 정크 메일 분류 수집된 정크 메일 정보를 이용하여 정크 메일 차단 전자서명

등록된 룰에 기초한 필터링 방법은 사용자가 등록한 특정 단어를 제목이나 본문에서 포함하고 있는 전자메일에 대해 특정 방법으로 처리하는 기법이다. 이를 위해 수신자 혹은 수신 메일 서버는 수신하기를 원하지 않는 단어와 분류된 전자 메일에 대한 처리 방법을 설정하여야 한다. 차단될 메

일에 대한 처리방법으로는 특정 폴더로 분류하거나 복사 또는 완전히 삭제하는 방법이 있다[8]. 등록된 주소를 수신 거부함으로써 정크 메일을 차단하는 방법은 세부적으로 두 가지 모델로 구분할 수 있는데, 첫째는 수신을 거부하고자 하는 발신자 전자메일 주소나 도메인 주소를 수신 메일 서버 혹은 수신자 메일 클라이언트에 등록하여 정크 메일을 차단하는 방법이고, 둘째는 수신하고자 하는 발신자의 전자메일 주소나 도메인 주소를 메일 수신서버나 메일 클라이언트에 등록하여 등록된 주소 이외의 전자메일은 모두 삭제하거나 차단하는 모델이 있다. 이를 위해서는 수신자는 발신자의 주소를 등록하고, 분류되는 메일을 위한 처리방법을 설정하여야 한다[9]. 이 외에도 정크 메일이 가질 수 있는 특성에 가산 점을 부여하여 점수에 의해 정크 메일을 분류 및 차단하는 방법과 수집된 정크 메일 정보를 이용하여 차단하는 방법, 전자서명을 이용한 차단 방법이 있다. 이러한 대부분의 정크 메일 분류 방법들은 수신자가 입력하는 룰이나 나이브 베이지안 알고리즘[10]과 같이 이미 알고 있는 사전 지식에 기초한 확률적 분포로써 정크 메일을 분류 하는 등의 룰 베이스(Rule Base) 패턴 매치 기반의 컨텍스트 필터링(Context Filtering) 분류 방법 이다. 이러한 방법들은 정크 메일의 증가에 따른 관리자의 정보입력 부담이 커지고 등록되지 않은 송신자나, 제목 및 내용에 비중이 높지 않은 정크 메일은 차단에 어려움이 있으며, 정상적인 메일 또한 차단될 가능성이 높다는 문제점이 있다.

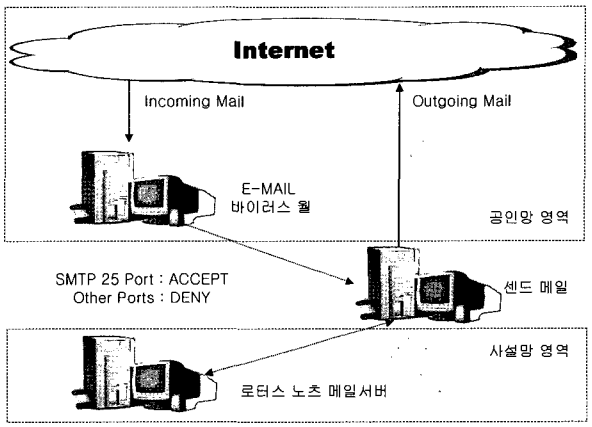
3. 정크 메일 차단을 위한 FQDN 확인 시스템

본 연구는 관련 연구에서 언급된 일반적으로 가장 많이 정크 메일 차단 방법으로 사용되는 룰 베이스 패턴 매치 기반의 차단 방법이 과연 실제 상황에서 얼마나 효율적인지에 대한 의문에서 시작 되었다. 따라서 현재 발생하고 있는 정크 메일이 어떤 특징을 갖고 있는지 파악하는 것이 중요하며, 파악된 특징에 가장 확실하게 대응할 수 있는 방법을 본 연구에서 제안하였다.

3.1 정크 메일 분석

본 연구에서는 정크 메일을 분류 할 수 있는 가장 주된 특징을 파악하기 위해 연구자가 소속된 회사의 메일 시스템에 기록된 메일 로그와 메일 사용자에게서 신고된 정크 메일을 분석 하였다. 본 연구에 사용된 메일 시스템 연구 환경은 (그림 3)과 같다.

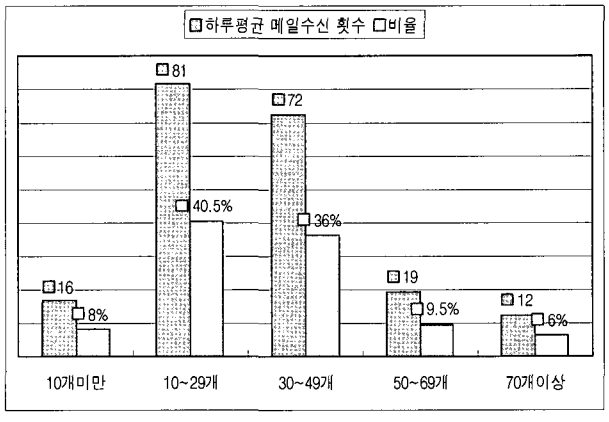
실험환경의 메일 시스템은 시스템을 보호하기 위하여 사설 망 안쪽에 위치하며, 인터넷 망으로의 메일 송수신은 (그림 3)에서와 같이 화살표의 방향으로 전달되고 발송된다. 인터넷 망으로부터 메일서버로 전달되는 메일은 우선 E-Mail 바이러스 월을 거치면서 메일에 바이러스가 포함되어 있는지를 검사 받게 되고, 사설 망 안쪽의 메일 서버로 메일을 전달해 주는 두 개의 네트워크 인터페이스를 갖고 있는 리눅스 기반의 센드메일로 메일이 포워드 된다. 리눅스 기반



(그림 3) 메일로그 수집에 사용된 메일 송수신 시스템 환경

의 센드메일 서버는 외부 인터넷 망에서 내부 사설 망으로 전달되는 메일과 내부 사설 망에서 외부 인터넷 망으로 발송되는 메일을 전달하는 기능을 담당하고 있으며, 리눅스에서 제공하는 iptable 패킷 필터링을 이용하여 E-Mail 바이러스 월과 센드메일 서버, 센드메일 서버와 로터스 노츠 메일 서버간에 25번 포트를 이용한 통신만 허용 되도록 설정되어 있다. 이때 E-Mail 바이러스 월과 센드메일 서버간의 통신은 단 방향 통신이고, 센드메일 서버에서 로터스 노츠 메일 서버간의 통신은 25번 포트를 이용한 양 방향 통신으로 설정 되어 있다. (그림 3)과 같이 실제 서비스 메일 서버를 사설 망에 두고, 수신 메일 서버와 발송 메일 서버를 이원화 하여 메일 패킷을 정해진 보안 장비를 경유해서만 통과하도록 흐름을 통제하여 운영함으로써 메일 릴레이에 의한 공격을 원천적으로 차단하고 메일에 편승해서 들어오는 악성 코드 및 바이러스로부터 메일 서버를 보호 할 수 있다.

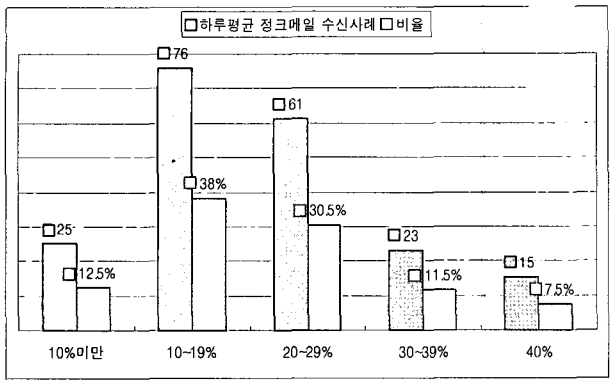
실험 환경의 메일 시스템에 정크 메일을 차단하기 위한 방법을 추가하기 위하여, 3개월의 조사기간 동안 수신 된 메일에 대한 로그와 메일 사용자에서 신고된 정크 메일의 로그 정보를 분석하여 몇 가지 중요 특징을 파악하였다.



(그림 4) 하루 평균 수신되는 전자메일 수

(그림 4)는 하루 평균 수신되는 메일의 수를 전체 메일 사용자수에 대비해 비율로 표시한 그림이다. 하루 평균 10

개에서 50개 정도의 메일을 수신하는 사용자가 전체 메일 사용자중 약 71%정도를 차지하고 있음을 알 수 있다.



(그림 5) 전제 수신 메일 중 정크메일 수신비율

(그림 5)는 전체 수신 메일 수중 정크 메일이 차지하는 비율을 나타낸 그림으로 정크 메일의 수신 비율을 수식 (1)에 의하여 계산하였다[5].

정크메일의 수신 율=(1계정당 수신되는 정크 메일 수 /1계정당 수신되는 전체 전자 메일 수)×100 (1)

정크 메일 수신 비율 조사 결과 수신되는 메일 중 약 10%~30%가 정크 메일인 사용자가 전체 메일 사용자중 약 61% 정도임을 알 수 있다. 실험 환경의 메일시스템에 메일이 도착하게 되면, 먼저 바이러스 월에 메일이 수신되고 어디서 발송된 메일인지 메일로그를 남기게 된다. 조사기간 동안 수신된 정크 메일의 특성을 결정하기 위하여, 수신된 정크 메일의 헤더를 분석하여 보았다.

```
from mx15.mx.voyager.net (mx15.mx.voyager.net [216.93.66.102])
with ESMTP id iA10WZI03220 ; Tue, 1 Jun 2004 09:32:36 +0900"

Field Name: Received
Data Type: Text List
Data Length: 119 bytes
Seq Num: 1
Dup Item ID: 3
Field Flags:
```

(그림 6) 정크메일 헤더 분석

(그림 6)은 정크 메일의 헤더를 분석한 결과다. 메일 헤더를 분석해 보면 먼저 눈에 띄는 정보가 메일을 어디에서 보냈는가 하는 "From" 구문인데, 조사기간 동안에 수신된 정크 메일의 "From" 구문에서 추출한 도메인과 IP주소를 분석한 결과 정상적으로 도메인 관리 기관에 등록된 도메인으로부터 발송되는 메일은 하나도 없었다. (그림 6)의 예에서 메일 발신 도메인은 "mx15.mx.voyager.net"이고 이 도메인의 IP 주소는 "216.93.66.102" 이지만, 실제로 도메인 관리기관에 도메인 검색 명령어인 "dig"나 "nslookup"으로 "216.93.66.102"의 IP주소를 검색해 보면, 메일 발신지의 도메인 정보나 도메인의 메일 교환 레코드 정보인 "MX" 레코드 정보를 찾을 수 없다. 이는 정크 메일의 정의인 "본인이 원하지도 않고 요청하지도 않았음에도 전송되는 송신자가 불명확한 메일"[5]과도 일치하는 것으로, 정크 메일을 보내는 송신자는 대량의 메일을 인터넷상의 불특정 공인 IP에 대용량의 수신자 정보를 갖고 있는 데이터베이스를 이용한 발신전용 메일서버를 이용하여 메일을 발송하기 때문에 발생하는 현상이다. 대부분의 정크 메일 차단 방법에서는 특정 IP에서 발생하는 메일 트래픽이 어느 기준이상 발생하면, IP 차단 리스트에 등록시켜 메일 수신을 거부하거나 아니면 관리자가 직접 차단 리스트에 입력하여 정크 메일을 차단 하고 있다. 하지만 유동 IP를 이용하여 대량의 정크 메일을 보내면서 지속적으로 IP를 바꾸게 되면, IP 추적을 통한 차단은 사실상 무용지물이 된다. 본 논문에서는 정크 메일의 발신지 정보의 IP 주소가 정식 도메인으로서 역할을 하지 않고 도메인 영역의 메일 교환 레코드인 "MX"레코드를 갖고 있지 않다는 것에 착안하여 메일이 인터넷 영역에서 도착하게 되면, 메일의 헤더부분에 포함된 발신자의 IP 주소를 역으로 검색하여, FQDN(Fully Qualified Domain Name)인지 아닌지를 판별한 후, 정상적인 도메인에서 보내는 메일로 판별된 경우 메일 전송을 허락하고 그렇지 않은 경우 송신자에게 반송시켜 송신자 등록 절차를 거쳐 메일 전송을 허가 받도록 시스템을 구성하여 정크 메일 차단 시스템을 구현 하였다[11], [12].

3.2 FQDN 확인을 이용한 정크메일 차단

본 논문에서 구현한 FQDN 확인을 이용한 정크메일의 차단 모듈은 (그림 7)과 같다. 먼저 인터넷 영역에서 메일이 도달하게 되면, FQDN 분석모듈은 수신된 메일의 헤더에서 메일 발신자 IP 주소를 분리해 낸다. 추출해 낸 IP 주소가 메일 수신 거부 IP 리스트에 포함되어 있으면 차단하고, 메일 수신 IP 리스트에 포함되어 있으면, 다음 단계를 거치지 않고 바로 바이러스 탐색 머신으로 포워드 한다.

메일 수신 IP 리스트나 메일 거부 IP 리스트 어느 쪽에도 속하지 않은 IP 주소는 <표 2>에 열거된 기본적인 정크메일 차단 필터 룰에 적용되는지 아닌지를 검사 받고, 도메인의 PTR 레코드 검색을 거쳐 FQDN(Fully Qualified Domain Name)인지 아닌지를 판별 받는다. FQDN의 판별 여부는 (그림 8)의 FQDN 확인 알고리즘의 절차에 의해 확인 된다. 추출해낸 메일 발신자의 IP 주소에서 네트워크 존과 호스트 주소를 분리해 낸 다음 해당 네트워크 존의 DNS 서버를 찾게 된다. 해당 DNS 서버를 찾았으면, IP 에서 호스트 이름을 찾는 정보를 제공하는 IN-ADDR.ARPA 정보를 검사하게 된다. IN-ADDR.ARPA의 존 정보에서 호스트의 IP 주소에 해당하는 PTR 레코드와 도메인 이름이 검색되면 FQDN(Fully Qualified Domain Name)이 되며, 그렇지 않고 상위 DNS 시스템에서도 PTR 레코드 정보가 검색되지 않으면 FQDN이 아니다.

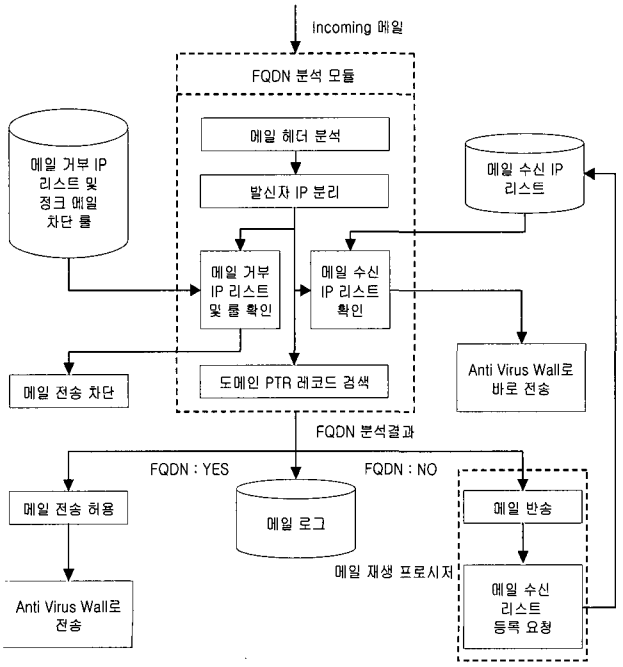
FQDN으로 판정 받으면 메일 전송에 대한 허가를 하고 다음 단계인 바이러스 검색 머신으로 포워드 한다. 만약

FQDN으로 판정 받지 못한 메일 발신자에게는 메일을 반송시키되, 메일 관리 홈페이지에 메일 수신 리스트로 등록 요청을 하도록 내용을 추가하여 메일을 반송 시킨다. 메일 수신 리스트에 등록된 메일은 FQDN 확인 절차를 거치지 않고 메일 전송 허가를 받도록 한다. 이와 같이 별도의 메일 수신 리스트를 만들도록 한 이유는 정상적인 메일이지만 IP 주소를 검색하여 PTR 레코드 검색 결과가 성공적이지 못할 경우를 대비하여 준비한 서브 프로시저이며, 발신 전용의 대량의 정크메일의 경우 반송되는 메일에 대한 처리를 못하기 때문에 정크메일에 대해서는 자동으로 차단하게 된다.

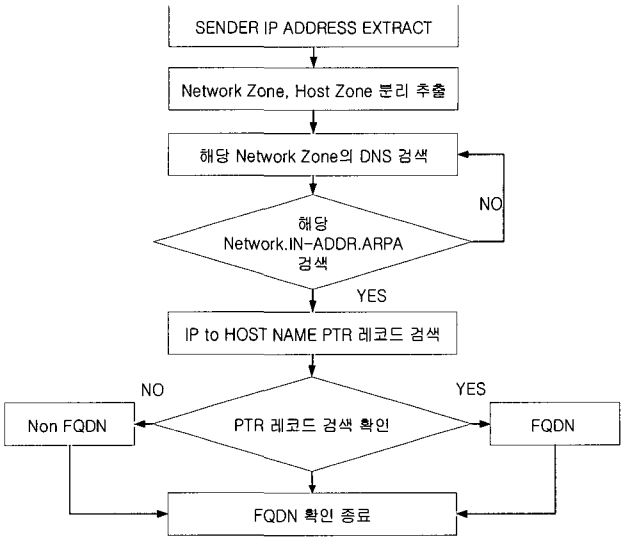
(그림 9)는 FQDN 확인 시스템을 이용하여 정크메일을 차단하도록 구성한 메일 송수신 시스템이다. FQDN 확인 시스템을 메일을 수신하게 되는 최전방에 위치 시킴으로써 일차적으로 정크메일을 차단 하고, 정상적인 메일로 판별 된 메일에 대한 부분만 메일 바이러스 탐색 머신 에서 바이러스 검사를 받도록 하여, 바이러스 검색에 필요한 시간을 최소화 한다. 따라서 정크메일 차단에서 바이러스 검사까지 (그림 3)과 같은 기존의 메일 시스템보다 메일 전송 지연에 따른 문제를 보완 하여 사설 망 내부에서의 메일 송수신 효율을 강화 시켰다.

〈표 2〉 정크메일 필터 룰

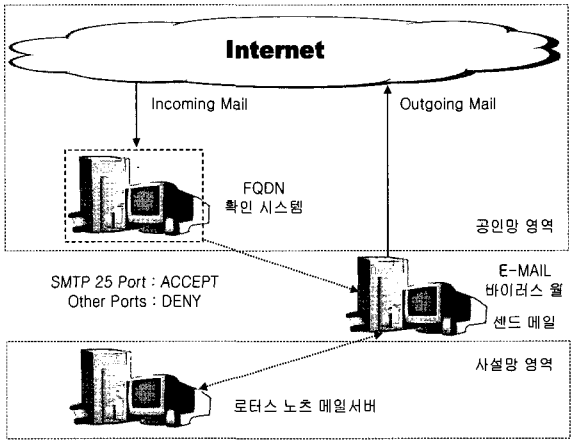
No	정크메일 필터 룰
1	From, To, 구문이 표준에 맞지 않는 메일 차단
2	From, To, Subject, Date에 값이 없는 메일 차단
3	To 이하의 수신자 도메인이 수신 서버 도메인과 일치하지 않을 경우 차단
4	메일 릴레이 시도 차단



(그림 7) FQDN 확인을 이용한 정크메일 차단 모듈 구성도



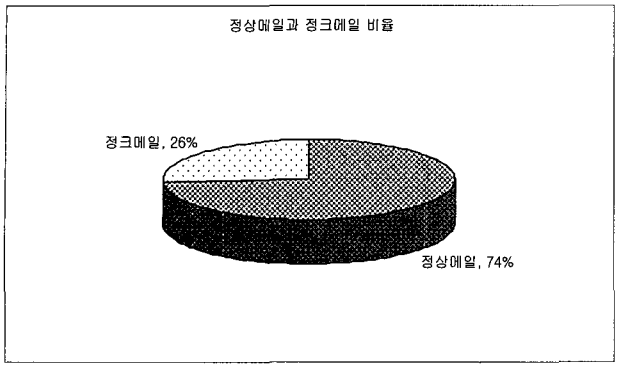
(그림 8) FQDN 확인 알고리즘



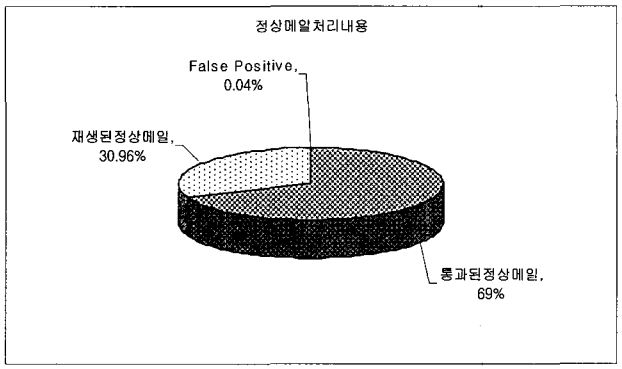
(그림 9) FQDN 확인을 이용한 개선된 정크메일 차단 시스템 구성

4. 사용결과 및 평가

본 논문에서 구현한 FQDN 확인을 이용한 정크메일 차단 시스템의 특징은 일부 정크메일 차단 시스템이 제공하는 내용별 메일 분류 기능에 초점을 맞추지 않았다. 이는 메일 시스템을 이용하는 사용자들의 요청에 의한 것으로 업무에 필요한 것 이외의 메일을 읽어 보고 분류하는데 소요되는 시간을 줄이고, 정크메일에 편승해서 침입하는 악성코드를 메일 시스템에 도달되기 이전에 차단하도록 하여 시스템을 보호하고 메일 시스템의 처리 효율을 높이기 위한 것이다. 본 논문에서 구현한 FQDN 확인을 이용한 정크메일 차단 시스템에 대한 평가는, 이 시스템이 전달되는 메일들 중 얼마만큼 정확히 정크메일을 차단하는가에 초점을 맞추었다. 본 논문에서 구현한 시스템의 평가를 위하여 시스템에서 처리한 3756개의 메일과 시스템에서 수집한 메일로그에 대하여 분석 한 결과는 다음과 같다. (그림 10)에서와 같이 메일 시스템에서 처리한 3756개의 메일 중 정크메일과 정상 메일의 비율은 각각 26%와 74%를 차지하였다.

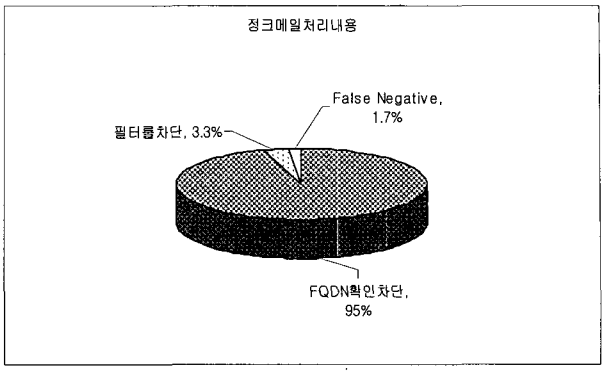


(그림 10) FQDN 확인을 이용한 정크메일 차단 결과 정상메일과 정크메일 비율



(그림 11) FQDN 확인을 이용한 정크메일 차단 결과 정상메일 처리내용

(그림 11)에서와 같이 정상 메일로 판별되어 사용자에게 전달된 메일 중 <표 2>의 필터 룰과 FQDN 확인 과정을 통과한 메일은 69%였으며 정상 메일 중 31%에 해당하는 메일이 FQDN 확인에 실패하여 메일 재생 프로시저를 거쳐 최종 수신자에게 전달 되었다. 정상적인 메일 중 31%나 되는 메일들이 FQDN의 확인 절차에서 실패했던 이유는 해당 메일 서버들이 공인 IP 부족이나 메일시스템의 보호를 위해 메일 시스템을 사설 망 안쪽에 두었을 경우, 메일을 교환해 주는 메일 서버와 DNS Name Resolution 설정을 할 때 PTR 레코드[11]에 대한 설정을 제대로 하지 않았거나, 인터넷 망의 공인 IP 대역에서 C Class 가 아닌 Subnet을 이용하여 메일 서버와 DNS Server를 구현 하였을 경우 Classless Name Resolution[12], [13] 설정을 해주지 않아서 PTR 레코드를 이용한 Reverse Name Resolution이 불가능하였기 때문이다. 이러한 메일들은 FQDN 확인에는 실패하였지만, 메일 재생 프로시저를 통해 최종 수신자에게 성공적으로 도달하도록 하였다. 이외에 정상 메일이 FQDN 확인에도 실패하고, 정크메일 필터 룰에도 해당하여 전달이 실패한 건수가 12건, 전체 수신된 정상 메일 대비 0.04%가 발생하였는데, 12건 모두 영세한 업체에서 메일 서비스를 호스팅 하여 사용하거나 메일 서비스를 임대하여 사용할 경우에 DNS 설정 오류나 메일 형식에 오류가 발생하여 전달에 실패한 경우였다.



(그림 12) FQDN 확인을 이용한 정크메일 차단 결과 정크메일 처리내용

(그림 12)와 같이 전체 수신 메일 대비 26%를 차지한 정크메일의 경우, <표 2>의 필터 룰에 의해 차단된 정크메일이 3.8%, FQDN 확인을 거쳐 반송되고, 메일 재생 프로시저에 응하지 않아서 차단된 메일이 95%를 차지 하였다. 정크메일이 정상 메일로 사용자들에게 전달된 건수는 15건이 발생하여 전체 차단된 정크메일 대비 1.7%를 차지하였는데, 분석 결과 이 경우는 모두 정상적인 도메인 설정 값과 메일 서비스를 하는 시스템이 정크메일 릴레이 공격에 이용되었거나 DNS Spoofing 공격에 의해 수신자의 메일 주소가 정크메일 공격자에게 유출되어 완벽한 정상 메일의 형태로 전달된 경우였다. 이 경우 메일 수신 거부 리스트에 정크메일 릴레이 공격에 이용되는 해당 메일 서버의 주소를 입력하여 정크메일 공격을 차단하도록 하고 해당 메일 서버 관리자에게 정크메일 릴레이 공격을 받고 있음을 알렸다. 본 논문에서 구현한 시스템을 평가한 결과 정크메일 차단 시스템의 성능 평가 척도가 되는 정크메일 차단 정확도는 98.3%로서 기존의 정크메일 필터링 방법들과의 비교를 위하여 기존에 발표된 논문[14], [15], [16]의 결론 부분을 인용하여 비교한 결과는 <표 3>과 같다.

〈표 3〉 정크메일 필터링 성능 비교

Filter Used	False Positive (%)	False Negative (%)	Spam Recall (%)	Spam Precision (%)
본 연구에서 제안된 시스템	0.04	1.7	97.2	98.3
Keyword Patterns	N/A	N/A	53.1	95.15
Naïve Bayesian	5.0	8.0	95.8	95.0

*Junk Mail Precision= Number of actual Junk Mail / Number of Classified Junk Mail

*Junk Mail Recall= Number of actual Junk Mail / Number of Total Junk Mail

제안된 방법과 기존 방식의 성능 평가 방법은 정크메일 정확도(Junk Mail Precision)와 정크메일 재현율(Junk Mail Recall)을 사용하였는데, 본 연구에서 제안한 시스템은 정크

메일 차단 정확도에서는 키워드 패턴 매치 방법 보다는 좋은 성능을 보였으며, 나이브 베이지안 방법과는 근소하게 좋은 성능을 나타낸 것을 확인 하였다. 또한 정크메일 재현율에서는 나이브 베이지안 방법이나 키워드 패턴 매치 방법보다 향상된 성능을 보임을 알 수 있다. 이는 기존의 시스템과는 달리 조건에 부합하지 않는 메일은 일단 리턴 후 재등록 하는 제안된 시스템의 메일 재생 처리절차 때문에 룰 베이스에 기초한 기존의 시스템 보다는 높은 차단 성공률을 보이는 것으로 분석된다.

5. 결 론

본 논문에서 구현한 FQDN(Fully Qualified Domain Name) 확인을 이용한 정크메일 차단 시스템은 대부분의 정크메일이 인터넷 망의 공인 IP에서, 수집된 대량의 메일링 리스트를 이용하여 발신 전용으로 발송 된다는 특성에 착안하여, 별도의 메일 분류에 필요한 알고리즘 없이 단순히 메일 헤더를 분석하여 정해진 형식에 맞게 메일이 발송 되었는지, 수신된 메일이 정상적인 도메인으로부터 발송 되었는지 확인하는 기능으로 정크메일을 차단 하였다. 기존의 정크메일 차단 시스템이 정확한 분류를 위해 정크메일에 대한 학습 기간을 거쳐 정크메일 차단 룰을 생성하고[17], 룰에 기반한 패턴 매치 차단 방법을 사용하기 때문에 메일 차단 시스템에서 수신자에게 정상적인 메일이 도달하기까지 많은 절차와 연산 작용을 거쳐야 되는 반면, 본 논문에서 구현한 시스템은 메일 헤더의 IP주소만 확인하여 반송시키거나 차단시키기 때문에 메일 전달 절차를 간소화 시켰고, 텍스트 기반이 아닌 정크메일과 특히 유동 IP를 이용해 발신 메일 주소를 주기적으로 바꾸는 정크메일의 경우 메일의 수신빈도나 IP를 추적해야 하는 정크메일 패턴 학습과정 없이 손쉽게 차단할 수 있어 우수한 효과를 보였다. 하지만, 정상적인 메일도 FQDN 조건을 만족하지 않으면, 메일 재생 프로시저를 통해 수신 메일 리스트에 등록해야 하므로 송신자에게 약간의 불편함을 주었지만, 종래의 수신자가 겪었던 메일 관리 및 삭제 등의 부담을 시스템의 안정성 유지와 사용자의 편의성간에 존재하는 상호 트레이드 오프 차원에서 메일 송신자의 이해를 유도하였다. 본 연구결과에서도 알 수 있듯이 공인 IP 주소 부족과 시스템 엔지니어의 전문적인 지식 결여로 인해 많은 수의 시스템들이 도메인 네임 서비스나 메일 서비스 설정 시 FQDN의 조건을 충족시키지 못했다. 따라서 향후 연구 과제로는 FQDN의 조건을 충족시키지 못하는 시스템들의 신뢰성을 향상시키는 방법에 대한 연구가 필요하다.

참 고 문 헌

[1] S. Atkins, "Size and cost of the problem," in Proceedings of the Fifty sixth Internet Engineering Task Force (IETF) Meeting, (SanFrancisco, CA), Spam Con Foundation, March 16-21, 2000.

[2] 박정선, 김창민, 김용기, "퍼지 관계 곱을 이용한 내용기반 정크 메일 분류 모델" 정보과학회 논문지 소프트웨어 및 응용 제 29권 제10호 pp.726-734, Oct., 2002..

[3] M. Salib, "Heuristics in the blender," in Proceedings of the 2003 Spam Conference, (Cambridge, US), 2002

[4] J. Weaver, "AOL escalates Spam warfare" MSNBC, March 5, 2003.

[5] 박광진, 공진동, 황성원, "2003년 정보화 역기능 실태조사" 한국정보보호 진흥원, 2003년 개인 인터넷 이용자의 정보화 역기능 살태 조사 보고서 pp.43-63, Dec., 2003.

[6] 정옥란, 조동석, "개인화된 분류를 위한 웹 메일 필터링 에이전트" 정보처리학회논문지B, 제10-B권 제 7호, pp.853-861, Dec., 2003.

[7] P. Pantel and D. Lin, "Spam corp. : A Spam classification & organization program," in Learning for Text Categorization : Papers from the 2000 Workshop, (Madison, Wisconsin), AAAI Technical Report WS-98-05, 2000.

[8] 서정우, 손태식, 서정택, 문종섭, "n-Gram 색인화 Support Vector Machine을 사용한 스팸메일 필터링에 대한 연구" 정보보호학회논문지, 제14권 제2호, pp.23-31, April, 2004.

[9] T. Oda and T. White, "Developing an immunity to Spam", in Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2003), (Chicago), July, 2003.

[10] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A Bayesian approach to filtering junk E-mail," in Learning for Text Categorization : Papers from the 2000 Workshop, (Madison, Wisconsin), AAAI Technical Report WS-98-05, 2000.

[11] G. Lindberg, "Anti-Spam Recommendations for SMTP MTAs," Chalmers University of Technology, RFC2505, February, 2000.

[12] D.Eastlake, C. Kaufman, "Domain Name System Security Extensions," RFC 2065, January, 2000.

[13] H. Eidnes, G. de Grouts, "Classless IN-ADDR.ARPA delegation" RFC 2317, March, 1998.

[14] Ion. A, Georgios. P, Vangelis. K, Georgios. S, Constantine. D, "Learning to Filter Spam E-Mail : A Comparison of a Naïve Bayesian and a Memory-Based Approach", PKDD 2000, pp.1-13, Sep., 2000.

[15] Mehran. S, Susan. D, David. H, Eric. H, " A Bayesian Approach to Filtering Junk E-Mail". In AAAI-98 Workshop on Learning for Text Categorization. 1998.

[16] Yanlei. D, Hongjun. L, and Dekai. W,"A Comparative Study of Classification Based Personal E-Mail Filtering", 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'00). pp.408-419, 2000.

[17] S. Forrest, S. A. Hofmeyr, and A. Somayaji, "Computer immunology," Communications of the ACM, Vol.40, No.10, pp.88~96, 2001.

김 성 찬

e-mail : viper72@chol.com

2000년~현재 ㈜유코레일 정보시스템부 과장

2002년 숭실대학교 정보과학 대학원 정보통신학과 석사

2002년~현재 숭실대학교 대학원 컴퓨터과학 박사과정

관심분야 : 정보보호, 유무선 PKI, VPN, MPLS, 트래픽 엔지니어링, QoS 라우팅, DiffServ, 액티브 네트워크, NGN, BcN

이 상 훈

e-mail : iam@leesanghun.pe.kr

2001년 숭실대학교 컴퓨터학부(학사)

2003년 숭실대학교 컴퓨터학부(석사)

2003~현재 숭실대학교 컴퓨터과학 박사과정

관심분야 : 정보보호, 컴퓨터 바이러스

전 문 석

e-mail : mjun@computing.ssu.ac.kr

1980년 숭실대학교 전자계산학과(학사)

1996년 University of Maryland 전산과(석사)

1989년 University of Maryland 전산과(박사)

1989년 Morgan State University 전산수학과 조교수

1989년~1991년 New Mexico State University 부설 Physical Science Lab. 책임연구원

1991년~현재 숭실대학교 정보과학대학 부교수

관심분야 : 컴퓨터 알고리즘, 병렬처리, VLSI 설계, 암호학