논문번호 05-02-08

Privacy 속성 기반의 오인된 메일 복구 알고리즘 개발

Development of A Recovery-algorithm of False-Positive Mail based on the Property of the Privacy

徐 相 鎭*, 陳 鉉 埈*★, 朴 魯 京*
Sangjin Seo*, Hyunjoon Jin*★, Nohkyung Park*

요 약

현재 전자우편은 정보통신 사회의 중요한 의사소통 수단이 되고 있으나, 스팸 메일의 증가로 인해 각종 사회 문제를 발생시키고 있다. 스팸 차단을 위해 관련 기관 및 기업이 다양한 형태의 스팸 차단 기술을 개발하고 있으나, 다양한 스팸 유형에 따른 대응이 각각 이루어져야 하므로 많은 처리 비용과 시스템의 복잡도가 가중되고 있다. 그리고 스팸 차단 기술 적용 시, 스팸 방지 필터의 적용 순서에 따라 차단 오인율(False-Positive Error)이 달라져 이용 신뢰도에 큰 영향을 끼친다. 본 논문에서는 스팸 차단 필터의 이용 신뢰도를 향상시키기 위해, Privacy 정보 기반의 False-Positive 메일 복구 기법을 제안 및 구현하였다. 구현된 프로토 타입을 통해 False-Positive 메일 복구 과정을 검증하고, 처리 결과를 분석 및 요약한다.

Abstract

While E-mail has become an important way of communications in IT societies, it creates various social problems due to increase of spam mails. Even though many organizations and corperations have been doing researches to develop spam mail blocking technologies, more cost and system complexities are required because of varieties of blocking technologies. In case of adopting spam blocking technologies, system reliability largely relies on the False-positive error rate with the order of employing spam blocking filters. In this paper, a False-positive mail recovery technique based on privacy information is proposed and implemented in order to improve the reliability of spam locking filters. Through the implemented prototype, recovery procedure for False-positive mails is verified and the results are summarized and analyzed.

Keywords : 스팸 메일 , 스팸 메일 차단, 오인된 메일 복구, Privacy, 스팸 차단 필터

Ⅰ. 서 론

정보 통신 분야에서 전자우편(E-Mail)은 인터넷 이용자 중 84.6%가 이용하는 중요한 의사소통 수단이다. 그러나 전자우편의 증가로 인해 스팸 메일을 이용한 온라인 마케팅이 활성화되고 있으며, 불법 음란 스팸 메일과 같은 유해 정보를 가진 스팸 메일이 성인뿐만 아니라 아동 및 청소년에게 무차별 전송되어 사회 문제가 되고 있다. 이를 규제하기 위해 정보통신

* 호서대학교 전기정보통신공학부 정보통신공학전공
(Department of Electrical and Info. & Comm. Engineering, Hoseo University)
★ 교신저자 (Correspondence Author)
接受日:2005年 8月 4日, 修正完了日:2005年 12月 26日

※ 본 과제(결과물)는 교육인적자원부와 산업자원부의 출연금 및 보조금으로 수행한 산학협력 중심대학 육성사업의 연구결과입니다.

망이용촉진 및 정보보호 등에 관한 법률 등을 강화하고 있지만, 대부분의 스팸 발송자들이 해외에 서버를 두고 스팸 메일을 발송하기 때문에 국내법에 따라 규제하기는 사실상 불가능하다[1].

스팸 메일을 차단하기 위해 관련 기관의 노력으로 다수의 안티 스팸 솔루션이 공개되어 있다. 스팸 메일은 불원성(不願性), 상업성(商業性), 대량성(大量性)의 특성을 가지고 있다. 이러한 특성을 이용한 다수의 스팸 메일 차단 기술들이 등장하고 있으나, 다양한 스팸 유형에 따른 대응 필터의 수가 늘어 스팸 메일 시스템의 복잡도와 처리 비용을 가중시키고 있다. 그리고 스팸 차단 필터의 처리 순서에 따라 정상 메일이 스팸 메일로 오인되어[2] 스팸 차단 신뢰도가 크게 하락하므로, 오인된 스팸 메일 복구 기법의 연구가 필요하다.

본 논문에서는 스팸 차단 시스템의 신뢰도를 향상시키기 위해 Privacy 속성 기반의 오인된 스팸 메일(False- Positive Spam Mail)의 복구 기법을 제안한다. Privacy 속성은 유출되지 않는 개인 정보로 정의된다[3]. 제안된 복구 기법의 검증을 위해 프로토 타입을 구현하고 처리과정을 분석한다. 논문의 구성은 2장에서 Privacy 기반의 오인된 메일 복구 기법을 제안한다. 그리고 3장에서 제안한 복구 기법의 수행을 위한 프로토 타입을 설계 및 구현하고, 처리 결과를 분석한다. 4장에서는 결론 및 향후 연구 과제를 기술한다.

II. Privacy 정보를 이용한 False-Positive 메일 복구 기법

2.1 오인된 스팸 메일 복구 기법의 개요

본 장에서는 Privacy 정보를 이용한 오인된 스팸 메일 복구를 위한 기본적인 처리 과정과 기존의 스팸 메일 차단 서버와의 연동 방식에 대해 개괄적으로 기술한다.

기존의 스팸 메일 차단 기술은 광고나 스팸 메일을 규정하는 스팸 키워드 DB와 필터 규칙(filtering rule-set)을 이용하여 내용 필터링[4]하여 스팸 메일을 판정한다. 그러나 신규 스팸 메일의 발생은 스팸 키워드 DB를 학습시키고, 차단 필터 규칙을 작성하는 것에 선행된다. 이는 발생과 대응 시간 간격 동안 신규 스팸 메일의 대응이 없음을 의미한다.

Privacy 정보는 개인 정보 항목이 키워드 리스트로 구성된 정보이다[3]. Privacy 정보는 스팸 정보에 비해 발생 빈도가 정적이며, 구축 정보의 크기도 광범위한 스팸 정보에 비해 작다. Privacy 정보 기반의 복구 기법은 스팸 메일 발신자의 개인 정보 참조 특성과 범위, 개인 정보 발생량에 대한 특성을 기반으로 처리한다.

스팸 메일 작성자는 각 개인에게 스팸 메일을 발송하기 위해 인터넷 사이트, 자동 메일 수집기 등 공개된 Privacy 정보 중 일부만 접근할 수 있으며, 대량의 메일 전송으로 인해 개별 발송자 메일에 대한 처리가 어려워 최소한의 Privacy 정보만 포함하고 있는 특성이 있다. 그러나 정상 메일은 공개된 Privacy 정보 이외에 추가적인 고유한 Privacy 정보를 포함하고 있다.

기존 스팸 메일 차단 기술에서 처리를 위해 참조하는 DB의 학습(Training)을 위해 베이시안 기법[5-6]을 사용한다. 베이시안 기법은 입력된 문장 중 관심 단어만 영역별 분류하여, 분류된 키워드의 포함 여부를 통계를 통해 판정하는 기법이다. 본 논문의 Privacy 메일 판정을 DB 학습을 위해 베이시안 기법을 이용한다. DB 학습 대상 키워드는 수신자의 Privacy 키워드이며, 기 학습된 Privacy 키워드와 발송 메일 내에 구문 분석을 통해 추출된 Privacy 후보자 키워드와 비교하여 동일 키워드이거나 Privacy 정보를 동반하는 문장의 동사로부터 추출된 Privacy 키워드로 판정되면 저장된다. 학습되는 Privacy 키워드는 스팸 메일 발송자의 접근 경로가 차단된 속성인 별명, 개인과 관련된 인적 사항 등의 희소가치에 따라 가중치를 다르게 부여하여 판정 시 참조한다.

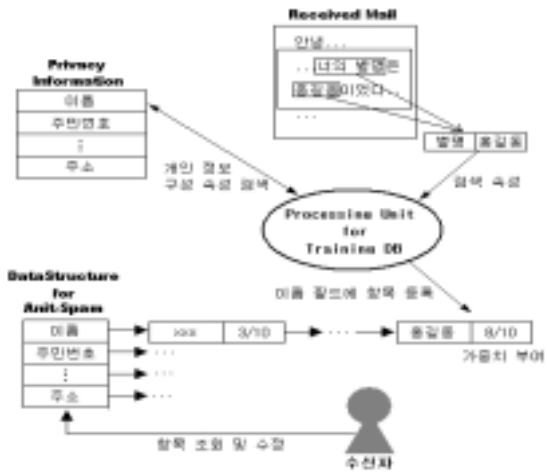

그림 1. 추출된 Privacy 정보의 DB-Training 과정
Fig. 1. DB-Training procedure of extracted privacy information.

본 논문의 복구 기법은 메일 수신 부하가 매우 높은 서버형 차단 프레임에 적용된다[7]. 그러므로, 시스템 부하를 줄이기 위해 스팸 메일 보관 모듈과 복구 모듈을 분리하여 연동시킨다. 이는 스팸 메일 차단을 위한 메일 수신 서버의 가용 자원과 분리시켜 병행처리의 부담을 분산시킬 수 있다.

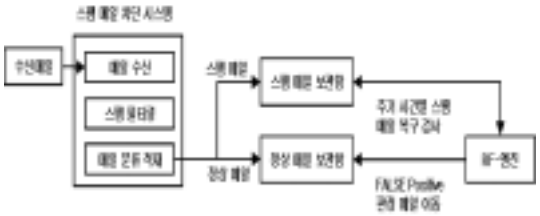

그림 2. False-positive 메일 복구
Fig. 2. False-positive mail recovery.

2.2 복구 기법을 위한 처리 구조 및 제안 알고리즘

Privacy 정보 기반의 오인된 메일 복구 기법은 수신 메일을 분석하여 고유하거나 특징적인 Privacy 정보를 추출 및 통계치를 환산하여 오인된 메일의 복구 여부를 판단한다. 그리고 메일 수신자의 Privacy 정보를 수집 및 관리하며 필요시 Privacy 정보가 복구에 참조 가능하도록 한다.

제안된 복구 기법은 폴 그램(Paul Gram)의 베이시안 조합(Bayesian combination) 기법을 이용한다[8]. 그림 3 은 베이시안 조합을 이용한 전반적인 오인된 메일 복구 과정을 간략 나타내고 있다.

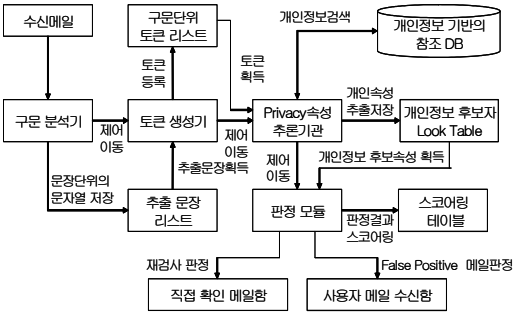

그림 3. Privacy 정보 기반의 False-positive 메일 복구 전체 처리 구성도
Fig. 3. Block diagram of the False-positive mail recovery based on privacy Information.

복구 기법은 크게 한국어 구문 분석기, 토큰 생성기, Privacy 속성 추론 기관, 판정 모듈 등으로 구성된다. 한국어 구문 분석기는 구문 분석된 키워드와 구문 품사 등의 정보가 제공된다. 토큰 생성기는 분석된 구문을 속성(Property) 형태로 저장한다. Privacy 속성 추론 기관은 분석된 구문 키워드와 품사를 이용하여, 기존의 개인정보로 학습된 Privacy DB를 참조하여 Privacy 속성 및 Privacy 속성 후보자를 추론한다. 판정 모듈은 추론 모듈로부터 추출된 Privacy 속성 및 Privacy 속성 후보자 키워드의 가중치를 누적하여 오인된 정상 메일(False-Positive Mail)로 판정한다. 본 논문에서는 그림 4 와 같이 오인된 메일 복구 처리를 위해 메일 구성 요소를 5개의 튜플로 정의한다. 여기서, P_k는 C_k에게 종속되는 Subset 이다.

그림 5에서 Privacy 기반의 오인된 메일 복구 알고리즘을 나타내고 있다.

N, S, FP, FN ∋ {S, P, C, V, E}
* 단, M = {N, S, FP, FN}
- M : {N, S, FP, FN}인 튜플들의 집합
- N : 정상 메일
- S : 스팸 메일
- FP : False-Positive 메일
- FN : False-Negative 메일
- S_k : {S_1, S_2, ... S_i}인 Spam Keyword들의 집합
- P_k : {P_1, P_2, ... P_i}인 Privacy Keyword들의 집합
- C_k : {C_1, C_2, ... C_i}인 Privacy Candidate Keyword들의 집합
- V_k : {V_1, V_2, ... V_i}인 Privacy Keyword 추출을 위한 문장의 동사 집합
- E_k : S, P, C, V 튜플에 속하지 않는 키워드

그림 4. 복구 기술 적용을 위한 메일 구성 튜플 정의
Fig. 4. Definition of tuples for mail elements used for recovery technique.

```
-Input: Tuple FP ∋ {S, P, C, V, E}
-Output: A judgement on a False-Positive mail

[0] Initialized the prototype.
[1] Copy a new e-mail.
[2] Analysis Body of Mail using KLT-Parser.
[3] Make Data-Structure for the protype.
[4] while(until reach end of a statement-list) {
[5]       read a statement from a statement-list.
[6]       Make a token-list for retrieving
           privacy-properties
[7] }
[8] while(until reach end of a token-list) {
[9]       read a pair of tokens from a token-list
[10]      retrieve properties of Ck in the privacy DB for
           privacy-information
[11]      if(a retrieved property is a pair of tokens for
           a candidate)
[12]            insert properties of Ck of candidate into
                a lookup table.
[13] }
[14] while(until reach end of a lookup table) {
[15]      read a property of Ck from a lookup table.
[16]      validate relation of Pk on a Ck.
[17]      if(Is a validated property)
                scoring a field related of privacy
[18] }
[19] processing analysis of a scoring table.
```

```
[20] judge a inputed e-mail FALSE-positive
[21] if(a result of judgment is a FALSE-positive mail)
[22]      move a mail to user mailbox
[23] else if(a result of judgment is not a
           FALSE-positive mail)
[24]      leave a mail in spam mailbox.
[25] else
[26]      move a mail to unsure-folder.
[27] free resources allocated prototype.
```

그림 5. Privacy 속성 기반의 False-positive 메일 복구 알고리즘
Fig. 5. False-positive mail recovery algorithm based on privacy information.

Privacy 속성 기반의 오인된 메일의 복구 알고리즘은 우선 검사 대상 메일을 복사하여 구문 분석하고, 문장을 토큰화된 Linked List로 구성한다. 토큰화된 Linked List로부터 Privacy 속성을 추론하여 P_k와 C_k를 추출한다. 추출된 Privacy 속성을 판정 및 스코어링하기 위해 Privacy 속성 저장 Lookup Table에 등록하여 기존에 등록된 Privacy 속성과의 비교를 통해 일치 판정을 수행한다. 속성으로 판정된 토큰은 등록된 Privacy 속성의 가중치를 누적하고, 누적치와 오인된 메일 판정 스코어와 비교하여False-Positive 메일을 결정한다.

본 논문에서는 베이시안 기반의 필터링 수행시 가능한 확률적인 오판을 최소화하고 이분법적인 분류를 배제하기 위해 직접 확인을 위한 False-Positive 후보자 메일 보관함(unsure mailbox)을 추가 적용하였다.

III. 프로토 타입 개발

본 장에서는 프로토 타입 구현 및 처리를 통해 False-Positive 메일의 복구과정을 살펴본다. 프로토 타입의 실행 결과에 따라 오인된 스팸 메일의 처리 결과의 통계치를 이용하여 시스템 이용 신뢰 여부를 분석한다.

3.1 프로토 타입 설계 및 구현

Privacy 정보 기반의 False-Positive 메일 복구 처리를 수행하는 프로토 타입을 구현 및 실행하기 위해 윈도우 2000 Server 환경에서 Visual Studio 6.0, KLT 2.0 파서[9]를 이용하였다. 실험을 위해 텍스트 기반의 스팸 메일 샘플 90개와 False-Positive 메일 10개를 임의의 순서로 입력 소스를 구성하였다. 100개의 입력 메일 소스를 리눅스 Redhat 9.0 버전의 Sendmail 메일 서버를 통해 스팸 필터링을 수행하였다. 스팸 필터링은 한글 처리를 위한 hcode 버전과 Procmail에 스팸 필터 스크립팅을 이용하여 스팸 차단을 수행하여, 10개의 의도된 False-Positive 메일이 발생하였다. 발생된 10개 메일을 윈도우 2000 Server에 제목과 본문만을 전송하여 복구 프로토타입이 개별 입력 처리하였다. 그림 6는 특정 광고 키워드를 스팸 필터링하기 위해 sendmail 메일 서버의 사용자 메일 계정인 "/home/ssjworld/"에 procmailrc의 스팸 필터 스크립팅 중 일부이다.

프로토타입의 처리 과정을 살펴보기 위해 그림 7와 같은 False-Positive 메일을 처리 과정을 분석한다. 그림 7의 False-Positive 메일의 발생 원인은 특정 광고를 차단하기 위한 내용 필터 스크립팅에 의해 오인된 메일이다.

```
*^(Subject|From|Cc):.*=\?EUC-KR\?(B|Q)\?
|formail -c | hcode -dk -m
*^Subject:.*(광고|홍보|기구|목록입니다|리스트입니다|
성인정보|제품|...)
/var/mail/spam
```

그림 6. procmailrc의 광고 키워드 필터 스크립팅
Fig. 6. Filter scripting for advertisement keywords of procmailrc.

안녕하세요.
주임님께서 보내신 물품 내역을 잘 받았습니다.
A사의 a**제품**은 이미 시장에서 잘 **홍보**되고 있습니다.
그러나 본사의 b **제품**은 A사의 a**제품**보다 더욱 우수한 성능의 **제품**입니다.
그래서 본사의 **제품**이 주임님이 계신 S사의 영업부서에 더 적합할 듯합니다.
추운날씨에 옥체 보존하십시오.
-B사 드림-

그림 7. 실험을 위한 Sample False-positive 메일
Fig. 7. A sample False-positive mail for experiments.

프로토타입은 스팸 메일이 보관된 리눅스 내에 "/var/mail/spam." 폴터의 메일 제목과 내용을 복사한다. 복사된 메일은 행단위로 구분되어 statement-list에 저장된 후 KLT 구문 분석기를 통해 구분의 품사 분석이 수행한다. 그림 8은 예제로 만들어진 False-Positive 메일의 KLT 구문 분석 결과를 나타내고 있다. 분석된 구문의 형태소 정보는 "순서", "품사", "키워드" 순으로 노드가 구성된다. KLT 구문 분석기에 의해 지정된 포맷으로 중간 출력 파일 "KLTAnalysis.dat"를 생성한다. 프로토 타입은 KLT 구문 분석기의 중간 출력 결과를 분석하여 P_k 정보 추출을 위한 품사와 키 값으로 구성된 키워드 노드들의 연결 리스트를 구축한다.

분석된 구문은 KLT에서 정의된 품사에 따라 분류된다. 처리 과정에서 분석된 품사 중 'N'은 체언, 'K'는 고유명사, 'A'는 알파벳으로 시작되는 고유명사, 'P'는 대명사,

'C'는 복합명사를 나타낸다. 구문 분석된 각 노드는 P_k 추출을 위해 구조체 노드에 분류된다. 각 구문을 구성하는 노드들의 품사와 할당된 값은 Privacy-DB를 이용하여 C_k를 추출한다. 추출된 C_k는 특정 개인 정보 속성 노드를 판정하기 위해 Lookup 테이블에 복사된다. Lookup 테이블은 P_k를 추출하기 위한 C_k의 연결 리스트로 구성된 임시 저장 공간이다. P_k를 추출하기 위해 Privacy-DB에 등록된 수신자의 개인 정보 속성을 참조한다. 그림 9은 실험을 위해 구축된 수신자의 Privacy 정보 영역 및 속성을 간략히 나타내고 있다.

0:N:안녕	1:P:제품	5:P:a	6:C:영업부서
0:K:주임님	3:N:시장	5:P:제품	6:P:영업
2:N:물품	5:N:홍보	7:N:우수	6:P:부서
3:N:내역	1:N:본사	8:N:성능	8:N:적합
0:@:A사	3:N:제품	9:N:제품	0:K:추운날
0:P:A	4:@:A사	1:N:본사	1:N:옥체
0:P:사	4:P:A	2:N:제품	2:N:보존
0:@:A사의	4:P:사	3:K:주임님	1:@:B사
0:P:사의	4:@:A사의	5:C:S사	1:P:B
1:@:a제품	4:P:사의	5:P:S	1:P:사
1:P:a	5:@:a제품	5:P:사	2:N:드림

그림 8. KLT 분석기를 이용한 Prototype의 구문 분석결과
Fig. 8. Syntatic analysis results using KLT analyzer in prototype.

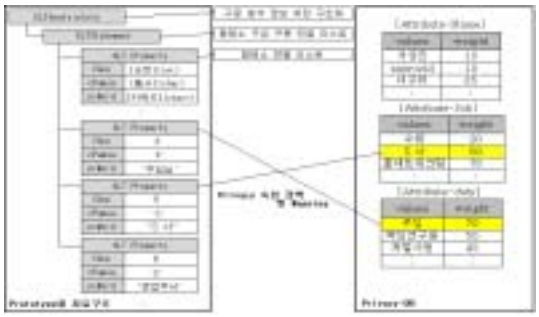

그림 9. 가중치 참조를 위한 Privacy Attribute DB
Fig. 9. Privacy attribute DB for reference of weight factors.

가중치는 스팸 메일 발신자가 접근하기 어려운 Privacy 정보의 고유성에 따라 높아지며, 시간에 따라 사용 빈도가 낮아지거나 스팸 메일 발신자가 직접 접근 가능한 Privacy 속성(직장주소 등)들의 가중치는 떨어지게 구성된다. 추출된 Lookup 테이블의 Privacy 속성은 "직책-주임", "직장-S사"이며, 각각 70, 60의 가중치를 가진다. 모든 가중치는 130으로 합산되어 False-Positive 메일 판정 기준 스코어인 70점 이상을 취득하여 False-Positive 메일로 판정한다. 그리고 "직장-영업부서"는 신규 Privacy 속성 P_k로 등록되며, 가중치는 최초 30점이 부여된다. 그림 10은 메일 복구를 위한 프로토 타입의 개별 처리 과정을 보여주고 있다.

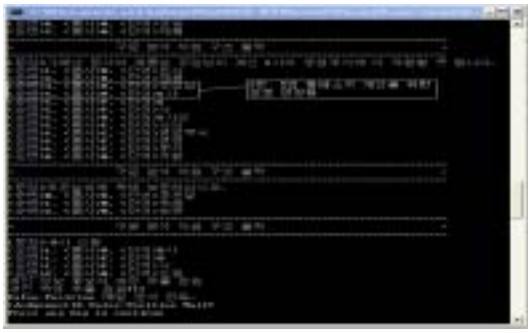

그림 10. Prototype 처리 결과 화면
Fig. 10. Screen capture of prototype processing results.

3.2 처리 결과

프로토 타입 실행을 통해 False-Positive 유도 메일을 포함한 100개의 텍스트 기반 메일을 리눅스 Sendmail의 스팸 메일 차단 필터를 이용하여 검사하였다. False-Positive 메일은 유도된 10개가 발생하였으며, 본 논문에서 제안하는 Privacy 정보 기반의 스팸 차단 프로토 타입은 6개의 False-Positive 메일을 복구하였다.

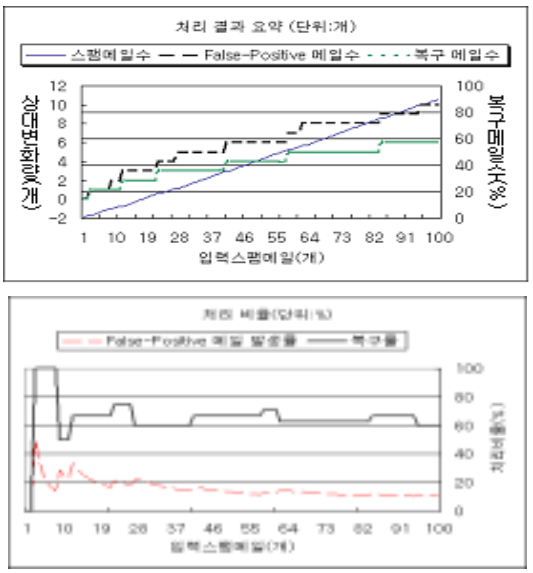

그림 11. 실험 결과 요약 그래프
Fig. 11. Summary graphs of experiment results.

그림 11은 실험 결과에 따라 스팸 메일 판정에 따른 False-Positive 메일 발생수를 간략히 나타내고 있다. 100개의 스팸 메일을 처리하는 과정에서 점진적으로 False-Positive 메일이 증가하는 것을 알 수 있다. 최종 프로토 타입의 False-Positive 메일 복구률은 최대 60% 정도임을 그래프를 통해 알 수 있다.

처리 결과에 따라 복구되지 않은 4개의 메일은 메일 내 P_k 속성 판정을 위한 키워드가 Privacy-DB에 학습되지 않거나, 동사에 의한 개인 정보 추론이 가능한 메일로 분석된다. 본 논문에서 사용한 KLT 구문 분석기는 형태소 중심의 구문 분석만을 수행하며, Privacy 속성 추론을 위한 동사 기반 키워드 DB는 일부만 참조하여 완전한 FALSE-Positive 메일 복구 기법의 상용화를 위해 더욱 많은 Privacy-DB의 학습이 필요하다.

IV. 결론

본 논문에서는 Privacy 정보 기반의 False-positive 메일 복구 기법을 제안하고, 제안한 복구 기법의 처리를 위해 프로토 타입을 설계 및 구현하였다. 복구 성능을 검증하기 위해 10개의 유도된 False-Positive 메일을 포함한 100개의 샘플 스팸 메일을 처리하였다. 유도된 10개의 False-Positive 메일이 Sendmail 서버의 Procmail의 차단 필터에 의해 오인되었으며, 오인된 10개의 메일 중 6개의 메일에서 Privacy 키워드를 추출하여 60%의 False-Positive 메일을 복구하였다. 그러나, 복구 성능에 직접적인 영향을 주는 Privacy-DB 학습 알고리즘의 구현에서 KLT 구문 분석기의 부적절한 품사 분석과 학습 키워드 부족으로 인해 Privacy-DB에 등록되지 않는 C_k (Privacy Candidate Keyword) 추출의 정확도가 떨어졌다.

Privacy 정보 기반의 오인된 메일 복구 기법은 실제 적용을 통한 성능 및 안정성 평가에 따라 단계별 스팸 차단 필터로 대치 가능하다. 수신된 메일의 내용 분석을 통해 형식 변형 메일의 원문 추출 후, 직접 Privacy 기반의 개인 메일로 판정 및 결과에 따라 정상 메일함에 저장하여 기존 스팸 메일 차단 필터들과 대치할 수 있다. Privacy DB는 스팸 차단을 위한 스팸 DB에 비해 상대적으로 크기가 작고 정적인 정보들로 구성되어, 하드웨어 장치화를 위한 모듈 구분과 정적인 정보 특성으로 인해 전체적인 처리 능력을 강화할 수 있다.

향후 연구 과제로 더욱 정교한 한국어 구문 분석기와 베이시안 조합 기술을 연동하여 동사의 의미에 따른 Privacy 속성 추론 엔진을 개발한다. 그리고, 전체 시스템 성능 향상을 위해 Privacy 속성 기반의 스팸 메일 차단 장치를 Embedded 장치에 이식하여 분산 처리 환경을 구축한다.

참 고 문 헌

[1] 주덕규, “스팸메일의 현황 및 대책”, 한국정보보호진흥원, 정보통신윤리, 2003. 6.

[2] 불법스팸대응센터, “스팸메일 규제방식에 대한 검토”, 2002. 5.

[3] T. Saito, K. Umesawa, and H.G. Okuno, “A Privacy-Enhanced Access Control”, 日本電子情報通信學會論文誌, Nov., 2001.

[4] 김진만, 장종욱, "컨텐츠 필터를 이용한 스팸메일 차단 시스템 설계 및 구현", 한국해양정보통신학회, 2003년도 춘계종합학술대회, 2003.

[5] Mehran Sahami, Susan Dumais, David Heckerman, Eric Horvitz, “A Bayesian Approach to Filtering Junk E-Mail”, Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, Aug., 2001.

[6] Ion Androutsopoulos, Georgios Paliouras, Vangelis Karkaletsis, Georgios Sakkis, Constantine D. Spyropoulos and Panagiotis Stamatopoulos, “Learning to Filter Spam E-Mail: A Comparison of a Naïve Bayesian and a Memory-Based Approach”, Proceedings of the "Machine Learning and Textual Information Access", Workshop of the 4th European Conference on Principles and Practice of Knowledge Discovery in Databases(PKDD), 2000.

[7] 불법스팸대응센터, “이메일서비스 업체별 스팸메일 방지기술", 2002. 6.

[8] 조한철, 조근식, "나이브 베이지안 분류자와 메시지 규칙을 이용한 스팸메일 필터링 시스템", 한국정보과학회, 제29회 춘계학술대회, 2002.4.27.

[9] http://nlp.kookmin.ac.kr

저 자 소 개

서 상 진 (정회원)

1999년 2월 부경대학교 전자계산학과 학사.
2001년 2월 부경대학교 전자계산학과 석사.
2005년 3월 ~ 현재 호서대 일반대학원 정보통신학과 박사과정

<주관심분야> 임베디드 시스템, 멀티미디어 응용, 모바일 응용

진 현 준 (정회원)

1984년 2월 고려대학교 전자공학과 학사.
1986년 2월 고려대학교 전자공학과 석사.
1998년 1월 Lehigh 대학교 전산학 박사.
1998년 3월 ~ 현재 호서대학교 전기정보통신공학부 부교수

<주관심분야> 시스템프로그램, 멀티미디어 소프트웨어, 무선 LAN 프로토콜

박 노 경 (정회원)

1984년 2월 고려대학교 전자공학과 학사
1986년 2월 고려대학교 전자공학과 석사
1990년 2월 고려대학교 전자공학과 공학박사
1988년 4월 ~ 현재 호서대학교 전기정보통신공학부 교수

1999년 3월 ~ 2000년 2월 미국 오레곤 주립대학교 ECE 연구교수
<주관심분야> 회로 및 시스템 설계, SoC 설계