Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.25, NO.1, Feb. 2015 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2015.25.1.107

사례기반추론기법을 적용한 해킹메일 프로파일링*

박 형 수,[1†] 김 휘 강,[1] 김 은 진[2‡]
[1]고려대학교, [2]경기대학교

Hacking Mail Profiling by Applying Case Based Reasoning*

Hyong-su Park,[1†] Huy-kang Kim,[1] Eun-jin Kim[2‡]
[1]Korea University, [2]Kyonggi University

요 약

방어기법은 계속 진화하고 있지만, 여전히 이메일을 통한 APT 공격의 경우에는 탐지와 방어가 어렵다. 국내의 공공기관 및 많은 민간기업들이 지능적이고 지속적인 해킹메일 공격으로 인해 피해가 지속되고 있다. 본 논문에서는 최근 몇 년간 공공기관을 대상으로 유입된 실제 해킹메일들을 데이터베이스로 구축한 뒤, 해킹메일들의 특징을 추출한 뒤 해킹메일을 사례기반추론기법을 이용하여 유사한 사례들을 쉽게 탐색할 수 있도록 사례벡터를 설계하였다. 또한 해킹메일을 이용한 과거 공격사례들을 프로파일링 해 본 결과 특정 지역과 집단에 공격이 집중되었음을 확인할 수 있었다. 이를 통해 사례추론기법이 신규 이메일이 유입될 경우, 과거 해킹메일과의 유사성을 비교하여 악성 여부를 탐지하는데 응용할 수 있을 뿐 아니라, 침해사고 대응에 있어서도 효율적으로 활용 가능함을 제시하였다.

ABSTRACT

Many defensive mechanisms have been evolved as new attack methods are developed. However, APT attacks using e-mail are still hard to detect and prevent. Recently, many organizations in the government sector or private sector have been hacked by malicious e-mail based APT attacks.

In this paper, first, we built hacking e-mail database based on the real e-mail data which were used in attacks on the Korean government organizations in recent years. Then, we extracted features from the hacking e-mails for profiling them. We design a case vector that can describe the specific characteristics of hacking e-mails well. Finally, based on case based reasoning, we made an algorithm for retrieving the most similar case from the hacking e-mail database when a new hacking e-mail is found.

As a result, hacking e-mails have common characteristics in several features such as geo-location information, and these features can be used for classifying benign e-mails and malicious e-mails. Furthermore, this proposed case based reasoning algorithm can be useful for making a decision to analyze suspicious e-mails.

Keywords: Hacking Mail, CBR, Profiling

I. 서 론

이메일 사용 증가와 함께 지능적이고 지속적인 해킹메일공격 으로 인한 피해가 해마다 증가하고 있다. 해킹메일은 시스템파괴, 주요자료 및 개인정보유출, 기밀자료 유출 등 사이버 범죄에 광범위하게 사용되어 개인의 차원을 넘어 정부, 방위산업, 에너지, 제조업 등 국가 주요 산업과 기반시설을 그 표적으로 하고 있다.

접수일(2014년 11월 6일), 수정일(2015년 1월 2일), 게재확정일(2015년 1월 22일)
* 본 논문은 미래창조과학부 및 정보통신산업진흥원의 '지식정보보안인력양성 최고정보보안전문가과정' 사업의 연구결과로 수행되었음(과제번호 : NIPA-H2102-13-1002)
† 주저자, boraman3@hanmail.net
‡ 교신저자, ejkim777@kyonggi.ac.kr(Corresponding author)

정상적인 문서에 악성코드를 은닉 시킨 후 국가기관의 주요인사 또는 특정인에게 유포시켜 PC에 저장된 중요 자료들을 빼내가는 해킹이 지속적으로 발생하고 있다. 또한 2013년에는 국가 주요기관 및 민간분야에 대한 대규모 사이버공격이 연이어 발생하였으며 2013년 '3.20 사이버테러'는 약 1주일에 걸쳐 6개 방송·금융사를 대상으로 발생하였다. 이러한 대규모 사이버 테러 유발에 있어 해킹메일을 통한 악성코드 유포가 그 중심에 있는 것으로 판단된다[1].

때문에 이러한 해킹메일 공격에 대해 누가 공격했는지, 왜 공격을 수행했는지에 대한 정보를 파악하고 이러한 공격을 예방하고 사전에 차단하는 것이 더 큰 피해를 막는 방법이라 할 것이다. 이에 본 논문은 해킹메일 프로파일링을 통해 해킹메일 공격에 대한 효과적인 탐지 방안을 제시하고자 한다.

II. 관련연구

2.1 해킹메일 공격 사례분석

해킹메일 공격은 시간이 지나감에 따라 더 지능적이고 다양해졌으나 정찰, 전달, 설치, 명령 및 제어 등과 같은 여러 가지 디지털 단서를 남긴다. 그 이유는 각 단계마다 공격자와 표적 사이에 어떤 접촉점이 필요하기 때문이다. 이러한 접촉들은 공격자에 대해 더 상세하게 알아볼 수 있으며 추적이 가능하도록 한다. 해킹메일에서 단서로 활용할 수 있는 정보들은 다음과 같다. 해킹 메일의 헤더에 대한 '문자집합'의 특성을 조사해 보면, 특정한 악성코드를 작성하기 위해 사용하는 악성코드 키보드의 배열을 확인할 수 있다. 대부분의 피싱 시도는 특정한 국가를 나타내지 않는 표준 키보드 배열을 사용한다. 그러나 비표준 키보드를 사용했다는 것이 명확할 경우 중요한 디지털 단서가 된다. 공격자가 악성코드 공격을 할 때 중국에서 사용하는 표준 중국어(GB2312) 키보드로 입력하는 특징있으며, 공격자들이 북한의 KPS 9566 문자집합을 사용하는 경우 그 지역에서 시도된 악성코드 공격을 식별하는데 활용할 수 있다. 또한 공격에 사용된 도메인을 분석하면 공격자의 위치를 정확하게 알아낼 수 있다. 허위 이름과 주소를 사용하여 DNS에 등록하였더라도 위치정보를 찾아낼 수 있으며, 다수의 도메인에 대해 연락처 정보를 분석하여 공격자 행위를 연계 파악할 수 있다[2,3].

해킹메일에 의한 공격은 다양한 정보를 가지고 있으며, 유럽과 미국에서는 페이스북, 링크드인 등으로부터 정보를 수집하고 있다고 알려지고 있다. 공격자가 노리는 정보는 이름, 이메일 주소, 회사명, 직책, 인간관계 등의 정보들이다. 이러한 정보를 기반으로 공격자는 위장 이메일을 작성한다. 여러 사례를 보면 최초 공격 이메일의 발신처는 대부분의 정보를 인터넷상에서 정보를 입수하고 있는 경우가 많다 또한 잘 알려지지 않은 듯하지만, 휴대폰 및 스마트폰 이메일 주소에 해킹메일이 오는 사례도 있고 SMS 및 카카오톡 등 애플리케이션을 통한 공격도 등장할 것이다.

본 연구에서는 00기관에서 수집된 실제 해킹메일 500여건의 정보를 활용하여 해킹메일의 특성과 실험데이터로 활용하였다. Table 1. 의 해킹메일 발송 현황을 살펴보면 해킹메일(스파이 이메일, 바이러스가 첨부된 이메일 등)은 수신자에게 흥미가 있을 만한 내용을 본문에 기재하여 보내고 있다. 예를 들어 월드컵이 개최되는 해에는 월드컵과 관련된 내용을 기재하거나, 특정 개인의 연구 과제물에 대한 세부적인 사항까지 파악하여 수신자가 해킹메일을 확인하도록 메일을 발송하고 있다[3].

최근 해킹메일의 가장 큰 특징은 외교·안보업무

Table 1. Hacking mail

Sender	Subject
UnO (rheac0@daum.net)	Please save the North Korean defectors
KimOO (jammy00@yahoo.com)	National Crisis Management
JOO (000_reader@daum.net)	hello JOO
HongOO (tongil2000@daum.net)	Dear Hong○○
GoenOO (daum.server@aol.com)	Subject research report (No ○○)
GoenOO (naver.server@aol.com)	Invitation (Annual General Meeting)
KimOO (kimjung000@yahoo.co.kr)	Happy New Year seniors
Naver Server (naver.server@aol.com)	Please change your password right time
Daum Server (daum.server@aol.com)	Daum member-Note stability
researcher (leeyo000@hanmail.net)	Security issues analysis

관계자 및 해외공관·주재원 등을 대상으로 이메일을 통한 해킹시도가 빈발하고 있다. 발신자는 주로 상용 이메일(한메일·네이버 등)을 이용하며 정부 주요인사 실명 또는 지인 등을 사칭하여 업무관련 메일로 위장(해킹한 메일계정 사용)하고 있다.

이에 따른 주요 해킹 메일 공격의 실제사례는 다음과 같다.

- 사례1 : Fig. 1. 메일 관리자로 속여 회원정보 변경 유도 - 메일 수신자가 로그인 화면에서 다시 ID·비밀번호 입력 시 해킹 경유지로 접속 ID 및 비밀번호가 유출

Fig.1. 의 해킹메일은 메일 관리자로 속이고 사용자가 아이디와 패스워드를 입력하여 계정을 유출하도록 하는 링크 URL 포함형 해킹메일이다. 발송자명은 Daum Server이지만 발송지 메일계정업체가 미국의 aol.com 인 것을 확인할 수 있으며, 수신자의 계정이 저장되는 서버는 추적이 어려운 미국에 있는 무료 호스팅 업체인 것을 확인할 수 있다.

- 사례2 : Fig.2. 00기관 명의 도용 해킹메일로 특정기관의 실제 담당자 또는 주요 직위자들의 정보를 도용하여 해킹메일 유포

Fig.2.의 명의 도용 해킹메일은 00기관의 실제 존재하는 인물의 이름을 도용하여 악성코드가 담긴 문서를 발송한 명의 도용 해킹메일이다. 해킹메일 유포자는 사전에 송신자와, 수신자와의 관계를 알고 있으며, 수신자가 믿을 수 있는 송신자로 위장하여 메일을 발송하였다. Fig.2. 의 해킹메일의 주요 특징은 발송자의 계정명과 실제 발송지(중국)로 해킹메일로 판단할 수 있는 요소가 담겨져 있다.

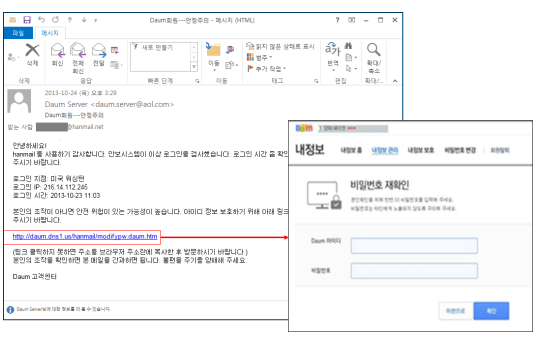

Fig.1. Hacking Mail #1

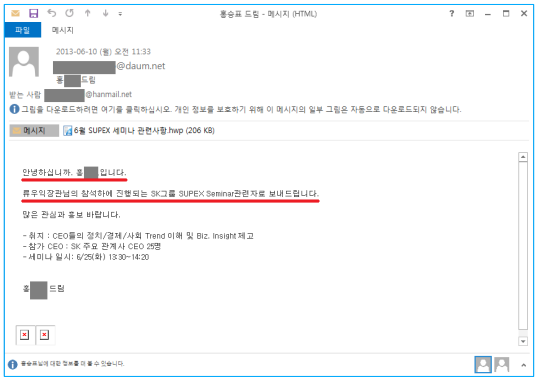

Fig.2. Hacking Mail #2

- 사례3 : Fig.3. 첨부파일 열람유도 해킹메일 첨부파일을 열람 시, PC에 악성코드에 감염 되도록 수신자 흥미유도

Fig.3. 해킹메일은 수신자가 어떤 분야에 종사하고 있으며 관심분야가 어떤 분야를 사전에 파악한 후 수신자가 관심을 가지고 있는 첨부파일명으로 메일을 발송하여 첨부파일 열람 시 악성코드에 감염되도록 유도한 첨부파일 열람 유도형 해킹메일이다. Fig.3.의 해킹메일의 주요 특징은 수신자가 00기관의 연구원장이며, 실제 메일발송지는 중국으로 확인할 수 있다. 위와 같은 해킹메일 공격은 메일내용에 '○○장관'·'○○국장' 등 실제 직책을 암시하며 메일 제목 클릭 혹은 첨부문서 열람토록 유도 하며, 해킹프로그램에 감염된 인터넷PC에 USB 등 외장 저장장치 사용 시 보관 자료가 전량 유출되는 피해가 발생 한다. 이러한 해킹공격의 공격자는 글의 표면적으로 드러나는 뜻뿐만 아니라 수신자의 인간관계까지 파악한 후에 이메일을 보내기 때문에 상당히 주의하지 않는다면 위장된 것을 눈

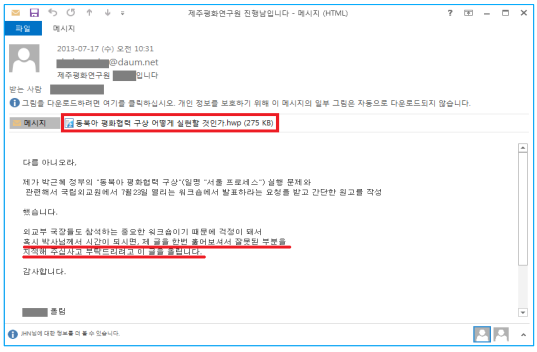

Fig.3. Hacking Mail #3

치 채기는 어려울 것이다. 인간관계의 경우 공격자가 검색 엔진 등을 이용하여 조사하거나 이미 침입한 PC에 저장되어 있는 이메일을 분석하여 파악하게 된다. 또한 키로거로 이메일의 내용을 취득하는 경우도 있다. 실제로 커뮤니티 및 특정 학회 회원을 노리는 경우가 많고, 어떠한 수단을 이용해 전달된 이메일을 훔쳐 읽는 경우가 많다

공격자가 수집한 이메일 주소는 매우 다양하다. 업무에 이용하고 있는 이메일 주소 외에도, 개인의 사적인 이메일의 주소가 표적이 되는 경우도 있다. 이러한 해킹 메일의 발신이 증가하고 있는 이유는 렌탈 서버나 프리 이메일 등 개인을 특정하기 어려운 환경을 이용하기 때문이다[3].

Table 2.는 이메일에서 사용하는 주요 키워드 목록이다. 이메일 메시지는 빈 줄에 의해 두 부분으로 구분되는 ASCII 텍스트로 구성된다. Header의 첫 번째 부분은 발신이자, 수신자, 메시지가 전송되는 날짜, 내용의 형식과 같은 메시지에 대한 정보를 포함하고 있다.

다음의 Fig.4.는 실제 해킹메일의 헤더 정보이다. 헤더 정보를 통해 해킹메일 공격의 목적과 특징 등의 여러 가지 요소들을 분석할 수 있는 단서들을 확인 할 수 있다.

Fig.4.의 실제 해킹메일의 헤더 정보를 보면 우선 "Received" 정보는 발송자가 보낸 메일이 발송자의 IP와 메일이 수신자에게 도달할 때까지 거쳐 간 Routing 정보를 확인할 수 있다. 여러 개의 "Received" 로 시작되는 문구 중 가장 아래 부분의 "Received ~ by" 이하 정보가 발송자가 메일을 보낼 때 사용한 메일서버를 나타내며, for 이하 정보는 수신자의 전자메일 주소 정보를 알 수 있다. 즉 Received : 〔"from"발송 호스트〕

Table 2. E-mail Header Keyword

Keyword	Mean
From	Sender Address
To	Recipient Address
Date	Date the message was sent
Subject	Message Subject
Reply-To	Reply-to address
X-Charset	Character Set
X-Mailer	Message SW
X-Originating-IP	Real IP Address

```
x-beehive-trace: ○○○_reader@daum.net
○○○@○○○.re.kr 10.201.1.9
Received: from daum.net
by ○○○.re.kr with ESMTP (Beehive_○○○.re.kr)
for ○○○@○○○.re.kr; Wed, 17 Jul 2013
13:24:13 +0900 (KST)
x-beehive-kind: normal
x-beehive-modified: received kind
Received: from unknown (HELO smail-34.
hanmail.net) (211.43.197.216)
by 10.201.1.9 with ESMTP; 17 Jul 2013
10:31:39 +0900
Received: from wwl1753.hanmail.net ([117.52.3.183])
by smail-34.hanmail.net (8.12.1/8.9.1) with
ESMTP id r6H1V852026913;
Wed, 17 Jul 2013 10:31:10 +0900
Received: (from hanadmin@localhost)
by wwl1753.hanmail.net (8.12.9/8.9.1) id
r6H1V6NS022350
for 〈○○○@○○○.re.kr〉; Wed, 17 Jul 2013
10:31:06 +0900
Errors-To: 〈○○○@daum.net〉
Date: 17 Jul 2013 10:31:06 +0900
Message-Id:
〈20130717103106.HM.00000000000001w@○○○
_reader.wwl1753.hanmail.net〉
From: "JHN" 〈○○○_reader@daum.net〉
Sender: ○○○_reader@daum.net
To: "=?utf-8?B?7KCE6rK966eMIOuwleyCrOuLmA
==?=" 〈○○○@○○○.re.kr〉
X-Original-SENDERIP: 211.43.197.216
X-Original-MAILFROM: ○○○_reader@daum.net
X-Originating-IP: [175.167.128.9]
Organization:
Subject:=?utf-8?B?7KCc7KO87Y+J7ZmU7Je
w6rWs7JuQIOynhO2WieuCqOyeheuLiOuLpA==
?=
X-Mailer: Daum Web Mailer 1.2
X-HM-UT:qnIqdClQyCewqZ3okLQmBC3tq
HSIvMfn2+5XRe1d7hc=
X-HM-FIGURE:qnIqdClQyCeL8ku99jccCBhms
x FTTpCl
MIME-Version: 1.0
X-Hanmail-Attr: fc=1
Content-Type:multipart/mixed;
boundary="1374024666.DaumWebMailer."
X-SCRAP: Proceed 20130717132338
```

Fig.4. Hacking Mail Header

"by" 수신 호스트["with 메일 프로토콜"] "id" 문자열["for" 수신자 메일주소]";" 날짜 및 시간 정보 순으로 나열되어 있다. 또한 Date 정보에서의 "+0900"는 타임존(Time Zone)를 의미하며 국내에서 이메일이 보내졌다면, 타임존은 "+9시간", 이며 "+8시간"으로

되어 있는 경우 이 타임존은 주로 중국, 러시아, 몽골, 말레이시아, 오스트레일리아 등의 지역이다. 그리고 "X-Original-MAILFROM" 에서 실제적으로 메일을 발송한 송신자의 메일주소를 확인 할 수 있으며, 가장 중요한 "X-Originating-IP" 는 실제 메일을 발송한 송신자의 PC의 IP 정보가 확인 가능하며, "X-Mailer" 에서는 메일을 발송한 메일 제품명과 버전을 확인할 수 있다. 이렇게 메일 헤더 정보에는 여러 가지 의심스러운 정보를 분석하며 세부정보를 확인할 수 있다. 이런 해킹메일을 아래와 같이 분석 유추할 수 있다[3].

우선 해킹메일은 발신처가 국내 기업인 것처럼 보이는 경우가 많다. 하지만 실제는 미국이나 제 3국인 발신경로인 경우가 많다. 국내 지인에게서 온 이메일이었을 경우, 이메일이 작성된 문자코드가 한국어가 아닌 것은 부자연스럽다. 맞춤법이 틀리거나 ?gb2312?가 있는 문자열이 있는 경우도 있다. 이런 문자코드는 중국어(간체) 환경에서 작성된 이메일로 추측할 수 있다. 국내에서 이메일이 보내졌다면, 타임존은 "+9시간"이 될 것이다. 하지만 Date를 확인하며 "+8시간"으로 되어 있는 경우가 있다. 이 타임존은 주로 중국, 러시아, 몽골, 말레이시아, 오스트레일리아 등의 지역이다. 일률적으로 어느 나라에서 보냈다고 말할 수 없더라도 적어도 국내에서 보낸 메일은 아닌 것이다. 이 세 가지 이유로도, 이 헤더 정보를 포함한 이메일은 해킹메일일 가능성이 높다고 추측할 수 있다[3].

해커가 이메일 발송용 주소를 확보하기 위해 사용하는 수단은 다음과 같다 무료 메일에 타인의 명의를 도용, 위장가입과 유관기관 직원이나 공격 대상자 친지의 메일계정을 해킹, 타인 명의로 메일이 발송되었을 경우, 스팸메일로 보일 수 있기 때문에 공격 대상자가 메일 첨부파일을 열람하지 않을 가능성이 있다. 따라서 메일 헤더를 조작하거나 메일내용을 통해 지인을 가장하는 방법을 사용하기도 한다. 두 번째 방법을 사용할 경우는 피해자가 해킹메일의 첨부파일을 열람할 가능성이 매우 높지만 목표를 해킹하기 위해 장기적인 준비가 필요한 부담이 있다. 해커는 각종 문서파일, 압축파일, 실행파일 등 다양한 파일을 첨부하거나 클릭 시 해킹 프로그램이 자동 설치되는 인터넷 링크를 포함하여 메일을 발송한다[4].

Fig.2.는 전형적인 해킹메일의 예이다. 메일 수신자가 이메일 첨부파일(한글, MS오피스, PDF 등)을 열람할 경우, 해킹프로그램이 PC에 설치된다. 첨부파일을 열어보도록 유도하기 위해 다양한 방법이 동원된다. Fig.2. 처럼 지인이 보낸 메일로 위장하는 방법부터 정상 이메일 위조, 유혹, 협박 등의 방법이 사용되기도 한다. 해킹메일 수신자가 메일에 첨부된 파일을 열어보거나 인터넷 링크를 클릭하여 해킹프로그램이 실행되면 해커가 사전에 확보한 자료유출 경유지로 해킹에 성공했다는 신호가 전송되며 해커의 역접속을 기다리게 된다.

피해자는 자신이 해킹되었다는 사실을 인지하기 매우 어렵다. 해킹프로그램이 숨겨진 문서파일을 열어도 정상적인 문서 내용이 표시되기 때문이다. 해킹프로그램이 자동으로 설치되는 인터넷 홈페이지에 접속하는 링크를 클릭했더라도 눈으로 보기에는 정상적인 홈페이지로 보인다. 해커는 공격대상자의 PC가 해킹되었음을 확인하고 피해자의 PC에 은밀히 접속하여 각종 악성프로그램을 설치, 하드디스크에 저장된 자료와 키보드에서 입력되는 내용을 미리 확보해둔 자료유출 경유지로 유출한다. 심지어 피해자 PC화면을 자신의 모니터로 보면서 원격 조종하는 해킹프로그램도 존재한다. 피해자의 PC와 메일, 게임 사이트, 인터넷뱅킹 로그인 정보 등은 2차 해킹이나 범죄에 악용될 수 있다.

2.2 사례기반추론

사례기반추론 알고리즘은 1977년 예일대학에서 연구가 시작되어 1980년 처음 Koton과 Bareiss에 의해 의학 분야에 적용되었다. CBR 알고리즘은 의학 분야뿐만 아니라 현재 마케팅, 과학, 사회학 등 여러 분야에 광범위하게 적용되고 있다. 범죄 수사 분야도 예외는 아니다. 범죄 수사를 통해 수집된 방대한 증거 자료들을 효과적으로 분석하기 위해 데이터베이스를 구축하고 검색하는데 적절한 방법이 될 수 있다. 범죄 방법에 대한 자료 수집과 저장 분류는 범죄 특성과 행위 패턴을 식별하는데 도움을 준다[3][5].

CBR 알고리즘은 유사한 과거 문제의 해결에 기초해서 새로운 문제를 해결하는 과정이다. 그 이론적 배경으로 역동적 기억이론(Schank, 1982, 1999)은 학습을 기존의 기억과 새롭게 받아들이는 경험이 상호작용을 거치면서 갱신되는 과정으로 보았다.

이런 과정은 추론을 통하지 않고는 유의미하게 이루어지지 않는다. 즉 새로운 것을 학습하고 이해하기 위해서는 기존의 기억 내용과 관련을 지어야 하는데 그 과정에서 추론은 필수적인 것이 된다는 것이다. 현재 발생한 문제가 과거의 사례와 정확하게 일치하지는

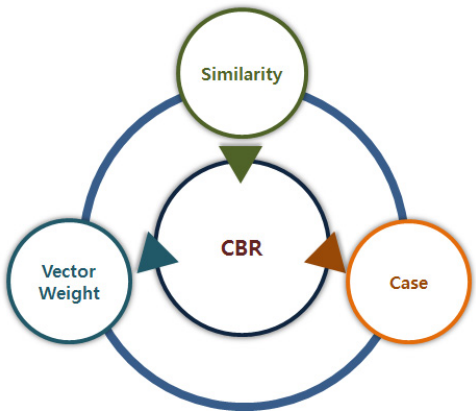

Fig.5. Hacking Mail of CBR model

않더라도 과거 사례나 경험은 현재 문제에 대한 부분적인 해결책을 제시할 수 있다. 사례기반추론은 새로운 문제를 해석하고 해결하는데 과거의 경험 사례를 이용하는 인간의 지적 능력이다. 즉, 자신의 과거 경험 속에 현재 직면하고 있는 문제와 유사한 사례가 있고 그 문제를 성공적으로 해결한 방안을 기억하고 있다면 인간은 그 사례를 현재의 문제에 적용하여 해결하려 한다는 것이다. 즉, 과거 사례와 지식들을 데이터베이스로 구축하여 새로운 문제가 발생했을 때 기존의 사례 데이터베이스에서 똑같거나 유사한 사례를 선택하여 그 사례가 제시하는 해결책으로 현 발생한 문제에 대한 답을 제시할 수 있다는 것이다[5].

2.3 프로파일링 분류

프로파일링이란 범죄 현장에서 범죄자가 나타낸 다양한 행동의 분석을 통해 범죄자의 배경 특성을 추론하여 범인 검거에 기여하고 범인의 협조를 얻게 하는 수사기법이다. 즉, 범죄자 프로파일링은 범죄 현장에는 범죄자의 평소 습성이나 일상적인 행동 방식이 고스란히 반영되어 있다는 것을 전제로 범죄 현장을 분석하여 범죄자의 유형을 밝혀내는 것이다. 이 정보에 따라 수사 전략을 세우고 수사망을 좁혀 범인 검거에 기여한다. 범인이 검거된 후에도 심문 전략을 세워 범인이 수사 기관의 조사에 협조 할 수 있도록 하여 다각적 기능을 수행한다. 이러한 프로파일링 목적은 어떠한 범죄 행위자를 특정 하는 것이 아니고 범죄 현장을 통해 그러한 범죄를 행할 만한 특정 대상자의 유형을 묘사, 설정하는 것으로, 크게 두 가지 활동을 한다. 첫째 범죄자의 사회심리학적 특성을 파악한다. 범죄자의 연령, 직업, 종교, 혼인 여부, 교육수준, 추가 범죄 가능성 등을 예측할 수 있는 다양한 심리학적인 정보를 수집하고 파악한다. 둘째, 범죄 현장에서 확보된 제반사항을 평가한다. 범죄자가 사용한 도구나 소유물을 통해 심리적 의미를 파악하는데, 용의자를 특정 하는데 매우 중요한 단서가 될 수 있다. 이와 같이 프로파일링을 통해 범죄자의 특성을 파악하여 혐의가 짙은 용의자를 식별하는데 도움을 줄 수 있고, 수사기관이 확실한 물적 증거를 확보하기 전에 범죄 가능성이 높은 용의자를 계속 추적할 근거를 제공할 수 있다[2,6].

해킹메일 발신자는 자신의 목적을 달성하기 위해 공격 대상기관을 선정하고 해당기관 소속 직원들의 이메일 주소를 홈페이지, 인터넷검색, 인명록, 명함, 관련자 PC 해킹 등 다양한 경로를 통해 입수한다. 그리고 해킹메일 발신자는 추적을 피하기 위해 여러 단계의 중간거점을 확보한다. 이러한 경유지로는 해커가 미리 해킹해둔 국내외 서버 및 PC 등이 이용된다. 이러한 공격방법을 분석하여 해킹메일 공격에 대한 프로파일링 하여 해킹메일 특징과 특성, 목적 등을 분류할 수 있다[3].

III. 제안알고리즘

CBR 알고리즘을 적용하기 위해 기존의 해킹메일 DB에서 유사도를 측정하기 위한 해킹 메일의 필수요소를 선별하여 각 요소들에 대하여 중요도에 따른 가중치와 점수화하여 유사도를 측정하여 데이터를 분석하였다. 해킹메일에서 의미 있는 특징값(feature)들을 찾기 위해, 과거 해킹메일 사례를 토대로 공격자가 사이버공격을 수행하기 위해 반드시 필요한 메일의 송신계정, IP 주소 및 동일한 첨부파일 및 악성코드를 식별할 수 있는 파일의 C&C서버의 URL 등 9종의 정보를 선별하였다. 이때 EML 헤더와 Contents 부분으로 나누어 구분하였으며, Body 부분에서는 첨부파일 유형과 링크 포함 유형으로 구분 하여 요소를 식별하였다. Table 3. 과 같이 해킹메일 식별 요소 분류하게 된 이유는 공격자가 사이버공격을 수행하기 위해 반드시 필요한 메일의 발신계정, IP, 호스트 정보 및 동일한 첨부파일 또는 URL 등 9종의 정보는 사이버공격에 사용하는 중요한 식별요소로서 선정하였다. Table 3. 의 해킹메일의 식별요소는 별도의 작업이 필요 없이 해킹메일 수집 시 최단시간에 확인 가능한 요소로서의 장점을 가지고 있다. 9종의 식별요소로만

Table 3. Hacking Mail Vector

EML Header	IP address, Date, Id, Sender Name, Domain Name, Characterset, Receive ID
Body	Subject, Filename or URL

으로도 공격집단이 갖는 유사성과 관계성을 분석할 수 있다.

- 발신자 IP 주소: 실제 해킹메일 발송 PC 정보 확인을 통한 공격자 정보 또는 공격자 집단 정보 유추 가능
- 발신 계정명: 도용 또는 해킹한 계정 정보 확인 을 통해 공격자 성향 정보 유추 가능
- 발신 일자: 유사한 일자에 발송된 해킹메일 공격은 동일 공격자 또는 공격집단의 공격으로 유추할 수 있음.
- 발신 이름: 발신 이름 통해 공격하려는 목표의 정보를 알 수 있어 공격 의도를 유추할 수 있음
- 도메인 이름: 발송된 해킹메일 도메인 정보를 통한 해킹메일 정보 특징 유추 가능
- 언어: 공격자 지역정보 유추 가능
- 수신 계정명: 공격 목표정보 유추가능으로 공격 자 의도 유추 가능
- 제목: 공격 목표에 대한 의도와 공격 성향 유추 가능
- 첨부파일명 or URL 정보: 공격 목표에 대한 연관성 정보 획득 가능

EML 헤더 (EML Header)에서 발신지 IP는 중요한 요소로서 실제 메일을 발신한 PC의 IP 주소를 확인하여 실제 공격자 정보를 수집할 수 있다. 수집된 해킹메일의 헤더 부분에 담겨진 〔X-Originating-IP〕 필드가 발신지 IP이다. 발신계정명은 〔from〕 필드에 있는 발송자 계정과 발신이름 도메인 정보를 추출하여 요소로 사용하며, 발신일자는 From 부분의 〔Date〕을 사용한다. 또한 To 필드에 있는 부분에 수신 계정명 정보가 담겨져 있다. EML에 담겨진 문자집합 정보, EML Contents 부분에서 메일의 종류에 따라 본문내용과 첨부파일 정보 등의 요소를 추출한다. 이러한 9가지의 해킹메일 정보의 요소를 추출하여 공격자의 성향과 다양한 정보를 분석할 수 있다. 추출한 요소들은 다시 공격자의 정보에 끼치는 영향의 정도에 따라 High, Medium, Low로 구분하여 중요도를 식별하였으며 이에 따라 가중치를 부여하였다. 이중 발신자 IP, 첨부파일명, URL 정보, 발신일자 등은 공격자 행위를 분석하는데 높은(High) 중요도로 식별하였다. High로 중요도를 식별한 요소들은 해킹메일 공격자 집단 식별과 공격유형을 판단하는데 가장 중요한 영향을 끼치는 요소로 식별하였다. 발신 계정명, 발신이름, 제목 등은 중간(Medium) 정도의 중요도 부여하였는데 Medium 레벨의 중요도는 반드시 필요한 해킹메일의 요소이지만 공격자가 조작이 가능한 요소로서 명확한 요소로 식별하기 제한사항이 있는 요소들이다. 그러나 변경 가능한 요소들이 모여 공격집단의 연관성을 식별할 수 있는 요소들이다. 도메인 이름, 언어, 수신 계정명은 낮은(Low)요소로 분류 하였다. Low로 분류한 요소들은 중복성이 많은 요소로서 가중치를 낮게 부여하였다〔7,8〕.

EML 헤더 영역에서는 요소들의 측정값은 두 가지 방법으로 측정할 수 있도록 설정하였다. 우선 발신자 IP 주소, 첨부파일명 or URL 정보, 발신일자는 유사도에 따라 1~0 사이에 5가지 값으로(1, 0.75, 0.5, 0.25, 0) 구분하여 각각의 값으로 부여한다.

발신 계정명, 발신이름, 제목, 도메인 이름, 언어,

Table 4. Hacking Mail Vector

Vector	IP address	File name / URL	Date	ID	Sender Name	Subject	Domain Name	Character set	Receive ID
Level	High			Medium			Low		
Weight	9	8	7	6	6	5	3	2	1
Score	1	1	1	0 OR 1	0 OR 1	1	0 OR 1	0 OR 1	0 OR 1
	0.75	0.75	0.75			0.75			
	0.5	0.5	0.5			0.5			
	0.25	0.25	0.25			0.25			
	0	0	0			0			

수신 계정명은 유사도에 따라 0 또는 1 값을 부여하여 측정값을 적용하였다. IP 주소 (IP address)에 대한 유사도 측정 방법은 참고문헌 [2]에서 제시한 방식을 이용하였다.

발송지의 IP 주소는 유사도를 측정하기 위해 4자리의 클래스를 다음과 같이 분리하여 계량화하였다. 예를 들어 1. 1. 1. 3 의 아이피는 A, B, C, D로 분류하여 수치화 하여 측정하여 최고점을 1부터 0까지 5단계로 나누어 유사도 점수를 할당한다.

- A IP Address : A, B, C, D
- B IP Address : a, b, c, d

첨부파일명 or URL 정보는 기존의 유사도 알고리즘 적용시 범위차이 과다발생과 전체의 문장을 완벽하게 수치화하여 유사도를 측정하기 어려워 중요단어의 유사도를 측정하여 단어의 유사도 개수에 의하여 Table 6.과 같이 유사도 점수를 부여한다.

Table 7.에서 보는 바와 같이 발신일자는 유사도를 정량화 측정하기 위해 발신일자에 대해 최소 7일부터 ~ 최대 6개월 이내까지 값을 측정하여 유사도 점수를 할당하여 유사값에 의해 분류될 수 있도록 하였다.

CBR 알고리즘을 적용하기 위한 해킹메일의 각 요소들에 대한 중요도에 가중치를 부여 점수화 하였다. 이러한 해킹메일 요소들을 Fig.6. 과 같은 해킹메일 CBR 요소들을 종합한 해킹메일 프로파일링 체계를 구성하였다.

Table 5. IP Address Similarity Measure

Case	Score
All	1
A = a, B = b, C = c	0.75
A = a, B = b	0.5
A = a	0.25
not All	0

Table 6. Attchfile Name or URL Similarity Measure

Case	Score
All	1
3 word Over	0.75
2 word Over	0.5
1 word Over	0.25
Miss Match	0

Table 7. Mail Date's Similarity Measure

Case	Score
7 day Under	1
1 Month Under	0.75
3 Month Under	0.5
6 Month Under	0.25
6 Month Over	0

IV. 적용 및 실험

4.1 시험환경 및 시나리오

본 논문에서는 해킹메일과 일반 메일이 유사도에서 차이를 보이는지를 파악하기 위하여 실제 스팸메일과 정상메일 적용을 통해 유사도 범위의 정확도를 실험하여 유사도 거리를 측정하였으며, 해킹메일 사례들 중 4개 그룹으로 분류한 해킹메일에 대해 사례기반추론 알고리즘 적용하여 유사도 값과 정확도를 측정하였다.

Table 8.에서 제시한 Test Case #1은 실제 해킹메일과의 유사도 거리를 측정하기 위한 스팸메일과 정상메일이다.

Table 9.은 실제 해킹메일 사례들을 4개 그룹으로 구분한 해킹메일로서 기존의 해킹메일과 유사도와 정확도를 측정하기 위해 계정 유출형과 악성코드 첨부파일형 두 가지로 구분하여 실험을 진행하였다.

위의 두 가지 실험 조건을 통한 결과값을 통해 실제

Table 8. Test Case #1

· Test Case #1 : Spam Mail and Mail
① Spam Mail : Yahoo Spam mail (2014)
② Mail : Naver mail (2013)

Table 9. Test Case #2

· Test Case #2 : 4 Groups Hacking Mail
① AA Group Hacking Mail (2014)
② BB Group Hacking Mail (2013)
③ CC Group Hacking Mail (2013)
④ DD Group Hacking Mail (2014)

해킹메일과 유사도와 정확도를 측정할 수 있다.

이러한 유사도 결과 뿐만 아니라 Fig.6.에 제시한 해킹메일 프로파일링 구성체계와 같이 DB화된 해킹메일 DB와 유사도를 검증하고 Virustotal, Maltego 등 다양한 도구들을 통해 공격자와 공격집단들의 유사성과 특징들을 조합하여 공격행위와 목적들을 측정 분류 프로파일링 할 수 있다[9,10].

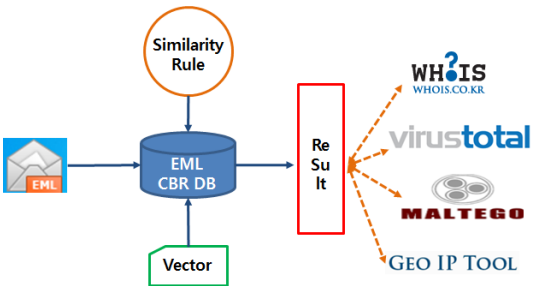

Fig.6. Hacking Mail of Profiling

4.2 시나리오 Ⅰ - 일반메일과 유사도 거리 측정

우선 해킹메일이 아닌 스팸메일과 정상메일 적용을 통해 유사도 범위의 정확도를 측정하기 위하여 국내와 외국의 메일업체에서 발송된 스팸메일과 정상메일과 기존 해킹메일과의 유사도를 측정하였다.

Table 10. 에서 제시한 해킹메일이 아닌 스팸메일과 정상메일을 실험조건데이터로 활용하여 실제 해킹메일과 유사도 거리를 측정하여 본 논문에서 제시한 알고리즘의 타당성을 제시하고 실제 해킹메일과 일반메일과의 유사도 차이를 식별할 수 있는지 측정할 수 있는 시험을 하였다. 우선 스팸메일은 2014년에 외국메일업체(gmail)을 통해 발송된 스팸메일이며, 정상메일은 2013년 국내 메일업체를 통해서 발송된 일반 정상메일이다.

우선 Fig.7.에 나타난 시험결과를 분석해 보면 2014년 3월 28일, gmail.com에서 발송된 스팸메일을 실험 데이터를 사용했을 때 기존 해킹메일 데이터

IP	ID	Date	Filename / URL	Sender Name	Subject	Domain Name	Character set	Receive ID
175.xx.xx.155	kmdjs12xx	2013-04-23	Panel configuration of Spring		Panel configuration...	daum.net	utf-8	xxx@hanmail.net
175.xx.xx.155	hrpak55xx	2014-03-07	http://hanmailxy-secre		Supination local IP...	daum.net	utf-8	xxx@hanmail.net
175.xx.xx.184	ypfjtrdgxx	2014-07-01	Issues data.hwp	Defense...	FW: Request data	daum.net	utf-8	xxx@daum.net
175.xx.xx.183	noreply-account	2014-03-02	(140302) Attached document	Military ...	Division commande...	daum.net	utf-8	xxx@daum.net
175.xx.xx.155	hrpak55xx	2014-07-01	http://hanmailxy-secreaty-manage	Park00	Educator Resources...	daum.net	utf-8	xxx@daum.net
175.xx.xx.184	ypfjtrdgxx	2014-03-01	Defense Policy Fact Sheet.hwp	Defence...	FW: Defence Policy...	daum.net	utf-8	xxx@daum.net
175.xx.xx.171	noreply-account	2014-03-03	Chinese President Hu...	Foreign...	Chinese President Hu...	daum.net	utf-8	xxx@daum.net
175.xx.xx.188	kaccsxx	2014-03-19	Conference Schedule.hwp		Conference Notification	daum.net	utf-8	xxx@hanmail.net
175.xx.xx.54	mjnglxx	2013-12-16	ional Action Plan (after interagency cons	Lee00	National Action...	kwdimail.re.kr	utf-8	xxx@daum.net
175.xx.xx.249	gdstuxx	2014-07-05	Security Strategy.hwp	Hwang00	Northeast Asian...	hanmail.net	utf-8	xxx@daum.net
175.xx.xx.198	webmail-helper	2014-06-08	http://daummail.dothome.co.kr	Daum Us	Suspicious activity...	hanmail.net	utf-8	xxx@hanmail.net
175.xx.xx.118	pscmwyxx	2014-08-27	http://daummailex.dothome.co.kr	Song00...	Defense Reform Institut...	daum.net	utf-8	xxx@daum.net
42.xx.xx.44	spammail_manaxx	2014-06-05	http://hanmailex-auth-system.bug	Daum	[Urgent!] Is your IP ...	daum.net	utf-8	xxx@hanmail.net
42.xx.xx.135	handaurmtexx	2013-09-10	http://hanmailadmin.my3gb.		Daum mail security...	daum.net	utf-8	xxx@hanmail.net

IP address	ID	Date	Filename / URL	Sender Name	Subject	Domain Name	Character set	Receive ID	Similarity	Accuracy
0	0	0	0	0	0	0	0	0	0	0.00
0	0	0.5	0	0	0	0	0	0	3.5	7.45
0	0	0	0	0	0	0	0	0	0	0.00
0	0	0	0	0	0	0	0	0	0	0.00
0	0	0.5	0	0	0	0	0	0	3.5	7.45
0	0	0.25	0	0	0	0	0	0	1.75	3.72
0	0	0.25	0	0	0	0	0	0	1.75	3.72
0	0	0.5	0	0	0	0	0	0	3.5	7.45
0	0	0.25	0	0	0	0	0	0	1.75	3.72
0	0	0.5	0	0	0	0	0	0	3.5	7.45
0	0	0	0	0	0	0	0	0	0	0.00
0	0	0.25	0	0	0	0	0	0	1.75	3.72
0	0	0.25	0	0	0	0	0	0	1.75	3.72
0	0	0.25	0	0	0	0	0	0	1.75	3.72
0	0	0.25	0	0	0	0	0	0	1.75	3.72

Fig.7. Similarity and accuracy of Spam Mail

Table 10. Spam Mail and Normal Mail Vector

	Spam Mail	Normal Mail
IP address	194.126.2.51	211.239.153.201
ID	formerlyu73	bir
Date	2014. 3. 28	2013. 8. 12
Sender Name	dear	research team
Domain Name	gmail.com	birbook.com
Characterset	en-US	euc-kr
Receive ID	○○○@○○.○○.kr	○○○@naver.com
Filename or URL	Fraud alert document 778-1.zip	data list.pdf
Subject	Attention! Your credit card is being used	loT data list

IP	ID	Date	Filename URL	Sender Name	Subject	Domain Name	Character set	Receive ID
175.xx.xx.207	master1	2013-07-04	http://mail-...		[Korea com Contact Us - Request a Mai...	korea.com	utf-8	xxx@korea.com
42.xx.xx.144	schyong1213	2014-02-26	Diplomatic ...		Program end	daum.net	utf-8	xxx@kndu.ac.kr
123.xx.xx.180	twomir	2013-07-10	Car air-c...	Hwanggyusik	Vehicle air conditioning - very important!	hanmail.net	utf-8	xxx@hanmail.net
175.167.128.68	mail.helper	2014-08-27	http://daummaile...	Seong...	Confidential	daum.net	utf-8	listenheaven@daum.net
31.xx.xx.91	helper_team	2014-02-10	http://mail...	Manager	[Naver Urgent!] Account confirmation...	naver.com	utf-8	xxx@naver.com
103.xx.xx.92	ccth2012	2013-11-15	http://han.a...	Yimye...	Forum progression of time planning...	aol.com	utf-8	xxx@naver.com
42.xx.xx.138	notice-mastar	2014-07-22	http://home...	Daum	[Daum] International IP Login blocked.	daum.net	utf-8	xxx@hanmail.net
42.xx.xx.138	64dmkim	2014-06-04	US Marine...	Sim deokbo	Append the US Marine Corp...	daum.net	utf-8	xxx@naver.com
42.xx.xx.72	schyong1213	2014-02-26	2014 Diplo...	Park Young...	Program end	hanmail.net	utf-8	xxx@kida.re.kr
42.xx.xx.151	post_mail_save	2014-05-06	http://hanmai...		[Urgent!] Is your IP is theft _...	daum.net	utf-8	xxx@hanmail.net
42.xx.xx.218	chinapoli	2013-11-29	Members...		[000] Member contacts send (Final)	daum.net	utf-8	xxx@hanmail.net
175.xx.xx.110	sunlight5678	2014-01-07	2014 Asso...		The [North Korean Studies Association...	daum.net	utf-8	xxx@kndu.ac.kr
175.xx.xx.221	ehdcngud	2014-02-25	Department ...	Kyengho Son	2014 East Strategic Assessment (resend)	naver.com	utf-8	xxx@hanmail.net
175.xx.xx.73	trsaekl7	2014-07-28	http://hanmaili...	Foreign policy	Strategic Resources (North...	daum.net	utf-8	xxx@daum.net
42.xx.xx.217	kaccsbz	2013-11-24	sociation list of participants.	ern Chinese So	Conference Notification	daum.net	utf-8	xxx@hanmail.net

IP address	ID	Date	Filename URL	Sender Name	Subject	Domain Name	Character set	Receive ID	Similarity	Accuracy
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0.75	0	0	0	0	1	0	7.25	15.43
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0	0	0	0	1	1	0	5	10.64
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0.25	0	0	0	0	1	0	3.75	7.98
0	0	0.5	0	0	0	0	1	0	5.5	11.70
0	0	0.5	0	0	0	0	1	0	5.5	11.70
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0.25	0	0	0	0	1	0	3.75	7.98
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0.75	0	0	0	0	1	0	7.25	15.43

Fig.8. Similarity and accuracy of Normal Mail

와의 유사도는 100%로 환산을 때 최고 7.45% 넘지 않는 결과 값을 얻을 수 있다. 그리고 일반 정상메일은 2013년 8월 12일, naver.com에서 발송된 최고 15.43%로 해킹메일과 유사하지 않다는 실험 결과를 확인할 수 있다. 이러한 실험 결과를 분석해 봤을 때, 중요도가 높은 발신지 IP 주소, 발신계정명, 첨부파일명 또는 URL 정보들의 값들이 전혀 유사성이 없는 것을 확인할 수 있으며, 유사도가 최대 9.4가 넘지 않는 것을 확인할 수 있다. 이러한 사항을 추론해 봤을 때 최대 유사도 9.4을 넘지 않는 값을 갖는 메일들은 해킹메일이 아닌 것을 확인할 수 있다.

4.3 시나리오 II - 해킹메일과 유사도 거리 측정

4개 분야의 해킹메일을 선정하여 본 논문에서 제시한 알고리즘을 사용하여 기존의 해킹메일과의 유사성을 검증하고자 한다. 우선 4개 분야의 해킹메일 중 악성코드 첨부파일과 본문내용에 링크 URL 포함형으로 구분하여 종류로 선정하였으며, 각각 비교할 기존의 해킹메일 역시 무작위로 선별하여 유사도를 측정하였다.

아래의 실험결과를 통하여, 4개 분야의 해킹메일들을 논문에서 제시한 알고리즘으로 적용한 결과 32.9 이상 유사도가 나타나는 것으로 확인되어 해킹메일로 확인된다. 이러한 결과를 확인 했을 때 유사도가 14.1 미만일 경우 스팸 또는 정상적인 일반메일로 확인할 수 있으며, 32.9 이상의 유사도 결과가 나타날 경우는 해킹메일로 확인할 수 있다. 이러한 결과들을 종합적으로 분석하여 봤을 때 아래와 같은 유사도 거리 측정 결과를 얻을 수 있다.

Table 11.에서 제시한 유사도 결과를 적용하여 해킹메일 여부를 검증하려고 할 때 14.1 이하의 유사도의 값을 갖는 메일은 일반적 메일로 분류 하고, 14.2 ~ 28.1의 유사도 값을 갖는 메일은 일반메일로 근접하나 다시 한 번 확인을 요하는 메일로 추정할 수 있으며, 28.2 ~ 32.8 사이의 메일은 해킹메일로 의심되어 추가적인 확인이 필요한 메일로 분류, 32.9 이상의 유사도의 값을 갖는 메일은 해킹메일로 분류할 수 있는 결과를 제시할 수 있다.

실험을 통해 각 그룹별 유사도가 높은 해킹메일을 확인할 수 있었으며, 그중 Fig.9.에서의 AA 그룹에 대한 해킹메일에 대한 분석을 해보면 Fig.13.과 같이 공격지 IP 주소와 수신자, 수집서버, C&C 관계가 서로 연관성을 확인 할 수 있으며 해당 해킹메일 공격자는 동일 지역에서 다양한 형태로의 해킹메일을 발송하

Table 11. Hacking Mail Vector

	AA Group Hacking Mail
IP address	175.167.144.0
ID	000
Date	2014. 7. 5
Sender Name	000
Domain Name	daum.net
Characterset	utf-8
Receive ID	000@hanmail.net
Filename or URL	http://hanmailxy-secreaty-manager.yupage.com/daum.htm
Subject	FW : 0000
	BB Group Hacking Mail
IP address	42.97.89.0
ID	000
Date	2013. 11. 26
Sender Name	000
Domain Name	daum.net
Characterset	utf-8
Receive ID	000@hanmail.net
Filename or URL	000list.cell
Subject	000 list
	CC Group Hacking Mail
IP address	
ID	daum.server
Date	2013. 9. 2.
Sender Name	Daum Server
Domain Name	aol.com
Characterset	utf-8
Receive ID	000@hanmail.net
Filename or URL	http://daum.dns1.us/hanmail/modifypw.daum.htm
Subject	Daum member
	DD Group Hacking Mail
IP address	36.97.16.0
ID	songw2020
Date	2014. 5. 23.
Sender Name	000
Domain Name	daum.net
Characterset	utf-8
Receive ID	000@daum.net
Filename or URL	time table.hwp
Subject	hello 000

Table 12. Similarity of Range Result

	14.1 Under	14.2 ~ 28.1	28.2 ~ 32.8	32.9 Over
Case	Normal Mail	Similar to Normal mail	Doubt to Hacking Mail	Hacking Mail

는 같은 공격자 또는 공격집단으로 자료 유출형 공격을 지속적으로 수행하는 공격 형태를 확인 수 있다[11,12].

V. 결 론

본 논문은 해킹메일 공격에 대한 효과적인 탐지와 대응 방안으로 해킹메일 프로파일링을 제안하였다. 해킹메일 공격은 "발신지 IP 주소, 발신 계정명, 발신 일자, 발신 이름, 도메인 이름, 언어 " 등 공격자를 식별할 수 정보를 남기게 되는데, 이러한 정보들의 DB化 및 사례기반추론 알고리즘 적용을 통해 공격성향과 집단의 특성을 프로파일링 할 수 있음을 보여주었다. 실제 해킹메일 공격사례 분석 결과 특정 지역과 집단의 공격의 집중되었음을 확인할 수 있어 프로파일링이 공격 탐지에 효과적임을 알 수 있었다.

보다 정확한 프로파일링을 위해 다량의 다양한 해킹메일의 DB가 요구됨에 따라 적극적인 방식으로 DB化에 접근하는 것이 필요하며 추후 연구에서는 보다 다양해지고 있는 해킹메일 공격에 대한 자료를 수집하고 분석 가능한 좀 더 포괄적인 프로파일링 체계를 제안하는 것이 필요할 것이다. 또한 프로파일링을 기반으로 한 해킹메일 탐지와 공격집단을 식별과 동시에 이에 따른 예방대책까지 제시될 수 있는 체계로까지 연구의 발전방향을 제시하고자 한다.

IP	ID	Date	Filename / URL	Sender Name	Subject	Domain Name	Character set	Receive ID
175,xx,xx,155	kmdjs12xx	2013-04-23	Panel configuration of Spring		Panel configuration...	daum,net	utf-8	xxx@hanmail.net
175,xx,xx,155	hrpak55xx	2014-03-07	http://hanmailxy-secre		Supination local IP...	daum,net	utf-8	xxx@hanmail.net
175,xx,xx,184	ypfjtrdgxx	2014-07-01	Issues data.hwp	Defense...	FW: Request data	daum,net	utf-8	xxx@daum,net
175,xx,xx,183	noreply-account	2014-03-02	(140302) Attached document	Military ...	Division commande...	daum,net	utf-8	xxx@daum,net
175,xx,xx,155	hrpak55xx	2014-07-01	http://hanmailxy-secreaty-manage	Park00	Educator Resources...	daum,net	utf-8	xxx@daum,net
175,xx,xx,184	ypfjtrdgxx	2014-03-01	Defense Policy Fact Sheet.hwp	Defence...	FW: Defence Policy...	daum,net	utf-8	xxx@daum,net
175,xx,xx,171	noreply-account	2014-03-03	Chinese President Hu...	Foreign...	Chinese President Hu...	daum,net	utf-8	xxx@daum,net
175,xx,xx,188	kaccsxx	2014-03-19	Conference Schedule.hwp		Conference Notification	daum,net	utf-8	xxx@hanmail.net
175,xx,xx,54	mjnglxx	2013-12-16	ional Action Plan (after interagency cons	Lee00	National Action...	kwdimail,re,kr	utf-8	xxx@daum,net
175,xx,xx,249	gdstuxx	2014-07-05	Security Strategy.hwp	Hwang00	Northeast Asian...	hanmail,net	utf-8	xxx@daum,net
175,xx,xx,198	webmail-helper	2014-06-08	http://daummail.dothome.co.kr	Daum Us	Suspicious activity...	hanmail,net	utf-8	xxx@hanmail.net
175,xx,xx,118	pscmwyxx	2014-08-27	http://daummailex.dothome.co.kr	Song00...	Defense Reform Institut...	daum,net	utf-8	xxx@daum,net
42,xx,xx,44	spammail_manaxx	2014-06-05	http://hanmailex-auth-system.bug	Daum	[Urgent!] Is your IP ...	daum,net	utf-8	xxx@hanmail.net
42,xx,xx,135	handaurmtexx	2013-09-10	http://hanmailadmin.my3gb.		Daum mail security...	daum,net	utf-8	xxx@hanmail.net

IP address	ID	Date	Filename / URL	Sender Name	Subject	Domain Name	Character set	Receive ID	Similarity	Accuracy
0.5	0	0	0	0	0	1	1	0	9.5	20.21
1	1	0.75	1	0	0	1	1	0	33.25	70.74
0.5	0	1	0	0	0.5	1	1	0	19	40.43
0.75	0	0.75	0	0	0	1	1	0	17	36.17
1	0	1	1	1	0	1	1	0	35	74.47
0.75	0	0.75	0	0	0.5	1	1	0	19.5	41.49
0.5	0	0.75	0	0	0	1	1	0	14.75	31.38
0.5	0	0.75	0	0	0	1	1	0	14.75	31.38
0.5	0	0	0	0	0	0	1	0	6.5	13.83
0.5	0	1	0	0	1	0	1	0	18.5	39.36
0.75	0	0.75	0	0	0	0	1	0	14	29.79
0.5	0	0.5	0	0	0	1	1	0	13	27.66
0	0	0.75	0	0	0	1	1	0	10.25	21.81
0	0	0	0	0	0	1	1	0	5	10.64

Fig.9. Similarity and accuracy of AA Group Hacking Mail

IP	ID	Date	Filename URL	Sender Name	Subject	Domain Name	Character set	Receive ID
210.xx.xx.166	nkjim_xxxx	2012-04-17	Defence Planning Forum.hwp		Defense Forum Request	daum.net	utf-8	xxxx@xx.xx.kr
175.xx.xx.188	kaccsxx	2014-03-19	http://www.kchina.or.kr/Notice/popup.asp?idx=50		Conference Notification	kchina.or.kr	utf-8	xxxx@hanmail.net
42.xx.xx.64	ljw231xxxx	2013-07-16			Vehicle air conditioning - a very major points!	daum.net	utf-8	xxxx@xx.xx.kr
106.xx.xx.197	mailsign_system	2014-01-29	http://hanmailix-auth-interface.bugs3.com/Daum_userprof		[Daum] was being blocked IP Login abroad.	daum.net	utf-8	xxxx@hanmail.net
	adviser_team	2014-02-11	http://mail-naver-login-account.tk/nover/naver_login.htm	Speaker	[Naver] is an emergency notification request mail	naver.com	utf-8	xxxx@naver.com
216.xx.xx.123	black_podo	2011-10-17	Employment Success Stories pagan type.hwp	lee00	The Tae-hyong Yi,	hanmail.net	utf-8	xxxx@lycos.co.kr
175.xx.xx.221	post_mail_sec	2013-12-18	http://hanmailxy-auth-sec.bugs3.com/Daum_userprofile.ht		[Urgent!] Hacked-up has been blocked.	daum.net	utf-8	xx@hanmail.net
	youngxiekxxxx	2013-11-01		Jo00	Invitation (Annual General Meeting)	aol.com	utf-8	xxxxxx@hanmail.net
	daum.server	2013-10-24	http://daum.dns1.us/hanmail/modifypw.daum.htm		Daum members ~~~ desire to change your password!	aol.com	utf-8	xx@hanmail.net
175.xx.xx.180	ydhrbgxx	2014-07-13		Daum Server	Emergency data will be sent to (DoD Deputy S@ng-iu Paek	daum.net	utf-8	xxxxxx@daum.net
	deyangxx	2013-03-13	Focus.pdf		kbs radio broadcasting Koreans start a cardinal	hanmail.net	utf-8	xxxx@xx.xx.kr
59.xx.xx.169	park-min-xx	2014-08-29	Defence Industry Exhibition Brochure (Korean) 8p.pdf	Son00	Defence Industry Exhibition held in 2014 advanced guidance	daum.net	utf-8	xxxx@naver.com
175.xx.xx.195	notice-manager	2014-07-30	http://clean-daum.yypage.com/iahyang55_login.h		[Urgent!] Your username is temporarily blocked.	daum.net	utf-8	xxxx@hanmail.net
137.xx.xx.170	daumservicer	2014-04-28	http://login.daum.net.accountslog inform.jumpingcrab.com/hanmail/lo	Daum	[Daum] Please reconfirm your password?	hanmail.net	utf-8	xxxxxx@hanmail.net
	daum.server	2013-07-29	http://daum.dns1.us/hanmail/modifypw.daum.htm	Daum Server	Daum members ~~~ stabilization care	aol.com	utf-8	xxxx@hanmail.net

IP address	ID	Date	Filename URL	Sender Name	Subject	Domain Name	Character set	Receive ID	Similarity	Accuracy
0	0	0.25	0	0	0	0	1	0	3.75	7.98
0.5	0	0	0	0	0	1	1	0	9.5	20.21
0	0	0.25	0	0	0	0	1	0	3.75	7.98
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0.5	0	0	0	0	1	0	5.5	11.70
0.5	0	0.25	0	0	0	1	1	0	11.25	23.94
0.5	0	0	0	0	0	1	1	0	9.5	20.21
0.5	0	0	0	0	0	0	1	0	6.5	13.83
0.5	0	0	0	0	0	1	1	0	9.5	20.21
1	1	1	0.75	0	0.75	1	1	0	36.75	78.19
0	0	0	0	0	0	1	1	0	5	10.64
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0	0	0	0	1	1	0	5	10.64
0.75	0	0.75	0	0	0	1	1	0	17	36.17

Fig.10. Similarity and accuracy of BB Group Hacking Mail

IP	ID	Date	Filename URL	Sender Name	Subject	Domain Name	Character set	Receive ID
175.xx.xx.207	master1	2013-07-04	http://mail-...		[Korea com Contact Us - Request a Mai...	korea.com	utf-8	xxx@korea.com
42.xx.xx.144	schyong1213	2014-02-26	Diplomatic ...		Program end	daum.net	utf-8	xxx@kndu.ac.kr
123.xx.xx.180	twomir	2013-07-10	Car air-c...	Hwanggyusik	Vehicle air conditioning - very important!	hanmail.net	utf-8	xxx@hanmail.net
175.167.128.68	mail.helper	2014-08-27	http://daummaile...	Seong...	Confidential	daum.net	utf-8	listenheaven@daum.net
31.xx.xx.91	helper_team	2014-02-10	http://mail...	Manager	[Naver Urgent!] Account confirmation...	naver.com	utf-8	xxx@naver.com
103.xx.xx.92	ccth2012	2013-11-15	http://han.a...	Yimye...	Forum progression of time planning...	aol.com	utf-8	xxx@naver.com
42.xx.xx.138	notice-mastar	2014-07-22	http://home...	Daum	[Daum] International IP Login blocked.	daum.net	utf-8	xxx@hanmail.net
42.xx.xx.138	64dmkim	2014-06-04	US Marine...	Sim deokbo	Append the US Marine Corp...	daum.net	utf-8	xxx@naver.com
42.xx.xx.72	schyong1213	2014-02-26	2014 Diplo...	Park Young...	Program end	hanmail.net	utf-8	xxx@kida.re.kr
42.xx.xx.151	post_mail_save	2014-05-06	http://hanmai...		[Urgent!] Is your IP is theft _...	daum.net	utf-8	xxx@hanmail.net
42.xx.xx.218	chinapoli	2013-11-29	Members...		[000] Member contacts send (Final)	daum.net	utf-8	xxx@hanmail.net
175.xx.xx.110	sunlight5678	2014-01-07	2014 Asso...		The [North Korean Studies Association...	daum.net	utf-8	xxx@kndu.ac.kr
175.xx.xx.221	ehdcngud	2014-02-25	Department ...	Kyengho Son	2014 East Strategic Assessment (resend)	naver.com	utf-8	xxx@hanmail.net
175.xx.xx.73	trsaekl7	2014-07-28	http://hanmaili...	Foreign policy	Strategic Resources (North...	daum.net	utf-8	xxx@daum.net
42.xx.xx.217	kaccsbz	2013-11-24	sociation list of participants.	ern Chinese So	Conference Notification	daum.net	utf-8	xxx@hanmail.net

IP address	ID	Date	Filename URL	Sender Name	Subject	Domain Name	Character set	Receive ID	Similarity	Accuracy
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0.5	0	0	0	0	1	0	5.5	11.70
0	0	0.25	0.25	0	0	0	1	0	5.75	12.23
1	0	0.25	0	0	0	0	1	0	12.75	27.13
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0.5	0.25	0	0	0	1	0	7.5	15.96
1	0	0.5	0	0	0	1	1	0	17.5	37.23
1	1	0.5	1	0	0.5	1	1	0	34	72.34
0	0	0	0	1	0	0	1	0	8	17.02
1	0	0	0	0	0	0	1	0	11	23.40
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0	0.25	0	0	0	1	0	4	8.51
1	1	0.5	1	1	1	1	1	0	42.5	90.43

Fig.11. Similarity and accuracy of CC Group Hacking Mail

IP	ID	Date	Filename URL	Sender Name	Subject	Domain Name	Character set	Receive ID
210.xx.xx.166	nkjim_xxxx	2012-04-17	Defence Planning Forum.hwp		Defense Forum Request	daum.net	utf-8	xxxx@xx.xx.kr
175.xx.xx.188	kaccsxx	2014-03-19	http://www.kchina.or.k...		Conference Notification	kchina.or.kr	utf-8	xxxx@hanmail.net
42.xx.xx.64	ljw231xxxx	2013-07-16			Vehicle air conditioning…	daum.net	utf-8	xxxx@xx.xx.kr
106.xx.xx.197	mailsign_system	2014-01-29	http://hanmailix...		[Daum] was being blocked…	daum.net	utf-8	xxxx@hanmail.net
	adviser_team	2014-02-11	http://mail-naver-logi...	Speaker	[Naver] is an emergency…	naver.com	utf-8	xxxx@naver.com
216.xx.xx.123	black_podo	2011-10-17	Employment Success...	lee00	The Tae-hyong Yi,	hanmail.net	utf-8	xxxx@lycos.co.kr
175.xx.xx.221	post_mail_sec	2013-12-18	http://hanmailxy-...		[Urgent!] Hacked-up has been blocked,	daum.net	utf-8	xx@hanmail.net
	youngxiekxxxx	2013-11-01		Jo00	Invitation (Annual…	aol.com	utf-8	xxxxxx@hanmail.net
	daum.server	2013-10-24	http://daum.dns1.u....		Daum members --- …	aol.com	utf-8	xx@hanmail.net
175.xx.xx.180	ydhrbgxx	2014-07-13		Daum Server	Emergency dat…	daum.net	utf-8	xxxxxx@daum.net
	deyangxx	2013-03-13	Focus.pdf		kbs radio broadcasting…	hanmail.net	utf-8	xxxx@xx.xx.kr
59.xx.xx.169	park-min-xx	2014-08-29	Defence Industry...	Son00	Defence Industry Exhi…	daum.net	utf-8	xxxx@naver.com
175.xx.xx.195	notice-manager	2014-07-30	http://clean-daum...		[Urgent!] Your username…	daum.net	utf-8	xxxx@hanmail.net
137.xx.xx.170	daumservicer	2014-04-28	http://login.daum.net....	Daum	[Daum] Please…	hanmail.net	utf-8	xxxxxx@hanmail.net
	daum.server	2013-07-29	http://daum.dns1.us/han...	Daum Server	Daum members ---…	aol.com	utf-8	xxxx@hanmail.net

IP address	ID	Date	Filename URL	Sender Name	Subject	Domain Name	Character set	Receive ID	Similarity	Accuracy
0	0	0	0	0	0	1	1	0	5	10.64
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0.25	0	0	0	0	1	0	3.75	7.98
0	0	0	0	0	0	1	1	0	5	10.64
0	0	0.25	0	0	0	1	1	0	6.75	14.36
0.5	1	1	0	1	1	1	1	0	33.5	71.28
0	0	0.5	0	0	0	0	1	0	5.5	11.70
0.75	0	0.75	0	1	0.75	1	1	0	26.75	56.91
0	0	0.25	0	0	0	0	1	0	3.75	7.98
0	0	0	0	0	0	0	1	0	2	4.26
0	0	0.5	0	0	0	1	1	0	8.5	18.09
0	0	0.75	0	0	0	1	1	0	10.25	21.81
0	0	0.5	0	0	0	0	1	0	5.5	11.70
0	0	0.75	0	0	0	1	1	0	10.25	21.81

Fig.12. Similarity and accuracy of DD Group Hacking Mail

Date	IP address	Sender Name	Subject	Server IP address	ID Owner	Similarity
14-7-5	175.000.000.155	000	FW:000	112.000.000.53	hqd****	
14-7-2	175.000.000.155		IP 000	112.000.000.90		33.25
14-7-2	175.000.000.155	000	data 000			35

Fig.13. Similarity and accuracy of Mail

References

[1] KISA, "2013 National Information Security White Pater," Apr. 2013.

[2] Mee Lan Han, Deok Jin Kim and Kim Huy Kang Kim, "Applying CBR algorithm for cyber infringement profiling system," Journal of The Korea Institute of information Security & Cryptology, 23(6), pp. 1069-1086, Dec. 2013.

[3] Wanju Kim, Changwook Park, Soojin Lee and Jaesung Lim, "Methods for Classification and Attack Prediction of Attack Groups based on Framework of Cyber Defense Operations," The Korean Institute of Information Scientists and Engineers, 20(6), pp. 317-328, Dec. 2013.

[4] Z. Yin, Y. Gao and B. Chen, "On

Development of Supplementary Criminal analysis System Based on CBR and Ontology," Computer Application and System Modeling (ICCASM), 2010 International Conference on, pp, V14-653-V14-655, Oct. 2010.

〔5〕 Changwook Park, Hyunji Chung, Kwangseok Seo and Sangjin Lee, "Research on the Classification Model of Similarity Malware using fuzzy Hash," Journal of The Korea Institute of information Security & Cryptology, 22(6), pp. 1325-1336, Dec. 2012.

〔6〕 US Army, "Open Source Intelligence", Field Manual Interim No. 2-22.9 HQ. Dept. Army, Dec. 2006.

〔7〕 MANDIANT, "APT1 : Exposing One of China's Cyber Espionage Units", http://intelreport.mandiant.com/Mandiant _APT1_Report.pdf, Jan. 2013.

〔8〕 Dmitry Tarakanov, "The "Kimsuky" Operation: A North Korean APT?", http://www.securelist.com/en/analysis/204792305/The_Kimsuky_Operation_A_North_Korean_APT, Sep. 2013.

〔9〕 FireEye, Inc. "Digital Bread Crumbs: Seven Clues To Identifying Who's Behind Advanced Cyber Attacks", http://www.fireeye.com/resources/pdfs/digital -bread-crumbs.pdf, July. 2013.

〔10〕 Daren Kindlund, "CFR Watering Hole Attack Details", FireEye Blog, Jan. 2012.

〔11〕 Joel Yonts, "Building a Malware Zoo," SANS Institute InfoSec Reading Room, Dec. 2010.

〔12〕 Q.Miao, Y.Wang, Y.Cao, X.Zhang, Z.Liu, "APICapture - a Tool for Monitoring the Behavior of Malware," Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering, pp. 390-394, Aug. 2010.

〈저자소개〉

박 형 수 (Hyongsu Park) 정회원
2000년 3월: 전주대학교 국어교육학과 졸업
2013년 9월~현재: 고려대학교 공공보안정책학과 석사과정
〈관심분야〉 침해대응, APT 공격 분석, 보안관제

김 휘 강 (Huy Kang Kim) 종신회원
1998년 2월: KAIST 산업경영학과 학사
2000년 2월: KAIST 산업공학과 석사
2009년 2월: KAIST 산업및시스템공학과 박사
2004년 5월~2010년 2월: 엔씨소프트 정보보안실장, Technical Director
2010년 3월~현재: 고려대학교 정보보대학원 조교수
〈관심분야〉 온라인게임 보안, 네트워크 보안, 네트워크 포렌식

김 은 진 (Eunjin Kim) 정회원
1999년 2월: KAIST 산업경영학과 졸업
2001년 2월: KAIST 경영공학과 석사 졸업
2007년 8월: KAIST 경영공학과 박사 졸업
2008년 9월~현재: 경기대학교 국제산업정보학과 부교수
〈관심분야〉 경영정보시스템, 보안경제학