Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.24, NO.6, Dec. 2014 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2014.24.6.1185

효율적인 보안관제 수행을 위한 다크넷 트래픽 기반 악성 URL 수집 및 분석방법 연구*

김 규 일,[1†] 최 상 수,[1] 박 학 수,[1] 고 상 준,[1,2] 송 중 석[1,2‡]
[1]한국과학기술정보연구원, [2]과학기술연합대학원대학교

A Study on Collection and Analysis Method of Malicious URLs Based on Darknet Traffic for Advanced Security Monitoring and Response*

Kyu-il Kim,[1†] Sang-soo Choi,[1] Hark-soo Park,[1] Sang-jun Ko,[1,2] Jung-suk Song[1,2‡]
[1]Korea Institute of Science and Technology Information,
[2]Korea University of Science & Technology

요 약

국내·외 해킹공격 전담 대응조직(CERTs)들은 침해사고 피해 최소화 및 사전예방을 위해 탐지패턴 기반의 보안장비 등을 활용하여 사이버공격에 대한 탐지·분석·대응(즉, 보안관제)을 수행하고 있다. 그러나 패턴기반의 보안관제체계는 해킹공격을 탐지 및 차단하기 위해 미리 정의된 탐지규칙에 근거하여 알려진 공격에 대해서만 대응이 가능하기 때문에 신·변종 공격에 대한 대응은 어려운 실정이다. 최근 국내·외에서는 기존 보안관제의 이러한 문제점을 극복하기 위해 다크넷이라는 기술을 활용한 연구가 주목을 받고 있다. 다크넷은 미사용 중인 IP주소의 집합을 의미하며, 실제 시스템이 존재하지 않는 다크넷으로 유입된 패킷들은 악성코드에 감염된 시스템이나 해커에 의한 공격행위로 간주 될 수 있다. 따라서 본 연구에서는 효율적인 보안관제 수행을 위한 다크넷 트래픽 기반의 악성 URL 수집 및 분석방법을 제안한다. 제안방법은 국내 연구기관의 협력을 통해 확보한 8,192개(C클래스 32개)의 다크넷으로 유입된 전체 패킷을 수집하였으며, 정규표현식을 사용하여 패킷에 포함된 모든 URL을 추출하고 이에 대한 심층 분석을 수행하였다. 본 연구의 분석을 통해 얻어진 결과는 대규모 네트워크에서 발생하고 있는 사이버 위협상황에 대한 신속·정확한 관측이 가능할 뿐만 아니라 추출한 악성 URL을 보안관제에 적용(보안장비 탐지패턴, DNS 싱크홀 등)함으로서 해킹공격에 대한 사이버위협 대응체계를 고도화하는데 목적을 둔다.

ABSTRACT

Domestic and international CERTs are carrying out security monitoring and response services based on security devices for intrusion incident prevention and damage minimization of the organizations. However, the security monitoring and response service has a fatal limitation in that it is unable to detect unknown attacks that are not matched to the predefined signatures. In recent, many approaches have adopted the darknet technique in order to overcome the limitation. Since the darknet means a set of unused IP addresses, no real systems connected to the darknet. Thus, all the incoming traffic to the darknet can be regarded as attack activities. In this paper, we present a collection and analysis method of malicious URLs based on darkent traffic for advanced security monitoring and response service. The proposed method prepared 8,192 darknet space and extracted

접수일(2014년 9월 11일), 수정일(2014년 10월 13일), 게재확정일(2014년 10월 24일)
* 본 연구는 2014년도 미래창조과학부의 수탁사업 「과학기술사이버안전센터 구축 및 운영사업」의 지원을 받아 수행된 연구임 (G-14-GM-IR02)
† 주저자, kisados@kisti.re.kr
‡ 교신저자, song@kisti.re.kr(Corresponding author)

all of URLs from the darknet traffic, and carried out in-depth analysis for the extracted URLs. The analysis results can contribute to the emergence response of large-scale cyber threats and it is able to improve the performance of the security monitoring and response if we apply the malicious URLs into the security devices, DNS sinkhole service, etc.
Keywords: Darknet, Security Monitoring and Response, Malicious URLs

I. 서 론

국내·외 해킹공격 전담 대응조직(CERTs)들은 침해사고 피해 최소화 및 사전예방을 위한 탐지패턴 기반의 보안장비 등을 활용하여 사이버 공격에 대한 탐지·분석·대응을 수행하고 있다. 그러나 패턴기반의 보안관제체계는 해킹공격을 탐지 및 차단하기 위해 미리 정의된 탐지규칙에 근거하여 알려진 공격(스캐닝, DDoS, 웜·바이러스)에 대해서만 대응이 가능할 뿐 알려지지 않은 새로운 해킹 공격에 대한 대응은 부족한 실정이다.

최근 국내·외에서는 이미 알려진 침해공격 및 사후 시에만 대응 가능한 기존 보안관제체제의 문제점을 해결하기 위해 다크넷(Darknet) 기술[1][4][8][12]을 적용하여 신·변종 악성행위 및 침해공격을 조기에 탐지하고자 하는 기술이 주목을 받고 있다.

다크넷은 미사용 중인 IP주소로 전송된 의심패킷들을 감시·분석하는 기술이다. Fig.1.은 다크넷의 기본 원리를 나타내며 공격자는 목표 시스템이 존재할 것이라 추측되는 IP대역에 악성트래픽을 보낸다. 악성트래픽은 스캔, 악성코드 및 악성URL 등이 포함되는 알려진 공격방법이거나 알려지지 않은 공격트래픽일 수 있다.

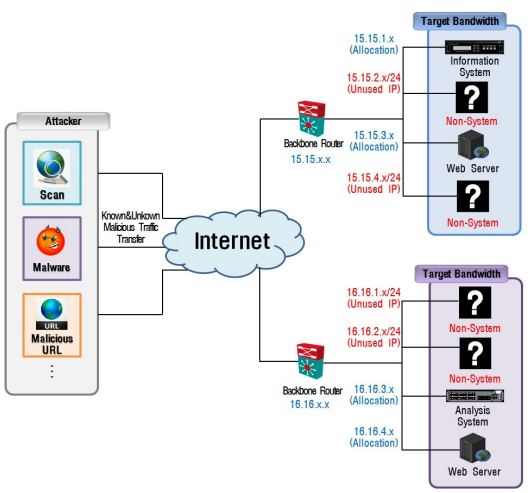


Fig. 1. Basic Concept of Darknet

만약 공격자가 목표 시스템을 초기에 발견하였을 경우, 해당 시스템의 취약점을 이용한 단계로 발전하게 되며 반대로 발견하지 못하였을 경우 공격자는 사전에 조사한 IP대역에 분명히 목표시스템이 존재할 것이라 예측하고 악성 트래픽을 지속적으로 보내는 행위를 하게 된다.

다크넷은 후자의 경우를 착안하여 해커들이 노리는 주요 네트워크 구간에 실제 시스템이 존재하지 않은 유휴IP 대역을 설정하여 이들로부터 유입된 공격자의 악성 트래픽을 관측 및 분석함으로서 특정 패킷만을 수집·분석하는 허니팟과 달리 대규모 네트워크에서 발생하는 전반적인 침해시도 현황을 파악할 수 있을 뿐만 아니라 기존 패턴기반에서 탐지가 불가능 하였던 신·변종 공격에 대한 행위를 사전에 탐지할 수 있다.

따라서 본 연구는 효율적인 보안관제체계를 위해 다크넷 트래픽기반의 악성 URL 수집 및 분석방법을 제안한다. 제안방법은 국내 연구기관의 협력을 통해 확보한 8,192개(C클래스 32개)의 다크넷으로 유입된 전체 패킷을 수집하였으며, 정규표현식을 사용하여 패킷에 포함된 모든 URL을 추출하고 이에 대한 심층 분석을 수행하였다.

본 연구의 분석을 통해 얻어진 결과는 대규모 네트워크에서 발생하고 있는 사이버 위협상황에 대한 신속·정확한 관측이 가능할 뿐만 아니라 추출한 악성 URL을 보안관제에 적용(보안장비 탐지패턴, DNS 싱크홀 등)함으로서 해킹공격에 대한 사이버위협 대응체계를 고도화하는데 목적을 둔다.

본 논문의 구성은 다음과 같다. 2장에서는 악성행위 공격행위를 탐지하기 위해 현재 진행되고 있는 기법들을 소개하고 3장은 제안 방법의 원리와 구성에 대해 기술한다. 4장에서는 제안 방법의 분석결과를 제시하고 5장에서는 본 논문의 최종 결론을 맺는다.

II. 관련 연구

본 장에서는 기존 패턴기반의 보안관제의 문제점을 보완하기 위해 현재 진행되고 있는 허니넷, DNS 싱

크홀, 악성코드 분석 및 그레이넷 연구들을 소개한다.

2.1 허니넷(Honey-Net)

허니넷[3][6][7]은 공격자의 관심을 유발하는 정보자원을 제공하여 침입을 유도하기 위한 네트워크를 의미하며 허니넷을 구성하고 있는 개별의 시스템을 허니팟이라 일컫는다. 허니넷의 목적은 보안상 취약점을 가진 정보시스템을 구축하여 공격자의 악성행위 정보 및 공격코드를 수집하기 위한 시스템이다. Fig.2.는 허니넷의 구성을 나타내며 방화벽 안에 이기종의 허니팟을 구축하여 해당 시스템으로 유입되는 악성 트래픽을 수집할 수 있는 기본적인 환경을 보여준다.

현재 허니넷의 기술수준은 하나의 시스템 상에서 다수의 운영체제를 설치 및 실행 가능한 가상환경기반의 허니넷이 주목을 받고 있다. 가상 허니넷은 비용절감과 유지관리가 용이하다는 장점을 가지고 있는 반면 허니팟에 유입되는 특정 패킷만을 수집하기 때문에 네트워크 전반에 대한 분석을 수행할 수 없는 단점을 지니고 있다. 또한 해당 시스템이 공격자들로 하여금 허니팟이라는 인지를 할 수 없도록 세심한 운영·관리가 요구된다.

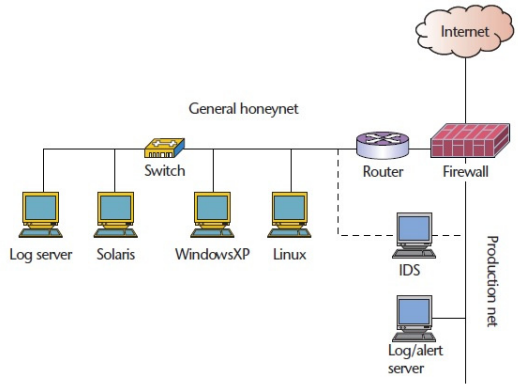


Fig. 2. Basic Concept of Honeynet

2.2 DNS 싱크홀

DNS 싱크홀[2][9][13]은 C&C 서버나 악성 도메인을 알고 있는 경우 좀비 PC의 DNS 질의에 대한 응답을 조정하여 C&C 서버로 접속할 시 DNS 싱크홀 서버로 우회시켜 공격 명령을 차단하는 기법이다. Fig.3.은 DNS 싱크홀 구축 전·후의 동작과정을 나타낸다. 우선, 왼쪽 화면은 좀비 PC가 DNS로 C&C 주소의 정보를 질의할 경우 DNS는 지령서버의 주소를 알려주고 좀비 PC는 C&C 서버로 접속이 이루어지게 된다. 그러나 우측 그림처럼 DNS 싱크홀을 적용할 경우 DNS 서버에 저장된 싱크홀 서버의 주소를 대신 전달하여 C&C와의 연결을 차단할 수 있다.

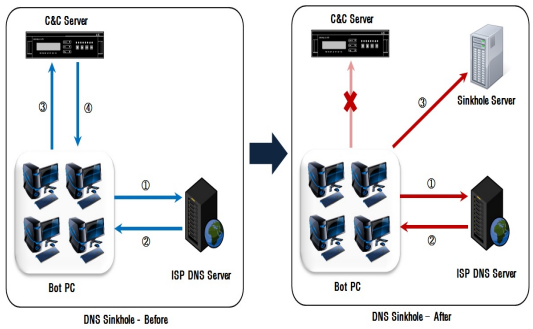


Fig. 3. Basic Concept of DNS Sinkhole

DNS 싱크홀은 기존 네트워크 시스템에 별다른 수정 없이 적용할 수 있으며 시스템 도입 비용이 저렴한데 비해 차단효과가 우수하여 현재 부문보안관제센터에서 활용하고 있다. 그러나 DNS 싱크홀은 좀비PC와 C&C 사이의 통신을 차단할 뿐 좀비PC의 취약점이나 악성코드를 제거할 수 없으며, 악성 도메인 또는 URL을 주기적으로 업데이트를 해야 한다는 문제점을 안고 있다.

2.3 악성코드 분석

최근 연구 및 산업계에서는 악성코드 수집을 위한 시스템을 자체적으로 구축하여 이들 행위를 분석하는 연구[5][10][11]들이 대내·외로 소개되고 있으며 분석도구 역시 점차 다양해지고 있다. 현재 주요 악성코드 분석동향은 해당 코드를 언팩킹하여 바이너리 코드 및 어셈블리 코드분석을 통해 특징(API, 명령어, 문자열, Byte, Block 등)을 추출하여 악성여부를 판별하는 정적분석 방법과 가상머신 상에서 악성코드를 직접 실행하여 API 호출 hooking 및 네트워크 패킷분석을 통해 코드의 특징을 추출하여 악성여부 및 신·변종 여부를 판별하는 동적분석 방법을 이용하고 있는 추세이다.

또한, 다양한 악성행위 정보(악성코드, 트래픽, 시스템, 로그, 악성 웹사이트, IDS 로그 등)에 대한 상관분석을 통해 감염경로, 실행형태, 공격대상, 공격행

위 및 위협수준 등 보다 정확하게 판별할 수 있는 방법들도 제안되고 있다.

이처럼 악성코드 분석이 정교해지고 다양한 형태의 코드를 분석할 수 있는 환경을 갖추어 가고 있으나 분석이전에 새로운 악성코드에 대한 지속적인 수집을 위해 수집환경 개선 및 확대방안 등이 선행되어야 하며 악성코드의 분석 정확성뿐만 아니라 신속성을 고려한 체계 역시 필요한 시점이라 하겠다.

2.4 그레이넷(Gerynet)

그레이넷(Greynet)[14][15]은 특정 기관에서 보유중인 IP 대역 중 사용하지 않은 유휴 IP를 기반으로 악성행위 정보를 수집하는 네트워크이다. 그레이넷의 일반적인 구조는 Fig.4.와 같으며 미 할당 IP 대역을 통해 정보를 수집한다는 점에서 다크넷과 유사한 개념으로 사용되지만 그레이넷은 추가적인 IP를 확보할 필요가 없다는 점에서 다크넷과 차이점을 둔다. 또한 그레이넷은 수집된 정보를 통해 네트워크상에서 발생하는 사이버 공격에 대한 관측이 가능하다는 장점을 가지고 있다.

그러나, 그레이넷은 오직 할당되고 남은 소수의 IP만을 정보수집에 활용하기 때문에 대규모 네트워크에서 발생하는 해킹공격 및 이상행위 징후를 정확히 판

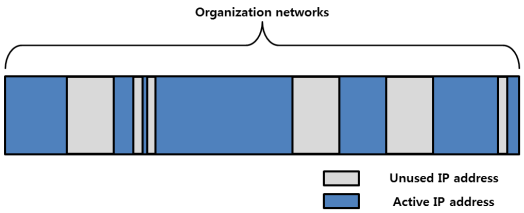


Fig. 4. Structure of Greynet

별하기 어렵다는 한계점을 지닌다.

III. URL 수집·분석을 위한 다크넷 시스템 구성

3.1 제안 시스템 구성

본 연구에서는 다크넷에 유입된 패킷을 관측·감시하기 위해서 국내 연구기관의 협력을 통해 확보한 8,192개(C클래스 32개) IP를 바탕으로 URL 수집·분석 시스템을 구축하였다. Fig.5.는 제안시스템의 프레임워크를 나타내며 주요 동작과정은 다음과 같다.

① 공격자는 타겟 시스템을 찾기 위해 네트워크 전반에 걸쳐 악성 트래픽을 전송한다.

②~③ 만약 악성 트래픽이 다크넷 대역으로 유입되었을 경우 해당 트래픽은 TAP 장비를 통해 수집 시스템으로 전송되며 수집된 패킷은 각 형식에 맞게

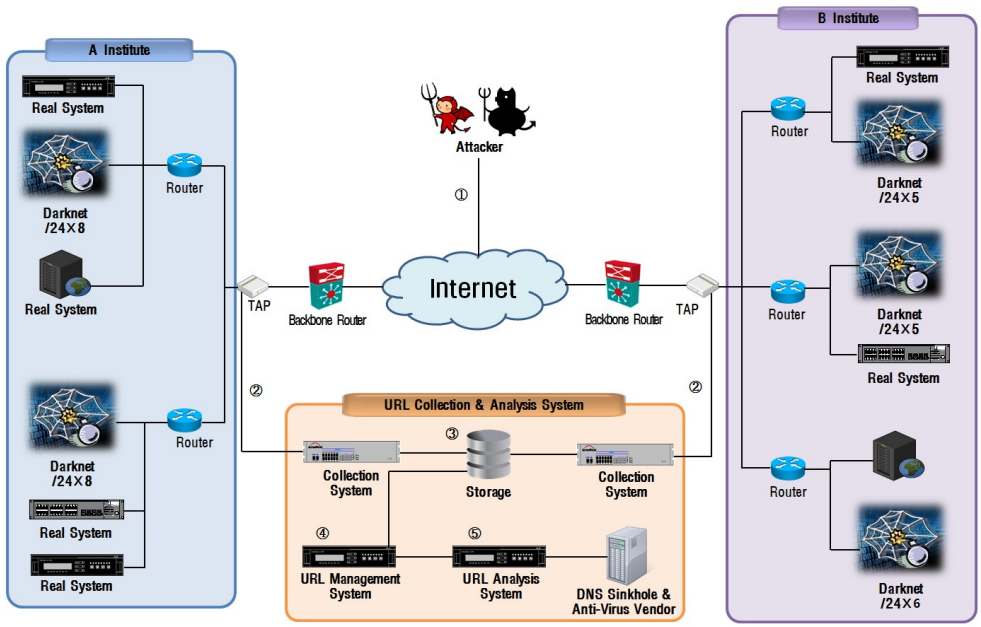


Fig. 5. Framework of Proposed System

스토리지에 저장된다.

④ URL 관리시스템은 저장된 다크넷 패킷 중에서 정규표현식을 이용하여 URL이 포함된 패킷만을 추출한다.

⑤ 추출된 URL은 데이터 정규화 과정을 거쳐 이와 밀접하게 연관된 DNS 정보를 기준으로 분류한 후 URL 분석을 수행한다.

3.2 수집시스템

수집시스템은 다크넷 대역으로 전송된 의심패킷들을 수집하기 위함이다. Fig.6.은 다크넷 수집시스템의 패킷형식을 나타내며 IP 헤더, 식별자(센서ID), 패킷 수신시간 및 다크넷 패킷(IP헤더, 이더넷 헤더, 프로토콜 헤더, 패킷 데이터)으로 구성된다.

Fig.7.은 다크넷 수집시스템의 순서도를 보여주며 동작과정은 다음과 같다.

① 각 지역망(A, B)의 수집시스템은 할당된 다크넷 주소정보를 설정파일로부터 불려온다.

② 수집시스템은 패킷캡쳐 핸들러를 생성 및 활성화하여 해당 네트워크에서 송·수신되는 모든 패킷에 대한 모니터링을 수행한다.

③ 설정된 다크넷 주소로 유입되는 패킷들을 수집하고 수집된 패킷을 스토리지에 저장한다.

IP header [40 byte]	Sensor ID [2 byte]	Receiving Time [4 byte]	Ethernet header [16 byte]	IP header [40 byte]	Protocol header	Data

Fig. 6. Packet Types of Darknet Collection System

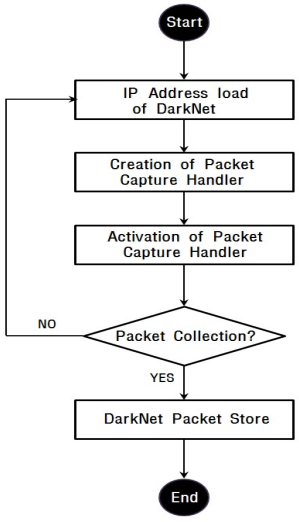


Fig. 7. Flowchart of Darknet Collection System

3.3 URL 관리시스템

URL 관리시스템은 스토리지에 저장된 다크넷 패킷 중에서 URL을 포함하는 데이터만을 추출하여 정규화를 수행하는 역할을 한다. 다크넷 패킷은 우선, 수집 프로토콜에 의해 수집된 16진 HEX값을 문자열 식별이 가능한 ASCII코드로 변환한다. 변환된 패킷은 영문자, 숫자 및 특수문자 외에도 많은 복수 조합들로 구성되어 있기 때문에 본 논문에서는 정규표현식을 사용하여 페이로드 안에 URL만을 정확히 추출한다.

만약 URL이 존재하는 패킷인 경우 효율적인 URL 분석을 위해 Fig.8.과 같이 8가지(전송시간, 출발지IP, 출발지 포트, 도착지 IP, 도착지 포트, 프로토콜, URL 및 페이로드) 타입으로 정규화한다. 정규화된 데이터는 URL 분석시스템에 전송되어 해당 URL의 악성여부를 판별하게 된다.

Fig.9.는 URL 관리시스템의 순서도를 나타내며 동작과정은 다음과 같다.

① 스토리지에서 저장된 다크넷 패킷을 불러온다.

Time	Src_ip	Src_port	Dst_ip	Dst_port	Protocol	url	Payload
2014-08-08 01:01:01	aa.bb.cc.dd	1234	bb.cc.dd.e	4567	UDP	www.xx.com	1EAB3E...

Fig. 8. Format of darknet packets

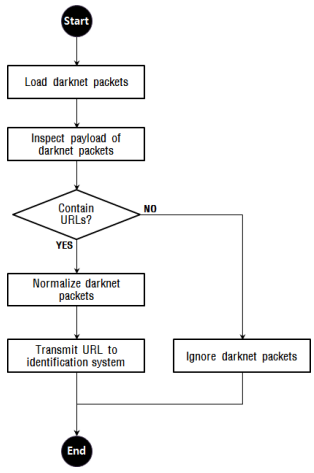


Fig. 9. Flowchart of Darknet URL Management System

② 정규표현식을 사용하여 URL을 포함하는 다크넷 패킷을 선별한다.

③ 다크넷 패킷의 페이로드 중에 URL을 추출하여 악성 URL 분석시스템으로 전송할 데이터를 정규화한다.

④ 정규화된 데이터를 악성 URL 분석 시스템으로 전송한다.

3.4 URL 분석시스템

URL 분석시스템은 추출된 URL를 분석하여 다크넷으로 유입된 원인과 현재 대규모 네트워크에서 발생하고 있는 위협상황에 대한 결과를 도출하기 위함이다. Fig.10.은 URL 분석시스템의 순서도를 나타내며 동작과정은 다음과 같다.

① 분석시스템은 URL 관리시스템으로부터 정규화된 URL 패킷을 수신한다.

② 분석시스템은 URL 패킷 중 Port 정보를 기준으로 가장 빈번하게 이용되는 패킷을 추출한다.

③ 대부분의 URL이 DNS 53포트를 사용하여 송·수신됨에 따라 분석시스템은 DNS 정보를 포함하는 URL과 DNS 정보를 포함하지 않은 URL로 구분한 후 분석을 수행한다.

④ 우선, DNS 정보를 포함하고 있는 URL이 악성 URL인지 여부를 판별하기 위해 DNS 싱크홀 및 안티 바이러스 벤더(Virus-Total)社의 블랙 도메인과 비교한 후 정상 및 악성 URL로 분류한다.

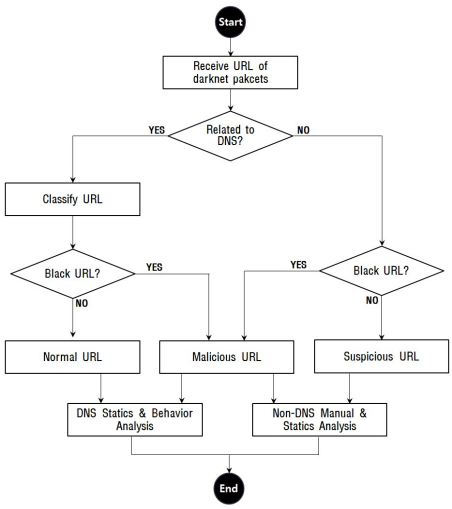


Fig. 10. Flowchart of Darknet URL Analysis System

⑤ 분석시스템은 URL 분류 데이터를 토대로 통계 및 행위분석을 통해 다크넷으로 유입된 원인과 현재 대규모 네트워크에서 발생하고 있는 위협상황에 대한 결과를 도출한다.

⑥ 반면 DNS를 포함하지 않은 URL은 위와 동일하게 악성 URL 여부를 확인하여 악성 및 의심스러운 URL로 분류한다.

⑦ 분류된 URL를 토대로 통계 및 수동분석을 통해 다크넷으로 유입된 배경과 현재 네트워크에서 발생하고 있는 이상징후에 대한 결과를 도출한다.

IV. 다크넷을 활용한 URL 분석

4.1 다크넷 URL 통계분석

본 연구에서는 다크넷에 유입된 패킷을 대상으로 URL을 추출하여 이들에 대한 통계분석을 수행하였다. 우리는 우선, 악성 URL을 정확히 분석 및 판별하기 위해 14년 8월 1일부터 14년 8월 18일까지 다크넷에서 수집된 최신 데이터를 바탕으로 분석을 수행하였다. Fig.11.은 수집된 다크넷 패킷에서 URL을 포함하지 않은 패킷과 URL을 포함하는 패킷의 분포차트를 보여준다. 먼저, URL 미포함 패킷은 전체의 약 98.7% (69,284,425개)로 다크넷의 유입된 대부분의 패킷이 URL를 미포함하고 있으며 반대로, 약 1.3%(930,587개)의 패킷만이 URL를 포함하는 것으로 나타났다.

우리는 URL를 포함하고 있는 패킷을 대상으로 다시 통계분석을 실시하였다. Fig.12.는 URL이 포함된 패킷 중에서 DNS 패킷과 DNS와 관련되지 않은 패킷의 분포차트를 보여준다. DNS 패킷은 출발지 및 도착지 포트가 53번인 경우에 해당되며 전체의 약 99.9% (929,671개)를 차지하고 있는 반면, DNS와 관련이 없는 패킷은 약 0.1%(916개)에 불과한 것으로 통계결과 나타났다.

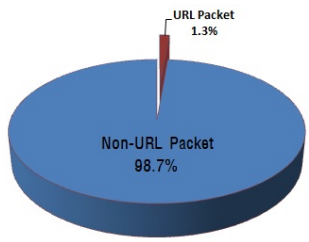


Fig. 11. Distributions of Darknet Packet

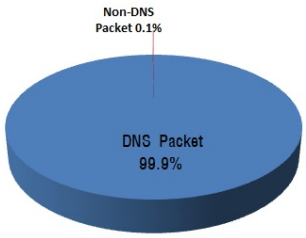


Fig. 12. Distributions of Darknet DNS Packet including URL

4.2 URL을 포함하는 DNS 패킷분석

4.2.1 출발지 DNS Port 분석

우리는 URL을 포함하는 DNS 패킷 중에서 출발지와 목적지 포트를 구분하여 분석을 수행하였다. 먼저, 출발지 포트가 53번인 패킷은 전체의 약 7%(67,576개, 1,290개 고유URL)로 목적지 포트(93%)보다 상당히 적은 수치를 기록하였다. Table.1은 해당 포트에 대한 상위 10위까지 URL를 보여주며 보안을 위해 URL 중간에 *를 기입하였다. 상위10 URL에 대한 비율은 전체패킷의 약 86%에 해당되며 이들은 일반적으로 사용자가 질의하지 않은 URL로서 해커에 의해 감염된 좀비PC들이 다크넷 IP로 스푸핑[1](Spoofing)하여 타겟 DNS서버로 해당 URL을 이용한 DDoS 공격을 시도했던 것으로 분석된다.

위의 분석결과를 뒷받침하는 요인으로 Table.1.의 8번과 9번의 URL인 경우, 바이러스 토탈(Virus Total)의 조회결과 악성 URL인 것으로 판명되었다. 또한, Table.2.는 다크넷으로 패킷을 전송한 이력이 있는 출발지 IP 순위를 나타낸 것이다. 이들 상위 10개 IP 모두 53번 포트로 접근이 가능하고 현재까지 DNS 서비스를 제공한 것으로 볼 때 DNS 서버로 사용되고 있음을 짐작할 수 있다.

Fig.13.은 해당 분석에 대한 공격과정을 보여준다.

① 우선, 공격자는 사용자가 빈번하게 이용하는 포털사이트 및 P2P사이트에 악성코드를 유포하여 해당 바이러스에 취약한 컴퓨터를 감염시킨다.

② 공격자는 감염된 컴퓨터를 완전히 장악하여 좀비PC로 만든 다음 다크넷 IP 및 그 주변의 네트워크 IP로 스푸핑하여 타겟 DNS 서버에 도메인을 질의한다.

③ 타겟 DNS는 순간적으로 많은 양의 도메인 질의를 받게 되며 또한, 요청 도메인이 자신에게 등록되지 않은 URL이라는 것을 확인하고 루트 DNS에게 질의를 요청하게 된다.

1) IP 스푸핑(IP Spoofing) : IP 프로토콜의 보안 취약점을 악용한 것으로 자신의 IP 주소를 변조하여 접속하는 방법

Table. 1 Top10 DNS(Source port 53) URL

Rank	URL(Source Port 53)	Number	Ratio
1	help.da****.com	38,281	56.65%
2	www.pinw****.com	8,041	11.90%
3	yctestweb.clo*****.net	2,690	3.98%
4	www.xs***.com	2,064	3.85%
5	www.jieyi****.com	2,108	3.12%
6	www.ad***.com	1,384	2.05%
7	www.longbufe****.com	975	1.44%
8	www.bai***.com	721	1.07%
9	www.ffe***.com	711	1.05%
10	www.kane***.cn	647	0.96%
Result Sum		58,162	86.07%

Table. 2 Top10 DNS(Source port 53) IP

Rank	IP(Source Port 53)	Number	Ratio
1	***.***.123.58	5,917	8.76%
2	***.***.123.221	5,853	8.66%
3	***.***.122.148	4,178	6.18%
4	***.***.122.221	4,148	6.14%
5	***.***.123.35	3,261	4.83%
6	***.***.123.37	3,235	4.79%
7	***.***.235.252	2,958	4.38%
8	***.***.235.251	2,933	4.34%
9	***.***.122.58	2,784	4.12%
10	***.***.122.35	2,754	4.08%

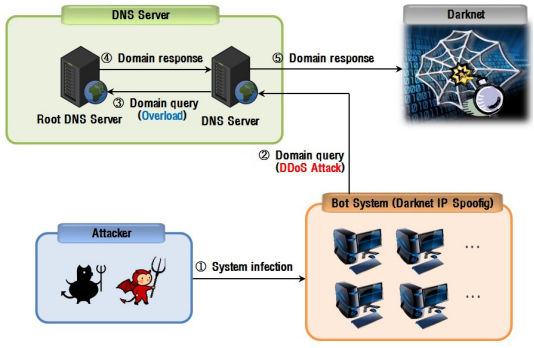


Fig. 13. Procedures of DNS DDoS Attacks

④~⑤ 루트 DNS는 질의에 대한 해당 결과를 리턴하게 되며 타겟 DNS는 원래의 IP 대역인 다크넷으로 응답패킷을 전송하게 된다.

결과적으로 공격자는 타켓 DNS의 서비스 제공을 막기 위해 DDoS 공격을 시도하였으며 일반적이지 않은 도메인을 질의하여 루트 DNS 서버까지 부하를 발생시키려는 행위를 하였음을 분석을 통해 알 수 있었다.

4.2.2 목적지 DNS Port 분석

URL을 포함하는 DNS 패킷 대부분이 목적지 포트가 53번이었으며 전체의 약 93%(862,095개, 6,746개 고유URL)을 차지하고 있는 것으로 나타났다. Table.3은 해당 포트에 대한 상위10위까지 URL를 보여주며 이들 10위까지의 패킷은 전체의 약 71%(608,668개)에 해당된다. 각 URL를 살펴보면 1번과 9번을 제외한 대부분이 사용자가 일반적으로 사용하는 URL인 것으로 확인되었다.

또한, Fig.14.는 다크넷 대역 중 해당 포트에서 패킷을 수신한 IP 현황을 그래프로 나타낸 것이다. 그래프를 살펴보면 구축한 다크넷 전체 네트워크에서 1번과 2번을 제외한 나머지 IP에서 일정 비율의 패킷을 수신한 것을 볼 수 있다. 이처럼 위의 사항을 종합하였을 때, 해커는 타겟 DNS 서버를 찾기 위해 대규모 네트워크를 대상으로 정상적인 URL를 사용하여 스캔공격을 시도하던 중 해당 네트워크에 포함된 다크넷 대역까지 공격패킷이 유입된 것으로 분석된다.

Fig.15.은 해당 분석에 대한 공격과정을 보여준다.

① 공격자는 사용자가 빈번하게 이용하는 사이트에 악성코드를 유포하여 사용자의 PC를 제어 가능한 좀비PC로 만든다.

② 생성된 좀비PC는 타겟 DNS 서버를 찾기 위해

Table. 3 Top10 DNS(Destination port 53) URL

Rank	URL(Destination Port 53)	Number	Ratio
1	dnsscan.shadow*****.org	169,870	19.70%
2	www.goog**.com	160,118	18.57%
3	www.goo***.it	143,317	16.62%
4	webpan**.sk	63,802	7.40%
5	try.ikri**y.cu.cc	20,269	2.35%
6	cs.washing***.edu	12,735	1.48%
7	cen***.gov	11,979	1.39%
8	wradish.com	9,775	1.13%
9	ddosfor***.pw	8,485	0.98%
10	net****.icsi.ber****.edu	8,318	0.96%
Result Sum		608,668	70.60%

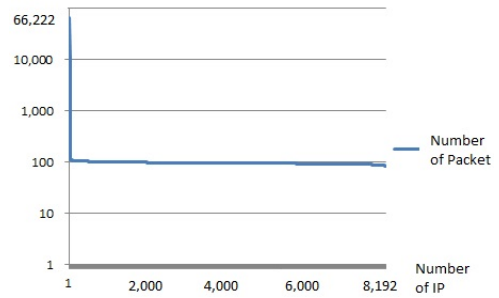


Fig. 14. Number of packet for DNS(Destination port 53) IP

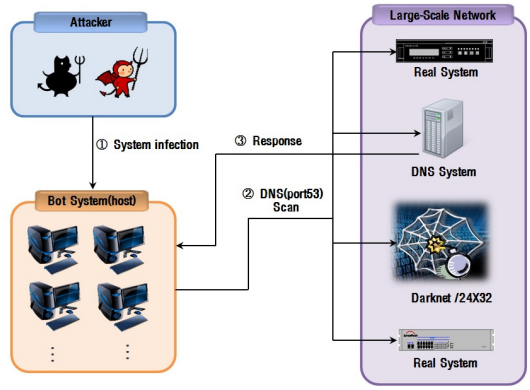


Fig. 15. Procedures of DNS Scan Attacks

광범위한 네트워크 구간에 53포트를 이용하여 공격 패킷을 전송하게 되며 다크넷으로 설정해 놓은 구간역시 해당 패킷이 유입된다.

③ 타겟 DNS 서버는 공격패킷에 대한 응답을 주게 되며 공격자는 응답 패킷을 통해 타겟 DNS 서버의 IP를 획득하게 된다.

4.3 Non-DNS URL 패킷분석

우리는 DNS 정보를 포함하지 않은 URL 패킷(916개) 중에서 중복제거를 통해 Table.4.와 같이 16개의 고유 URL를 추출하였다. 이들 URL를 토대로 분석을 수행한 결과 약 44% (1, 5, 8, 9, 12, 14 및 15번)가 악성 URL인 것으로 판명되었으며 3개의 프로토콜(ICMP, TCP 및 UDP)이 사용되었다.

먼저, ICMP 패킷의 경우 해당 패킷에 포함된 실제 페이로드의 목적지 포트가 53번인 점을 감안할 때 특정 서버를 찾기 위한 스캔행위를 시도하였던 것으로 분석된다.

둘째로, TCP 패킷의 경우 패킷의 플래그 특성이

Table. 4 Non-DNS URL Extraction List

Rank	Extraction URL	Contents & Protocol
1	www.ff***op.com	Malicious URL
2	web.asd***.cu.cc	UDP
3	domain.ikri***.cu.cc	UDP
4	www.wgaam***.net	UDP
5	m.root-ser****.net	Malicious URL
6	a.root-ser****.net	UDP
7	we***.com	ICMP
8	I.root-ser****.net	Malicious URL
9	g.root-ser****.net	Malicious URL
10	www.uc1***.com	ICMP
11	www.momo***.com	ICMP
12	www.wan****.com	Malicious URL
13	5.xyr***.com	ICMP
14	yun***.com	Malicious URL
15	www.baid***.com	Malicious URL
16	gwaewgni.youd***.com	ICMP

연결 요청에 대한 응답(Syn+Ack)임을 확인할 수 있었다. 이는 공격자가 다크넷 대역으로 IP를 스푸핑하여 타겟 시스템으로 공격 패킷을 보내는 행위를 시도하였던 것으로 분석된다.

끝으로, UDP 패킷의 경우 패킷에 포함된 URL의 웹 서버 IP 이력을 조회한 결과 경유지 및 악성코드 전송 서버로 사용된 것을 확인할 수 있었다.

V. 결 론

본 연구는 효율적인 보안관제체계를 위한 다크넷 트래픽기반의 URL 수집 및 분석방법을 제안하였다. 제안방법은 다크넷(8192개 IP)에 유입된 패킷을 대상으로 정규표현식을 사용하여 URL이 포함된 패킷만을 추출하였다. 추출된 URL은 정규화 과정을 거쳐 가장 밀접하게 연관된 DNS 포트를 기준으로 분류한 후 URL분석을 수행하여 아래와 같이 의미 있는 분석결과를 도출하였다.

첫째로 출발지 포트가 53번인 URL 패킷들은 해커에 의해 감염된 좀비PC들이 다크넷 IP 대역으로 스푸핑한 후 해당 URL를 이용하여 타겟 DNS 서버로 DDoS공격을 시도한 것으로 나타났다.

둘째로, 목적지 포트가 53번인 URL 패킷들은 해커가 대규모 네트워크를 대상으로 타겟 DNS 서버를 찾기 위해 해당 URL를 사용하여 스캔공격을 시도하던 중 다크넷 대역까지 공격패킷이 유입된 것을 알 수 있었다.

끝으로, DNS 정보를 포함하지 않은 URL 패킷들은 거의 대부분이 악성 URL 이거나 해킹공격에 사용되는 것으로 분석되었다.

이처럼 다크넷을 통해 기존에 알려지지 않은 신·변종 공격을 사전에 관측 및 탐지가 가능한 이유는 해커는 목표 시스템에 대한 실제 해킹공격을 수행하기 전에 해킹공격의 성공유무를 미리확인하기 위해 해당 시스템과 유사하거나 취약성이 있는 시스템을 타겟으로 설정하여 테스트를 진행하게 된다.

따라서 우리는 이들 테스트 구간에 다크넷을 구축하여 해당 대역으로 유입된 패킷들을 수집·분석함으로써 현재 대규모 네트워크에서 발생하는 최신 위협동향을 관측할 수 있게 되었으며 관측 데이터를 기반으로 해킹공격을 사전에 예측하는 새로운 시각을 제시하였다.

본 연구결과는 과학기술사이버안전센터(S&T-SEC)를 통해 실시간 보안관제 및 침해대응 서비스에 적극 활용할 계획이며 향후 연구로는 일본정보통신연구원(NICT)간의 국제협력을 통해 상당수의 다크넷 대역을 확보하여 글로벌 사이버위협에 대한 분석을 수행할 예정이다.

References

[1] Eto, M., Inoue. D., Song, J., Nakazato, J., Ohtaka, K., and Nakao, K., "nicter : A Large-Scale Network Incident Analysis System," Proc. of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security(BADGERS '11), pp. 37-45, Apr. 2011.

[2] Kim, H., Choi, S., and Song, J., "A Methodology for Multipurpose DNS Sinkhole Analyzing Double Bounce Emails," Proc. on ICONIP 2013, LNCS 8226, pp. 609-616, Nov. 2013.

[3] Spitzner, L., "The Honeynet Project: trapping the hackers," Magazine of Security & Privacy, IEEE pp.15-23, Mar. 2003.

[4] Choi, S., Kim, S., and Park, H., "A Fusion

Framework of IDS Alerts and Darknet Traffic for Effective Incident Monitoring and Response," Journal of Applied Mathematics & Information Science, pp.245-251, Dec. 2013.

[5] Egele, M., Scholte, T., Kirda, E., and Kruegel, C., "A survey on automated dynamic malware-analysis techniques and tools," Journal of ACM Computing Surveys (CSUR) Vol. 44, Issue 2, Feb. 2012.

[6] Abbasi, F., H, and Harris, R. J., "Experiences with a Generation III virtual Honeynet," Proc. of the Telecommunication Networks and Applications Conference(ATNAC'09), pp.1-6, Nov. 2009.

[7] Abbasi, F., H, and Harris, R. J., "Intrusion detection in Honeynets by compression and hashing," Proc. of the Telecommunication Networks and Application Conference (ATNAC'10), pp.96-101, Nov. 2010.

[8] Bailey, M., Cooke, E., Jahanian, F., Provos, N., Rosaen, K., and Watson, D., "Data Reduction for the Scalable Automated Analysis of Distributed Darknet Traffic," Proc. of the 5th ACM SIGCOMM conference on Internet Measurement(IMC'05), pp 239-252, Oct. 2005.

[9] Lee, H., Choi, S., Lee, Y., and Park, H., "Enhanced Sinkhole System by Improving Post-processing Mechanism," Proc. on FGIT 2010, LNCS 6485, pp. 469-480, Dec. 2010.

[10] Willenms, C., Holz, T., and Freiling, F., "Toward Automated Dynamic Malware Analysis Using CW Sandbox," Journal of IEEE Security and Privacy, Vol 5, Issue 2, Mar. 2007.

[11] Qiu, H., and Osoro F. C. C., "Static malware detection with Segmented Sandboxing," Proc. of 8th International Conference on the Malicious and Unwanted Software (MALWARE'13), pp. 132-141, Oct. 2013.

[12] Nakao, K., Inoue, D., Eto, M., and Yoshioka, K., "Practical Correlation Analysis Between Scan and Malware Proles Against Zero-day Attacks Based on Darknet Monitoring," Journal of IEICE Transactions on Information and System E 92D(5), pp.787-798, Dec. 2009.

[13] Kim, Y., and Youm, H., "A New Bot Disinfection Method Based on DNS Sinkhole," Journal of the Korea Institute of Information Security & Cryptology vol.18, no.6, pp. 107-114, Dec. 2008.

[14] Harrop, W., Armitage, G., "Defining and Evaluating Greynets(Sparse Darknets)," Proc. of the IEEE conference on Local Computer Networks 30th Anniversary(LCN'05), pp. 344-350, Nov. 2005.

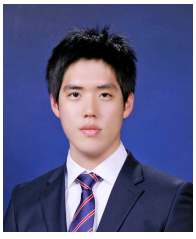
[15] Harrop, W., Armitage, G., "Gerynets: a definition and evaluation of sparsely populated darknets," Proc. of the ACM SIGCOMM workshop on Mining network data(MineNet'05), pp. 171-172, Aug. 2005.

〈저 자 소 개〉

김 규 일(Kyu-il Kim) 정회원
2005년 2월: 성균관대학교 컴퓨터공학과 석사
2010년 2월: 성균관대학교 컴퓨터공학과 박사
2010년 6월~현재: 한국과학기술정보연구원 과학기술정보보호실 선임연구원
〈관심분야〉 보안관제, 침해사고대응, 악성코드 분석

최 상 수 (Sang-soo Choi) 정회원
2001년 2월: 한남대학교 컴퓨터공학과 졸업
2003년 2월: 한남대학교 컴퓨터공학과 석사
2006년 2월: 한남대학교 컴퓨터공학 박사
2006년 2월~현재: 한국과학기술정보연구원 과학기술정보보호실 선임연구원
〈관심분야〉 정보보호, 보안관제, 침해사고대응

박 학 수 (Hark-soo Park) 정회원
1989년 2월: 한남대학교 전자계산학과 졸업
1991년 2월: 한남대학교 컴퓨터공학과 석사
2003년 2월: 한남대학교 컴퓨터공학 박사
1991년 3월~현재: 한국과학기술정보연구원 과학기술정보보호실 책임연구원
〈관심분야〉 정보보호, 보안관제, 침해사고대응



고 상 준 (Sang-jun Ko) 학생회원
2013년 2월: 한국항공대학교 정보통신공학 졸업
2013년 3월~현재: 과학기술연합대학원대학교 그리드 및 슈퍼컴퓨팅 석사과정
〈관심분야〉 정보보호, 네트워크 보안, 악성코드 분석

송 중 석 (Jung-suk Song) 정회원
2003년 2월: 한국항공대학교 통신정보공학 졸업
2005년 2월: 한국항공대학교 정보공학 석사
2009년 3월: 교토대학교(일본) 지능정보학 박사
2009년 4월~2010년 9월: 일본정보통신연구원 정보통신 보안 연구소 전문연구원
20010년 10월~2011년 9월: 일본정보통신연구원 네트워크 보안 연구소 선임연구원
2011년 10월~현재: 한국과학기술정보연구원 과학기술정보보호실 선임연구원
2012년 9월~현재: 과학기술연합대학원대학교 그리드 및 슈퍼컴퓨팅 부교수
〈관심분야〉 보안관제, 침해사고대응, 악성코드 분석, 네트워크 보안