JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지(J. Korea Inst. Inf. Commun. Eng.) Vol. 20, No. 7 : 1290~1295 Jul. 2016

악성코드 인젝션 사이트 탐지를 통한 방어효율 향상방안

백재종*

Enhanced Method for Preventing Malware by Detecting of Injection Site

Jaejong Baek*

Information and Communication School, Naval Education and Training Command, Changwon, 51699, Korea

요 약

최근 모바일 인터넷 이용률이 급증하면서 인터넷 이용자의 웹 브라우저를 통한 사회 공학적 또는 드라이브 바이 다운로드 방식으로 악성코드 유포 공격이 확산되고 있다. 현재 드라이브 바이 다운로드 공격 방어 초점은 최종 다운로드 사이트 및 유포 경로에 초점을 두어 진행되어 왔으나 공격 초기 악성코드를 주입하는 인젝션 사이트에 대한 특성 탐지 및 차단에 대해서는 충분히 연구되지 않았다. 본 논문에서는 이러한 악성 코드 다운로드 공격에 대한 방어메커니즘 향상을 목적으로, 악성코드 다운로드의 핵심 근원지인 인젝션 사이트를 탐지하는 방안에 대해서 연구한다. 결과적으로 악성코드의 확산을 방지하기 위해 다운로드 공격의 최종 사이트를 탐지 및 차단하는 현재의 URL 블랙리스트 기법에 추가하여, 악성코드를 주입하는 인젝션 사이트를 탐지 특징을 추출 하는 방안을 제시한다. 또한 URL 블랙리스트 기반의 접근법과 비교하여 악성코드 감염률을 효율적으로 최소화 할 수 있는 방안임을 보인다.

ABSTRACT

Recently, as mobile internet usage has been increasing rapidly, malware attacks through user's web browsers has been spreading in a way of social engineering or drive-by downloading. Existing defense mechanism against drive-by download attack mainly focused on final download sites and distribution paths. However, detection and prevention of injection sites to inject malicious code into the comprised websites have not been fully investigated. In this paper, for the purpose of improving defense mechanisms against these malware downloads attacks, we focus on detecting the injection site which is the key source of malware downloads spreading. As a result, in addition to the current URL blacklist techniques, we proposed the enhanced method which adds features of detecting the injection site to prevent the malware spreading. We empirically show that the proposed method can effectively minimize malware infections by blocking the source of the infection spreading, compared to other approaches of the URL blacklisting that directly uses the drive-by browser exploits.

키워드 : 멀웨어, 악성코드, 드라이브 바이 다운로드, 인젝션 사이트, 사회 공학적 공격, 해킹

Key word : Malware, Malicious Code, Drive-by Download, Injection Site, Social Engineering Attack, Hacking

Received 11 March 2016, **Revised** 15 March 2016, **Accepted** 30 March 2016

* **Corresponding Author** Jaejong Baek(E-mail:jbaek7@asu.edu, Tel:+82-55-549-6750)
Information and Communication School, Naval Education and Training Commander, Changwon 51699, Korea

Open Access http://dx.doi.org/10.6109/jkiice.2016.20.7.1290 print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅰ. 서 론

최근 모바일 인터넷 이용률이 급증하면서 인터넷 이용자의 웹 브라우저를 통한 사회 공학적 또는 드라이브 바이 다운로드(drive-by download) 방식으로 악성코드 유포 공격이 확산되고 있다. 사용자는 정상적인 웹사이트로 인식하고 방문했을 뿐인데, 해당 웹사이트가 공격자에 의해 침투당해 악의적인 스크립트를 유포하고 있었다면, 그 사용자는 자신도 모르게 악성코드에 공격당하게 된다. 이런 방법으로 많은 인터넷 사용자가 자신도 모르게 랜섬웨어에 감염되어 자신의 파일들을 모두 인질이 되는 사건이 빈번히 발생되고 있다[1]. 본 논문에서는 이러한 악성 코드 다운로드 공격에 대한 방어메커니즘 향상을 목적으로, 악성코드 다운로드의 핵심 근원지인 인젝션 사이트를 탐지하는 방안에 대해서 연구한다. 먼저 최종 악성코드의 다운로드에 선행하는 경유지(hopping site)를 역 추적하여 피해자가 어떻게 최종 다운로드 사이트에 도달하는지를 분석 한다. 결과적으로 악성코드의 확산을 방지하기 위해 다운로드 공격의 최종 사이트를 탐지 및 차단하는 현재의 URL 블랙리스트 기법에 추가하여, 악성코드를 주입하는 인젝션 사이트를 탐지하는 특징을 추출하여 악성코드 감염률을 최소화 할 수 있는 방안을 제시한다.

Ⅱ. 관련 연구

2.1. 탐지 기법 동향

드라이브 바이 다운로드 공격은 공격코드를 자동으로 생성하는 다양한 익스플로잇 킷을 활용하여 실버라이트, 어도비 플래시, 자바 등 웹 브라우저의 취약점을 통해 악성코드를 자동으로 유포하고 있다[2]. 이에 반해 대부분의 기존 악성코드 탐지시스템 연구 초점은 네트워크 관점에서의 URL 블랙 리스트 방식에 두고 있다.

도메인(URL) 기반의 정적블랙리스트를 이용하는 도메인 평판체계[3]는 도메인에 대한 악성 점수를 제공하여 사용자가 악의적인 웹사이트를 방문하는 것을 차단한다. 각 사이트에 대한 역할이나 기능을 정의하지는 않는다. 정적 블랙리스트의 대다수 도메인은 자주 변경되기 쉬운 악성 익스플로잇 명칭, 다운로드 도메인들 목록이다. 최근에는 시그니처 기반의 백신의 제한사항을 보완한 실행 가능한 평판시스템이 제안되었다[4]. 이는 탐지 기반을 콘텐츠 특징을 이용하는 것이 아니라 악성코드의 분배 구조 특성에 초점을 두었다. 하지만 이는 사용자가 어떻게 악의적인 실행파일을 다운받았는지 등의 정보는 제공하지 못한다.

클라이언트 허니팟[5]은 웹사이트에 방문하여 시스템 변경을 관찰하거나 악의적인 콘텐츠에 대한 반응분석을 통하여 드라이브 바이 다운로드를 탐지하나 수집할 수 있는 웹사이트 수에 제한이 있고 속도가 느린 단점이 있다. Static crawlers[6]은 휴리스틱 방법을 이용하여 악의적으로 판단되는 웹의 콘텐츠를 필터링하여 탐지한다. 또한 알려진 악의적인 콘텐츠와 유사한 것을 포함하는 웹사이트를 식별하기 위해 검색엔진을 활용한다.

리디렉션 체인에 초점을 둔 그래프화 접근은 브라우징 행위기반 트리 기법[7]이 있으며 Referrer, Location 헤더와 URL를 이용하여 그래프화 하여 해당 사이트의 행위가 악의적인지 판단한다. Referrer 헤더는 하이퍼링크를 사용하여 액세스한 경우에는 직전에 참조한 웹사이트를 나타내며 Location은 웹사이트를 표시한 직후에 재차 다른 웹사이트로 이동할 경우에 새로운 웹사이트를 지정하는 것을 의미한다(Redirection). 유사한 방법으로 WarningBird[8]는 트위터에 게시된 악의적인 웹사이트를 식별하는데 리디렉션 체인과 추출된 특징으로 악성유무를 구분한다.

이상과 같은 기존 방어기법은 악의적인 코드가 삽입된 근원지 URL을 차단한 것 보다는 주로 실제 드라이브 바이 익스플로잇 또는 악성코드 다운로드가 가능토록 하는 공격 사이트의 URL을 차단하는 것에 의존한다. 이와 달리 본 연구에서는 다운로드의 원인을 분류하고 공격에 포함된 도메인의 역할과 기능을 식별하여 어떻게 사용자가 공격사이트에 유도되었는지를 분석하는 방법을 제안한다. 본 기법은 URL 블랙리스트 기반 접근방안에 추가하여 드라이브 바이 다운로드공격의 피해율을 최소화할 수 있다.

2.2. 사회 공학적 vs. 드라이브 바이 다운로드 공격 비교

악성코드 다운로드를 개시하도록 버튼 클릭 등과 같은 명시적인 사용자의 개입이 포함된 경우 이를 사회 공학적 공격이라 정의한다. 이와 달리 브라우저 익스플

로잇을 통하여 사용자 모르게 또는 허락 없이 악성코드가 다운로드 되는 것을 드라이브 바이 다운로드 공격이라 한다.

Table. 1 Feature of Sociable Engineering and Drive-by download attack

Feature	Sociable Engineering	Drive-by
User interaction	○	×
exploit URL similarity	×	○
Download Domain Recurrence	○	×
Referrer header	○	×
User-Agent Popularity	○	○

표 1은 각 공격의 특성을 비교한 것으로 드라이브 바이 다운로드의 대부분은 익스플로잇 키트에 의해 제공된다. 따라서 대부분 드라이브 바이 다운로드 경로에 포함된 익스플로잇 전송 URL 들은 알려진 익스플로잇 키트 URL과 유사하다. 또한 악성 코드 다운로드를 제공하는 대부분의 도메인은 거의 통신하지 않고 공격의 한 시점에 특정 클라이언트와 통신한다. 이러한 특성을 기반으로 드라이브 바이 공격에서는 통신 빈도가 거의 발생하지 않지만 사회 공학적 공격에서는 악성코드 다운로드를 위한 무료파일 공유사이트로 사용되기 때문에 통신 빈도가 높게 발생한다. 사회 공학적 공격의 경우, 악의적인 파일 다운로드를 전달하는 HTTP 트랜잭션은 Referrer 헤더를 포함하는 경향이 있다. 이는 보통 그것들을 특징화시키는 직접적인 사용자 인터액션이기 때문이다. 반면에 드라이브 바이 공격의 악성코드 파일 전송은 브라우저 익스플로잇을 경유하여 발생한다. 일반적으로 쉘코드(shell code)로 부터 초기화되는 요청은 Referrer 헤더가 없다. 또한 각 공격의 다운로드 경로는 전형적으로 몇 개의 노드를 포함한다. 이는 피해자가 공격에 이르도록 브라우저를 사용할 때 일반적인 브라우저 User-Agent 스트링을 발생하는 공통점을 가지고 있다. 본 논문에서는 상기 연구동향에 따라 공격 특징이 명확하고 사용자의 심리적인 판단에 따른 사회 공학적 공격보다는 사용자가 미인지적으로 공격당하여 탐지하기 어려운 드라이브 바이 다운로드 공격에 대해서 분석한다.

Ⅲ. 제안하는 탐지 강화 기법

3.1. 악성코드 다운로드 경로 구분

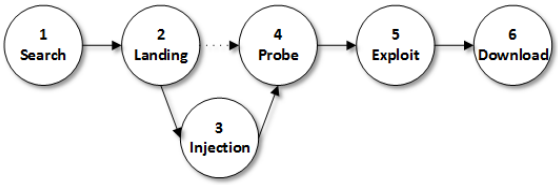

Fig. 1 Download Path stages

일반적으로 악성코드 다운로드 공격 경로를 분석할 때 경유하는 사이트 구분은 검색사이트로 시작되는 것을 가정할 때 그림 1과 같이 랜딩, 인젝션, 검사, 익스플로잇, 다운로드 사이트로 구분될 수 있다. 각 사이트의 역할과 기능은 다음과 같다[9].

3.1.1. 랜딩 사이트(Landing Site)

랜딩사이트는 드라이브 바이 공격경로가 시작되는 웹사이트이며 사전에 공격자에 의해 장악되었으나 정상적으로 운영되는 사이트이다. 보통 접속하는 사용자 수가 많고 많은 관심을 유도할 수 있는 사회적 이슈 등을 반영하는 웹 사이트들이 악용된다. iframe 등과 같이 지정된 사이트로 리디렉션 하는 코드나 대중적 혹은 피공격자의 맞춤형 이슈에 대한 링크를 악성 사이트로 연결하는 코드가 삽입된다.

3.1.2. 인젝션 사이트(Injection Site)

해킹된 랜딩사이트에 주입되는 악성코드의 근원이 되는 사이트로 다음 경유 사이트로 중계 해주는 역할을 한다. 관리가 부실한 웹 사이트들이 이용 되거나 공격자가 운영하는 사이트일 수 있으며 iframe 등과 같이 지정된 사이트로 리디렉션 하는 코드가 삽입 된다. 일반적인 경유사이트와 차이는 경유사이트의 시작 사이트라는 점이다.

3.1.3. 취약점 검사 사이트(Probing Site)

공격자가 운영하는 사이트이며 피공격자의 브라우저에 설치된 자바, 실버라이트, 플래쉬, 아도비 리더 등 플러그인 소프트웨어에 대한 버전과 취약점을 검사하여 이에 적합한 공격 익스플로잇을 결정하는 사이트다.

3.1.4. 익스플로잇 사이트(Exploit Site)

공격자가 운영하는 사이트이며 피공격자자의 시스템의 취약점을 이용하는 콘텐츠를 포함하는 사이트로 사용자 단말기 및 어플리케이션의 취약성을 공격하는 익스플로잇 코드가 삽입된다.

3.1.5. 유포 사이트(Distribution Site)

유포 사이트는 사용자 단말기에 악성코드를 설치하는 사이트로서 공격자가 직접 운영하는 경우와 해킹된 웹 사이트가 악용되는 경우가 있다. 피공격자가 접속하면 자동으로 악성코드(파일) 다운로드하여 설치하면서 사용자에게 설치 동의를 구하는 절차가 없다.

3.2. 기존 연구의 문제점

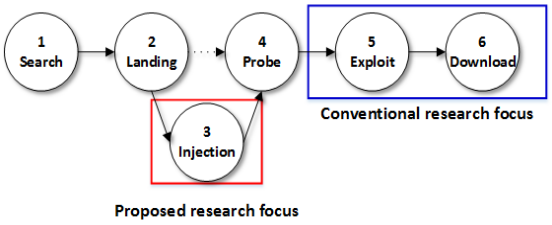

Fig. 2 Download Path stages

블랙리스트 또는 행위 추적 기반의 기존 방어기법은 주로 익스플로잇 사이트의 악성코드 자체 URL을 차단하는 것에 초점을 두어 다수 연구가 진행되었으나 악성코드 유포의 근원 역할을 하는 인젝션 사이트에 초점을 둔 탐지 및 차단 연구는 활성화되지 못했다. 그림 2에서 보듯이 5번, 6번 사이트에 대해서 악성코드나 URL을 탐지하여 차단 규칙에 포함한 것이 기존 블랙리스트 기반의 기법이었다. 이런 방법으로는 악성코드를 삭제하고 접속을 차단해도 공격자가 처음 사이트를 해킹하도록 허용한 근본적인 취약성은 해결되지 않는다.

또한 근본 원인을 해결하지 않으면 이후에 사이트가 다시 해킹당할 수 있다. 따라서 본 연구에서는 유포의 근원이 되는 3번 인젝션 사이트를 식별하기 위하여 인젝션 사이트의 특성을 분석하고 이를 기반으로 탐지 시그니처를 추가했을 때 전체적인 방어효율이 개선될 수 있음 보인다.

3.3 인젝션 사이트 특성 추출

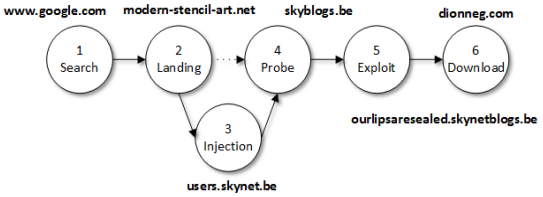

Fig. 3 Sample Analysis of drive-by Download Path

그림 3은 [10]에 공개된 악성 웹 도메인 목록(Immortal Malware Domains)중 한 사이트를 역추적 분석하여 각 경유 사이트의 기능을 식별한 결과를 나타내고 있다. 리디렉션 체인을 재구성하여 전체 웹 경로를 재구성하고 각 사이트 간 관계를 규명하였다. 이렇게 반복적인 분석을 100건을 수행한 결과 몇 가지 인젝션 노드가 갖는 공통적인 특징을 추출해낼 수 있었다. 그림 4는 분석 과정에서 인젝션사이트가 다른 사이트로 리디렉트하는 스크립트 일부를 나타낸다.

```
<iframe frameborder="0" height="0"
  src="http://skyblogs.be"
  style="display:none" width="0"></iframe>
……
<script type='text/javascript'
src='http://modern-stencil-art.net/xx/xxx.js'></script>
……
<script>
  if (document.referrer.match(/xxxxx\.com/)) {
    window.location("http://modern-stencil-art.net//"); }
</script>
```

Fig. 4 Redirect script source sample

인젝션 사이트 특성 분석을 위해 [10] 사이트에 공개된 악성 도메인 목록을 분석한 결과 다음과 같은 인젝션 사이트의 특성을 추출하였다.

① 랜딩사이트의 후속 사이트일 경유
② 광고와 같이 랜딩 사이트에 악성 코드를 주입하는 경우
③ 사이트를 오픈하고 검색엔진에 노출이 되기 시작한 도메인 에이지가 얼마 되지 않는 경우
④ 후속사이트 개수가 유일하거나 소수인 경우
⑤ 랜딩 사이트와 다른 도메인인 경우

분석 결과 대부분 인젝션 도메인은 랜딩 도메인과 상

이하며 랜딩사이트의 식별은 Mekky[7]의 기법을 활용하여 식별하였다. 인젝션 사이트는 iframe 태그가 코딩된 랜딩사이트의 소스 코드를 가지고 있는 저장소가 되며 웹브라우저의 취약성을 검사하는 프로빙 사이트나 유사한 또 다른 사이트로 리디렉션을 진행한다. 따라서 광고와 같은 특성을 가지고 있어 구분하기 힘들지만, 랜딩사이트 다음에 올 경우는 99% 인젝션 사이트로 식별되었다. 또한, 일반적으로 합법적인 사이트 소유자에 의해 악성코드가 노출될 수 있어 해킹된 사이트를 직접적으로 조정하지는 않는다.

인젝션 사이트는 해커가 호스팅한 신규 도메인으로 에이지가 작은 특성이 있다. 리디렉션 수는 대개의 경우 하나를 가지는 경향이 있지만 때에 따라서는 그 이상을 가지고 있을 수 있다. 이러한 수치는 특정 악성코드 도메인 경우 판단될 수 있는 기준치가 된다.

3.4. 탐지 흐름도

상기 5가지 특성을 가지고 인젝션 사이트를 탐지하는 흐름도는 다음 그림 5와 같다. 각 특성의 비교 우선순위는 분석 결과에 따른 종속관계를 의미한다. 스크립트가 난독화 된 경우 해독하는 절차와, URL이 실행파일(.EXE, .BAT 등) 경우의 악성코드여부를 확인하는 절차는 단순화를 위해 생략하였다. 이러한 방법들은 기존 연구에서 이미 많이 알려져 있다. 랜딩사이트면서 코드를 주입하는 사이트는 인젝션사이트가 확실하며, 리디렉션 수가 특정 수(N) 이상이며 도메인 에이지가 특정 수(D) 이하 일 경우 의심되는 사이트로 판단되어 재분석을 수행한다. 랜딩사이트와 동일한 도메인일 경우에도 의심되는 사이트로 판단되어 재분석을 수행한다. 특정수는 악성코드 도메인별로 상이함으로 본 흐름도 에서는 변수로 정의하였다.

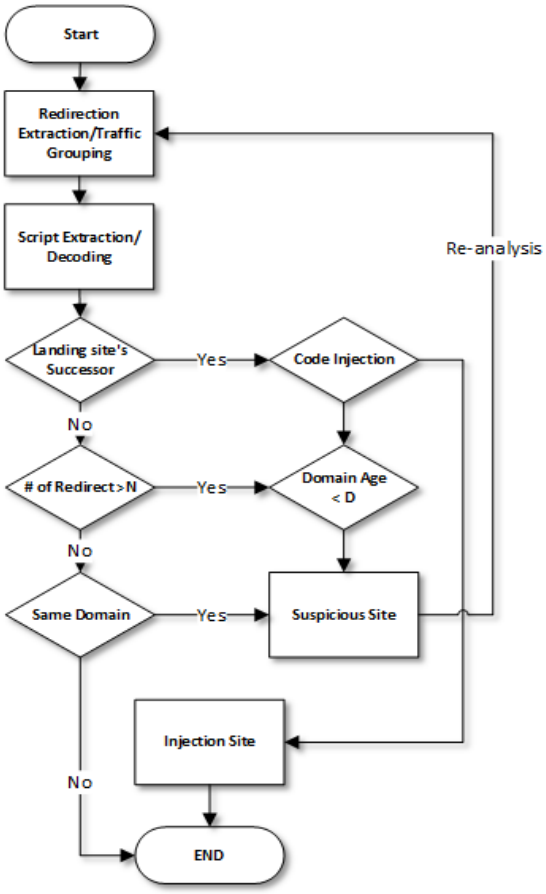

Fig. 5 Flow chart for Detecting Injection Site

Table. 2 Results in Comparing with Products

Feature	GSB	Eset	Proposed
Landing	38(76%)	26(52%)	41(82%)
Injection	30(60%)	0(40%)	44(88%)
Probing	38(76%)	26(52%)	39(78%)
Exploit	38(76%)	27(54%)	43(86%)
Download	38(76%)	25(50%)	44(88%)

Ⅳ. 성능평가

인젝션 사이트 탐지 특성의 성능을 측정하기 위해 인터넷 악성도메인 정보 사이트 [11] 를 참고하여 50개를 매칭 및 분석했다. 먼저 드라이브 바이 다운로드 공격의 경로 구분되는 각 단계별 탐지율을 구글의 세이프브라우징(GSB), ESET Smart Security와 제안한 기법과 비교한 결과는 표 2와 같다.

각 제품과의 비교상의 참고사항은 제안하는 기법을 적요할 때는 수동으로 웹 소스와 트래픽을 직접 분석하였기 때문에 모든 단계에서의 탐지율이 상대적으로 높게 나왔다고 볼 수도 있다. 이는 본 연구 목적상 인젝션 사이트 탐지에 대한 효율성을 보이기 위한 것임으로 향후 현재의 수동 분석 기법을 자동화하여 객관적으로 분석 비교하는 것이 필요하다.

제안하는 탐지기법은 다른 제품보다 특히 인젝션 사이트 탐지율 효율이 28 ~ 48% 높게 나옴을 알 수 있다. 표 3에서는 50개의 도메인을 분석할 때 인젝션 사이트를 탐지해내는 각 특성의 확률을 나타내고 있다. 이것은 각 특성이 인젝션 사이트를 판별하는 기준의 우선순위로 사용된다. 즉, 인젝션 사이트의 경우는 1위부터 5위순으로 특성을 가질 확률이 많다고 볼 수 있다.

Table. 3 Feature Ranking

Rank	Relative importance	Feature
1	100%	① Successor of Landing site
2	90%	② Code Injection
3	85%	③ # of Node's Successor
4	76%	④ Domain Age
5	43%	⑤ Same Domain

V. 결 론

현재 악성코드 연구동향은 드라이브 바이 다운로드 탐지 시 최종 다운로드 사이트 및 유포 경로에 초점을 두어 진행되어 왔으나 공격 초기 랜딩사이트의 후속 사이트인 인젝션 사이트에 대한 특성 탐지 및 차단에 대해서는 잘 알려지지 않았다.

본 논문에서는 악성코드 유포 근원지가 되는 인젝션 사이트를 탐지하는 특성을 다수의 공격경로를 분석 및 통계적으로 추출하여 악성코드 감염공격에 대한 방어 효율을 높이는 방안을 제시하였다. 본 연구에서는 수동분석으로 샘플 50개의 악성코드를 적용하였지만 자동화된 도구로 대규모 네트워크에 일정기간 적용하여 실제적인 피해감소율을 측정 및 비교하여 효율성에 대한 세부적인 연구를 향후 추진해야할 예정이다.

REFERENCES

[1] The Register' article. [Internet]. Available : http://www.theregister.co.uk/2016/03/09/trend_micro_ransomware_iot_threat_rise/

[2] Boan news's article [Internet]. Available: http://www.boannews.com/media/view.asp?idx=46385.

[3] M. Antonakakis, et al., "Detecting Malware Domains at the Upper DNS Hierarchy," *In USENIX Security*, vol. 11. pp. 1-16, 2011.

[4] P. Vadrevum et al., "Measuring and detecting malware downloads in live network traffic," In *ESORICS*. pp. 556-573, 2013.

[5] J. Nazario, et al., "A virtual client honeypot," In *Proceedings of the 2nd USENIX Conference on LEET.*, vol 9, pp 911-919, 2009.

[6] N. Provos, et al., "The ghost in the browser analysis of webbased malware," In *Proceedings of the First Conference on First Workshop on HotBots*, vol 7, pp 4-13, 2007.

[7] H. Mekky, et al., "Detecting malicious http redirections using trees of user browsing activity," In *INFOCOM*, pp. 1159-1167, 2014.

[8] S. Lee, et al., "A near real-time detection system for suspicious urls in twitter stream," *IEEE Trans. Dependable Secur. Comput.* vol. 10, no. 3, pp. 183-195, May 2013.

[9] N. Terry, et al,. "WebWitness: Investigating, Categorizing, and Mitigating Malware Download Paths," In *USENIX Security 15*, pp. 1025-1040, 2015.

[10] malware domains list. [Internet]. Available : http://mirror1.malwaredomains.com/files/immortal_domains.txt

[11] sample malicious domain list. [Internet]. Available : https://zeustracker.abuse.ch/blocklist.php?download=domainblocklist

백재종(Jaejong Baek)

1996년 2월 한밭대학교 전자계산학과 공학사
2001년 2월 연세대학교 컴퓨터과학과 공학석사
2011년 8월 연세대학교 컴퓨터과학과 공학박사
2014년 1월 ~ 2016년 6월 해군정보통신학교 학부장
2015년 9월 ~ 2016년 8월 호원대학교 국방과학기술학부 겸임교수
2016년 8월 ~ 애리조나주립대 컴퓨터과학과 박사후 연구원
※관심분야 : 무선 보안, 네트워크 보안, 시스템보안, 사이버전, 역공학