Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.27, NO.4, Aug. 2017 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2017.27.4.821

다크넷 트래픽의 목적지 포트를 활용한 블랙 IP 탐지에 관한 연구*

박 진 학,[1†] 권 태 웅,[1] 이 윤 수,[1] 최 상 수,[1] 송 중 석[1,2‡]
[1]한국과학기술정보연구원, [2]과학기술연합대학원대학교

A Study on Detecting Black IPs for Using Destination Ports of Darknet Traffic*

Jinhak Park,[1†] Taewoong Kwon,[1] Younsu Lee,[1] Sangsoo Choi,[1] Jungsuk Song[1,2‡]
[1]Korea Institute of Science and Technology Information
[2]Korea University of Science & Technology

요 약

인터넷은 우리나라의 경제·사회를 움직이는 중요한 인프라 자원이며 일상생활의 편리성·효율성을 제공하고 있다. 하지만, 인터넷 인프라 자원의 취약점을 이용하여 사용자를 위협하는 경우가 발생한다. 최근에 지속적으로 지능적이고 고도화된 새로운 공격 패턴이나 악성 코드들이 늘어나고 있는 추세이다. 현재 신·변종 공격을 막기 위한 연구로 다크넷이라는 기술이 주목받고 있다. 다크넷은 미사용 중인 IP 주소들의 집합을 의미하며, 실제 시스템이 존재하지 않는 다크넷으로 유입된 패킷들은 신규 악성코드에 감염된 시스템이나 해커에 의한 공격행위로 간주 될 수 있다. 따라서 본 연구는 다크넷에 수집된 트래픽의 포트 정보를 기반한 통계 데이터를 추출하고 알려지거나 알려지지 않은 블랙 IP를 찾기 위한 알고리즘을 제시하였다. 국내 미사용 중인 IP 주소 8,192개(C클래스 32개) 다크넷 IP에서 3개월간(2016. 6 ~ 2016. 8) 총 827,254,121건의 패킷을 수집하였다. 수집된 데이터를 제시한 알고리즘 적용 결과, 블랙 IP는 6월 19건, 7월 21건, 8월 17건이 탐지되었다. 본 연구의 분석을 통해 얻어진 결과는 기존 알려진 공격들의 블랙 IP 탐지 빈도를 알 수 있고 잠재적인 위협을 유발할 수 있는 새로운 블랙 IP를 찾아낼 수 있다.

ABSTRACT

The internet is an important infra resource that it controls the economy and society of our country. Also, it is providing convenience and efficiency of the everyday life. But, a case of various are occurred through an using vulnerability of an internet infra resource. Recently various attacks of unknown to the user are an increasing trend. Also, currently system of security control is focussing on patterns for detecting attacks. However, internet threats are consistently increasing by intelligent and advanced various attacks. In recent, the darknet is received attention to research for detecting unknown attacks. Since the darknet means a set of unused IP addresses, no real systems connected to the darknet. In this paper, we proposed an algorithm for finding black IPs through collected the darknet traffic based on a statistics data of port information. The proposed method prepared 8,192 darknet space and collected the darknet traffic during 3 months. It collected total 827,254,121 during 3 months of 2016. Applied results of the proposed algorithm, black IPs are June 19, July 21, and August 17. In this paper, results by analysis identify to detect frequency of black IPs and find new black IPs of caused potential cyber threats.

Keywords: Darknet, System of security control, Detecting black IPs

Received(06. 15. 2017), Modified(06. 19. 2017), Accepted(06. 20. 2017)

* 본 연구는 2017년도 과학기술정보통신부의 기본사업 「첨단연구망기반 정보보호체계 구축 및 서비스」의 지원을 받아 수행된 연구임(K-17-L01-C04-S03)

† 주저자, painstars@kisti.re.kr
‡ 교신저자, song@kisti.re.kr(Corresponding author)

I. 서 론

인터넷은 우리나라의 경제·사회를 움직이는 중요한 인프라 자원이며 일상생활의 편리성·효율성을 제공하고 있다. 따라서 급속도로 발전하고 있는 인터넷 환경을 악용하는 지능적이고 고도화된 다양한 위협들이 발생되고 있다. 최근에 이러한 위협을 제거하기 위한 연구가 활발히 진행되고 있는데, 대표적으로 미사용 IP 주소들의 집합을 활용한 다크넷(Darknet)[1-4], 취약한 환경을 구축하여 공격자를 끌어들이는 허니넷(Honeynet)[5-8], C&C (Command & Control Server) 서버와 통신을 막는 DNS(Domain Name System) 싱크홀(Sink-Hole)[9-11] 등 다양한 방법을 통해 위협 제거를 위한 연구가 진행되고 있다.

현재 국내 보안관제 체계는 국가사이버안전센터를 중심으로 국가공공분야 정보보호를 위해 부문/단위 보안관제 센터를 운영하고 있다. 각 센터들은 다양한 보안장비인 IDS, IPS, Firewall 등을 활용하여 보안관제 서비스를 제공하고 있다. 그러나 보안관제 시스템들의 대부분은 패턴기반 보안 관제를 기반으로 알려진 공격을 막는데 집중하고 있다. 하지만 최근에는 지능적이고 고도화된 새로운 공격이 지속적으로 발생하고 증가하고 있는 추세이다. 따라서 알려진 공격을 효율적으로 막는 것도 중요하지만, 잠재적인 위협을 갖는 정보를 찾기 위한 분석의 중요성이 대두되고 있다.

많은 연구자들이 잠재적인 위협을 유발하는 신·변종 공격에 대응하기 위한 기술의 하나인 블랙 IP 주소 기반 침입탐지에 대한 연구를 활발히 진행하고 있다. 왜냐하면 블랙 IP 주소를 통해 악성 URL, 악성코드 유포지·배포지, 피싱사이트, C&C 서버, Bot 등을 사전 탐지하여 차단할 수 있기 때문이다. 현재 연구되고 있는 블랙 IP 주소 탐지를 위한 대표적인 방법으로 Low/High Interaction Honeypot[17,18], Client Honeypot[19] 등이 있다. 첫 번째로 Low/High Interaction Honeypot의 경우, 공격자에게 쉽게 노출될 수 있는 가상 환경을 조성하여 접근한 Source IP 주소에서 블랙 IP 주소를 찾아내는 방법이다. 두 번째로 Client Honeypot의 경우, 가상 환경에서 다양한 웹사이트에 접근하여 이루어지는 행위를 분석하여 블랙 IP 주소를 판단하는 방법이다. 추가적으로 악성코드를 분석하는 방법들[12-14]이 있는데, Honeypot에서 수집한 실제 악성 코드 및 파일을 실행하여 외부의 악성 IP 주소로 접속하는 시도를 찾아내는 동적 분석과 악성 코드 및 파일들의 내부 코드와 구조를 확인하는 정적 분석 방법이 있다. 동적 분석이란 파일을 직접 실행시켜 그 행위를 분석하고 디버깅을 통해 코드 흐름과 메모리 상태를 직접 확인하는 것을 말하고 정적 분석이란 악성코드를 실행하지 않고 디스어셈블러를 이용하여 내부 코드와 구조를 확인하거나 헤더 정보와 내부 문자열, 실행 압축 여부, 등록 정보 등을 통하여 동작을 파악하는 것을 말한다. 하지만 기존에 연구되고 있는 블랙 IP 주소 탐지를 위한 방법들은 다양한 어플리케이션, 운영체제(O/S) 버전 운영·관리, 가상환경 은닉, 2차 피해 제어 등의 복잡한 구조를 필요로 하고 구축에 요구되는 비용이 많다. 또한, 관리 비용과 분석에 활용되는 환경의 보안성 검토도 지속적으로 유지해야하기 때문에 운영·관리의 어려움이 있다. 그리고 분석된 블랙 IP 주소에 대해 악성 유·무를 판단하기가 어렵다.

따라서 본 논문에서는 다크넷을 활용하여 복잡한 구조와 비용이 많이 드는 문제를 해결하였다. 왜냐하면, 미사용 IP 주소 집합과 유입된 패킷을 저장하는 서버만으로 쉽게 구축할 수 있기 때문이다. 또한, 구축된 다크넷 시스템은 운영·관리 측면에서 유리하고 실제 시스템이 존재하지 않기 때문에 다크넷으로 유입된 패킷들은 비정상행위로 간주할 수 있기 때문에 추가적인 분석·분류 작업을 필요로 하지 않으며 유입 패킷 그 자체로 블랙 IP 주소 추출에 활용할 수 있다.

본 연구에서는 8,192개(C클래스 32개) 다크넷 IP 주소에서 3개월간(2016. 6 ~ 2016. 8) 패킷을 수집하였다. 수집된 데이터의 양이 방대하기 때문에 다크넷을 통해 얻을 수 있는 정보들(출발지·목적지 IP 주소, 출발지·목적지 포트, Payload 등) 중 목적지 포트에 초점을 맞춘 알고리즘을 제시하였다. 월/일간 다크넷으로 유입된 패킷의 목적지 포트 상위 10위 비교를 통해 일별 목적지 포트 상위 10위에 신규로 진입한 포트를 분석하였다. 이를 통해, 알려지거나 알려지지 않은 블랙 IP 주소들을 찾아냈다. 따라서 이 결과를 통해 알려진 악성 IP 주소와 잠재적인 위협을 유발할 수 있는 의심 IP 주소를 찾았다.

본 논문의 구성은 다음과 같다. 2장에서는 블랙 IP 주소 탐지와 다크넷 트래픽 관련 연구를 소개하고 3장은 통계 데이터를 토대로 한 다크넷 트래픽 분석 구조에 대해서 설명한다. 4장에서는 제안된 방

법을 적용한 실험결과를 제시하고 5장에서는 본 논문의 최종 결론을 맺는다.

II. 관련 연구

2.1 블랙 IP 주소 탐지 관련 연구

블랙 IP 주소 탐지와 관련된 많은 연구들은 주로 특수한 환경을 조성하여 고의적으로 공격을 받아 공격자의 행위를 관찰하거나 알려진 사전 정보(악성코드, 악성 URL, C&C 서버 등)에 접근하여 블랙 IP 주소를 찾아내는 연구가 있다. 여기에는 다양한 방법들이 존재하는데 대표적으로 Low/High Interaction Honeypot, Client Honeypot 등이 있다.

첫 번째로 Low/High Interaction Honeypot은 외부의 공격을 유인해서 현재 벌어지고 있는 해킹 상황을 확인할 수 있도록 구성한 가상 네트워크의 일종이다. 따라서 허니팟은 침입자를 잡기 위해 설치하는 네트워크가 아니라 그들의 움직임을 감시하고 시스템에 침입하기 위해 사용하는 방법, 도구 등 최신의 해킹 경향을 배우기 위한 방법이다. 따라서 침입에 이용된 블랙 IP 주소를 찾아낼 수 있다.

두 번째로 Client Honeypot은 가상 환경 안에서 악성 웹사이트나 관심 대상 웹사이트에 주기적으로 접근하여 시스템 파일, 프로세스, 레지스트리 등의 변경 사항을 확인하는 방법이다. 이러한 방법들은 중요 시스템 파일들이 변경되거나 삭제될 수 있고 관리자가 인지하지 못하는 프로세스가 실행될 수 있다. 또한 레지스트리가 변경되거나 삭제됨으로써 해당 시스템이 동작하지 않거나 의도하지 않은 방향으로 동작할 수 있다. 따라서 다양한 정보의 변경 사항을 통해 Source IP 주소에서 블랙 IP 주소를 찾아낼 수 있다.

추가적으로 이러한 Honeypot에서 수집한 실제 악성 코드 및 파일들의 내부 구조를 확인하고 특징을 찾아내는 방법과 실제 알려진 악성 코드나 파일을 실행하여 실행한 시스템 환경이 어떻게 변하는지, 어떤 IP 주소로 접근하여 정보를 주고받고 명령을 수행하는지 등을 확인하는 방법들이 있다. 따라서 악성코드별 문자열 기반의 특이점을 추출하여 특정 IP 주소로 접근한 정보를 통해 블랙 IP 주소를 찾아낼 수 있다.

하지만 이러한 방법들은 어플리케이션을 어떻게 활용하는지, O/S 선택 및 버전별 구축, 악성 코드 및 블랙 IP 주소에 대한 접근이 이루어지기 때문에 보안을 고려하여 실제 시스템 환경과 분리 등의 복잡한 구조를 필요로 하고 구축에 요구되는 비용이 많이 든다. 또한, 지속적으로 어플리케이션, O/S의 버전, 분석 소스 등의 업데이트가 이루어져야 하므로 관리 비용에 대한 부담이 있다. 그리고 분석된 블랙 IP 주소에 대해 악성 유·무를 판단하기가 어렵다.

2.2 다크넷 트래픽 관련 연구

기존 다크넷 트래픽에 대한 연구들은 대부분 다크넷으로 유입되는 방대한 양의 트래픽을 분류[15]하기 위한 연구와 다크넷 트래픽으로 특정 공격 또는 패턴 등을 추출[16]하는 연구되고 있다. 따라서 기존 연구들의 경우, 블랙 IP 주소를 찾는데 집중하기보다 공격 시나리오를 추출하고 다크넷 트래픽 분류(Scan, Flooding, Backscatter, Misconfiguration 등)에 초점을 맞춰 연구되고 있다.

다크넷 시스템을 구축하기 위해서는 기본적으로 사용하지 않는 IP 주소 집합들과 다크넷 IP 주소에 들어오는 패킷들을 저장하는 서버만으로 구축 가능하다. 그리고 실제 존재하지 않는 시스템인 다크넷으로 접근하는 시도 자체가 명백한 공격 시도로 간주할 수 있기 때문에 분석에 활용할 수 있는 용이한 정보를 얻을 수 있다. 왜냐하면 공격자가 사전에 정보를 얻기 위해 스캐닝이나 플루딩 공격 시도를 하거나 악성코드에 감염되어 지속적인 접속 시도를 하기 때문이다.

따라서 본 논문에서는 알려지거나 알려지지 않은 블랙 IP 주소를 찾는데 초점을 맞춰 다크넷 트래픽의 포트 정보를 활용한 공격 패턴의 일종인 블랙 IP 주소를 찾기 위한 알고리즘과 실제 적용 결과를 제시한다.

III. 통계기반 다크넷 트래픽 분석 방법

다크넷에 유입된 패킷에는 다양한 정보(출발지·목적지 IP 주소, 출발지·목적지 포트, Payload 등)가 포함되어 있다. 본 연구는 이 중 목적지 포트에 초점을 맞춰 블랙 IP 주소를 찾는 알고리즘을 제시하였다.

기존의 사이버 위협은 주로 보안체제나 웹 서비스

등의 취약점을 타깃으로 하여 이루어졌으나, 최근 들어 취약한 어플리케이션을 타깃으로 하여 특정 포트에 집중적으로 공격을 시도하는 행위가 많아지고 있다. 대부분의 알려진 공격들은 특정 포트로 지속적인 접근 시도가 이루어져 쉽게 관측이 가능하지만, 사전 정보가 없고 간헐적으로 접근하는 포트 중 잠재적인 위협을 유발하는 신·변종 공격을 관측 하기는 어렵다. 따라서 본 논문에서는 다크넷으로 유입된 패킷을 각 포트별로 분류하고 분류된 데이터들 중 새롭게 관측된 포트 정보를 분석하여 잠재적인 위협을 유발하는 블랙 IP 주소를 찾는다.

3.1 제안방법 개요

Fig. 1.은 제안 방법의 전체적인 흐름이며 분석 과정은 다음과 같다. 본 논문에서는 분석의 효율성을 고려하여 Top 10을 선택하였지만 Top 5 또는 Top 20 등으로 선택할 경우, 결과가 달라질 수 있다.

① 월/일 단위 패킷 분류 : 사용하지 않는 IP 주소 집합인 다크넷으로 유입된 패킷을 월/일 단위로 패킷을 분류한다.

② 목적지 포트 분류 : 월/일 단위로 분류된 패킷을 각 포트별(0~65535번)로 유입된 패킷 수를 기준으로 하여 재분류한다.

③ 신규 목적지 포트 추출 : 월별 Top 10 목적지 포트와 해당 월의 일별 Top 10 목적지 포트를 비교하여 월별 Top 10 목적지 포트에 없는 포트 정보와 이에 해당하는 패킷을 추출한다.

④ 송신자 IP 주소 추출 : 신규로 발생된 일별 Top 10 목적지 포트로 접속 시도를 한 송신자 IP 주소를 추출한다.

⑤ 악성 IP 주소 유·무 판단 : 추출된 송신자 IP 주소를 Virus-Total을 이용하여 Antivirus solutions에서 악성 IP 주소로 판단된 이력 유·무를 확인하여 이력이 존재하면 악성 IP 주소로 존재하지 않는다면 의심 IP 주소로 판단한다.

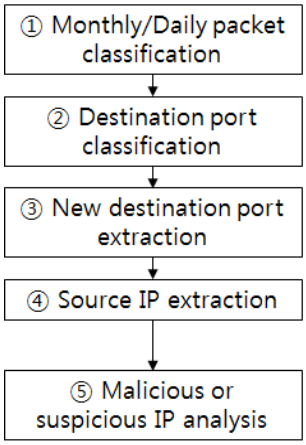


Fig. 1. Overall structure of proposed method

3.2 상세 알고리즘

제안한 목적지 포트 정보를 활용한 이상행위 검증 방법에 따라 이상행위 탐지 알고리즘을 구성할 수 있다. 이를 통해 고도화된 위협을 탐지하고 공격 전반에 대한 내용을 빠르게 파악할 수 있기 때문에 공격이 실질적인 영향을 미치기 전에 실효적인 대응을 할 수 있다. 제안하는 알고리즘의 구조는 Table 1.과 같다.

다크넷 IP 주소로 유입된 모든 패킷 P를 입력값으로 받아서 악성 IP 주소 E와 의심 IP 주소 F의 출력값을 도출하고 세부 동작 과정은 다음과 같다.

1. 사용하지 않는 IP 주소 집합인 다크넷으로 유입되는 모든 패킷을 수집한다. 수집된 다크넷 패킷 P를 $Port()^M$ 함수를 이용하여 월별로 동일한 목적지 포트를 갖는 패킷들을 멤버로 하는 단위(그룹)으로 분류한다.
2. 동일한 방법으로 수집된 다크넷 패킷 P를 $Port()^D$ 함수를 일별로 동일한 목적지 포트를 갖는 패킷들을 멤버로 하는 단위(그룹)으로 분류한다.

Table 1. Algorithm for detecting black IPs

Input	$P=\{P_1,\ldots,P_n\}(P_n : Darknet\ packet)$
Output	$E(Malicious\ black\ IP)$ $F(Suspicious\ black\ IP)$
1. $Port(P)^M \Rightarrow M=\{M_0,\ldots,M_k\}(k\in[0,65535])$	
2. $Port(P)^D \Rightarrow D=\{D_0,\ldots,D_k\}(k\in[0,65535])$	
3. $MaxMonth(M_k)=X \text{ for } X\in[1,10]$	
4. $MaxDay(D_k)=Y \text{ for } Y\in[1,10]$	
5. $Y\notin X \Rightarrow K$	
6. $K=A\cup B(A: Centralized\ IP, B: Distributed\ IP)$	
7. $Analysis(A)=C$	
8. $C=E\cup F\begin{pmatrix}E: Antivirus\ solutions\ O\\F: Antivirus\ solutions\ X\end{pmatrix}$	

3. 분류된 월별 단위(그룹)의 패킷 개수를 기준으로 M_k를 $MaxMonth()$ 함수에 입력하여 월별 유입된 패킷의 상위 10개에 해당하는 포트번호 리스트를 산출한다.
4. 동일한 방법으로 분류된 일별 단위(그룹)의 패킷 개수를 기준으로 D_k를 $MaxDay()$ 함수에 입력하여 일별 유입된 패킷의 상위 10개에 해당하는 포트번호 리스트를 산출한다.
5. 월별 Top 10 목적지 포트에 없는 포트 정보와 이에 해당하는 패킷을 추출하기 위해 $MaxDay()$ 집합 Y 중 $MaxMonth()$ 집합 X에 포함되지 않는 포트 번호 K에 해당하는 패킷을 추출한다.
6. K 집합은 편중 분포 IP 주소에서 지속적으로 패킷이 들어온 A 집합과 균등 분포 IP 주소에서 패킷이 들어온 B 집합으로 구성된다. 따라서 패킷의 송신자 IP 주소를 추출하고, 편중 분포를 나타내는 IP 주소를 산출한다. 왜냐하면 지속적인 접근 시도가 이루어지는 편중 분포 IP주소는 공격 시도나 악성코드에 감염되었을 확률이 높기 때문이다.
7. 신규 목적지 포트로 접근 시도를 한 편중 분포 IP 주소에서 패킷이 들어온 A 집합을 입력값으로 Virus-Total에서 Antivirus solutions의 탐지된 이력을 확인 할 수 있는 $Analysis()$ 함수를 이용하여 분석값 C를 산출한다.
8. 분석값 C는 Antivirus solutions에 탐지된 이력이 있는 E 집합과 탐지되지 않은 F 집합으로 구성된다. 따라서 분석값 C는 악성 IP 주소의 E 집합과 의심 IP 주소의 F 집합으로 분류된다. 따라서 악성 URL로 판단된 이력과 악성 첨부 파일의 링크 이력을 확인할 수 있다.

IV. 통계기반 다크넷 트래픽 분석 실험 결과

본 연구에서는 다크넷에 유입된 패킷을 대상으로 제안 방법의 성능 및 효율성을 검증하였다. 우선 2016년 6월부터 8월까지 3개월간 다크넷에서 수집된 최신 데이터를 바탕으로 분석을 실시하였다.

4.1 실험 환경 및 개요

다크넷에 유입된 트래픽을 관측·감시하기 위해서 국내 미사용 IP 주소 8,192개(C클래스 32개) IP 주소를 바탕으로 악성 패킷을 2016년 6월부터 8월까지 3개월간 수집하였다. 자세한 IP 주소를 공개할 경우, 해당 IP 주소에서의 공격 행위 관측에 영향을 줄 수 있기 때문에 자세한 다크넷 IP 주소 정보는 공개하지 않는다. 따라서 다크넷 IP 주소의 분포(A, B, C, D 클래스)에 따라 실험 결과 및 알고리즘 성능이 달라질 수 있다.

4.2 다크넷 트래픽 통계

사용하지 않는 IP 주소 집합으로 다크넷을 구축하고 유입되는 모든 패킷을 월/일 단위로 분류하였다.

월별 다크넷으로 유입된 패킷은 Table 2.와 같다. 세부적으로 6월은 254,440,763건, 7월 295,811,101건, 8월 277,002,257건의 패킷이 수집되었다. 월별 다크넷으로 유입된 패킷의 Unique Source IP 주소는 6월 4,540,729개, 7월 4,025,682개, 8월 8,392,962개로 나타났다.

일별 다크넷으로 유입된 패킷을 Table 3.과 같다. 일별로 평균 9,000,000건 정도의 패킷이 유입되었다. 대부분 동일한 payload를 가지는 중복되는 패킷들이었다. 월별 Unique Source IP 주소의 수와 일별 Unique Source IP 주소의 수 합계가 다른 이유는 일별 발생된 Unique Source IP 주소들 사이에 중복되는 IP 주소들이 존재하기 때문이다. 8월의 경우, APT 공격의 비중이 낮아지고 본격적인 표적형 공격이 진행됨에 따라 Unique Source IP 주소의 수가 늘어난 것으로 예상된다.

Table 2. Collected monthly information of darknet packets

	June	July	August
Number of darknet packets	254,440,763	295,811,101	277,002,257
Number of unique source IPs	4,540,729	4,025,682	8,392,962

4.3 월별/일별 목적지 포트 추출 및 비교

다크넷에서 수집한 패킷을 월별 목적지 포트 Top 10으로 분류한 자료는 Table 4.와 같다. 3개월간 수집된 데이터의 각 월별 순위는 비슷한 결과를 나타냈다. 그리고 몇몇의 포트(5060번, 8080번, 443번)는 Top 10에 속하지 않은 달이 있지만, Top 20 내에 포함되어 있었다. 전체적으로 Top 10에 해당하는 패킷이 전체의 약 80%정도를 차지하였고 특히 23번 포트(telnet)가 가장 많은 비중을 차지하였다.

Table 3. Collected daily information of darknet packets

Number of daily darknet packets								
June			July			August		
	Number of daily darknet packets	Number of unique source IPs		Number of daily darknet packets	Number of unique source IPs		Number of daily darknet packets	Number of unique source IPs
1	10,655,917	464,587	1	8,220,182	331,731	1	11,619,338	328,308
2	9,125,947	441,781	2	11,258,267	364,204	2	10,705,840	460,004
3	6,359,184	380,727	3	8,750,582	356,485	3	9,791,592	575,500
4	9,210,504	392,632	4	8,401,615	395,271	4	9,377,651	634,935
5	8,899,013	381,723	5	7,783,598	390,512	5	9,895,000	500,036
6	10,342,350	392,028	6	8,730,833	354,929	6	11,421,925	526,211
7	7,538,814	407,951	7	9,102,735	332,356	7	10,956,996	613,598
8	7,082,341	376,674	8	7,268,840	262,360	8	9,389,803	527,865
9	7,056,946	369,207	9	8,278,704	247,295	9	10,004,605	716,043
10	7,294,804	372,996	10	9,340,722	248,774	10	8,238,542	603,612
11	8,750,623	376,391	11	6,708,000	231,815	11	9,580,293	509,400
12	8,870,253	360,161	12	9,137,271	310,719	12	9,024,947	562,634
13	8,377,788	334,362	13	7,242,923	292,920	13	9,206,974	503,965
14	6,730,448	259,989	14	8,126,103	310,668	14	8,221,312	424,242
15	7,173,890	283,128	15	9,109,777	319,381	15	8,278,476	472,097
16	6,618,321	329,909	16	10,909,296	319,182	16	7,107,510	364,201
17	6,616,218	336,361	17	10,027,628	324,624	17	7,694,470	369,538
18	7,767,773	321,014	18	8,170,351	304,345	18	7,441,940	490,155
19	9,291,905	325,858	19	8,583,044	312,874	19	8,147,499	445,057
20	8,176,003	351,289	20	8,560,827	294,600	20	8,665,733	541,984
21	8,751,639	359,498	21	7,959,496	240,300	21	8,572,114	571,917
22	9,043,535	385,677	22	7,870,454	246,978	22	7,988,836	591,421
23	8,844,996	373,951	23	11,782,202	252,714	23	7,503,918	674,513
24	10,560,409	345,671	24	12,600,028	268,157	24	7,941,053	627,162
25	10,686,091	348,081	25	11,801,394	268,564	25	7,270,195	624,954
26	11,699,169	305,872	26	10,900,232	293,575	26	8,063,627	637,029
27	8,243,465	248,948	27	11,817,630	303,103	27	9,794,554	662,083
28	8,320,155	297,178	28	11,133,403	301,165	28	9,477,509	606,386
29	7,955,588	345,632	29	11,077,826	295,278	29	7,943,875	605,449
30	8,396,674	356,943	30	12,074,189	292,609	30	9,099,293	685,317
31	-	-	31	13,082,949	280,515	31	8,576,837	674,311
Total	**254,440,763**	**10,626,219**	**Total**	**295,811,101**	**9,348,003**	**Total**	**277,002,257**	**17,129,927**

Table 4. Experimental results from June to August, 2016

	June			July			August		
Rank	Destination Port	Count	Rate	Destination Port	Count	Rate	Destination Port	Count	Rate
1	**23**	**145,391,658**	**57.1%**	**23**	**164,914,904**	**55.8%**	**23**	**172,075,546**	**62.1%**
2	**53413**	**15,661,491**	**6.2%**	**53413**	**39,926,022**	**13.5%**	**53413**	**18,178,925**	**6.6%**
3	1433	10,478,965	4.1%	1433	10,789,158	3.6%	1433	9,680,314	3.5%
4	22	6,670,342	2.6%	445	6,745,920	2.3%	445	6,542,380	2.4%
5	445	6,327,365	2.5%	22	3,886,784	1.3%	80	3,747,821	1.4%
6	80	4,082,535	1.6%	80	3,630,507	1.2%	22	3,744,273	1.4%
7	3389	2,592,602	1.0%	3389	2,566,864	0.9%	3389	2,683,378	1.0%
8	3306	2,495,698	1.0%	3306	2,475,203	0.8%	5060	2,536,850	0.9%
9	5060	2,079,693	0.8%	443	1,774,757	0.6%	3306	2,106,770	0.8%
10	8080	1,920,132	0.8%	8080	1,254,974	0.4%	443	1,899,553	0.7%
	Total	**197,700,481**	**78%**	**Total**	**237,965,093**	**80.4%**	**Total**	**223,195,810**	**80.8%**

53413번 포트의 경우, 중국 내에서는 넷코어(Netcore)로 중국 외 국가에서는 네티스(Netis)라는 제품명으로 판매 중인 제품들에서 기기에 접근할 수 있는 백도어가 발견되어 많은 공격 시도가 발생됐다. 수집된 패킷의 대부분이 원격 접속(23번 Telnet, 1433번 Microsoft SQL Server, 22번 SSH Remote Login Protocol, 445번 NETBios, 3389번 MS WBT Server)과 관련되거나 HTTP 프로토콜(80번 World Wide Web HTTP, 8080번 HTTP Alternate, 443번 http protocol over TLS/SSL)과 관련된 포트이다.

수집된 일별 다크넷 패킷의 Top 10 목적지 포트를 추출하여 Table 4.의 월별 Top 10 목적지 포트와 비교했다. 그 결과 6월은 32개, 7월은 49개, 8월은 34개의 신규로 Top 10에 진입한 포트가 발생했다. 신규로 진입한 목적지 포트들은 잘 알려진 포트(443번), 등록된 포트들(4028번, 8888번, 12345번, 41630번 등), 동적 포트들(57209번, 57248번, 64749번 등)이 관측됐다. 일반적으로 사용자들은 알려지지 않은 다양한 포트 번호를 이용한다. 따라서 트로이 목마와 같은 프로그램들은 악의적인 목적으로 알려지지 않은 취약한 포트 번호를 찾아서 공격에 활용하는 경우가 있다.

4.4 분석 대상 IP 주소 분류

신규 목적지 포트로 접근 시도가 있는 편중 분포 IP 주소를 분석 대상으로 분류했다. 그 결과 6월 19개, 7월 21개, 8월 17개로 분류됐다. 편중 분포 IP들은 Virus-Total을 통해 악성 IP 주소(6월 4개, 7월 4개, 8월 3개)와 의심 IP 주소(6월 15개, 7월 17개, 8월 14개)로 분석되었다. 해당하는 악성 IP 주소가 발생된 날짜와 포트 번호는 Fig. 2.와 같다. 본 논문에서는 각 월의 한 개의 경우를 택하여 기술하였다. 6월 6일의 57248번 포트의 경우, 총 10개의 출발지 IP 주소에서 발생되었고 이중 특정 IP 주

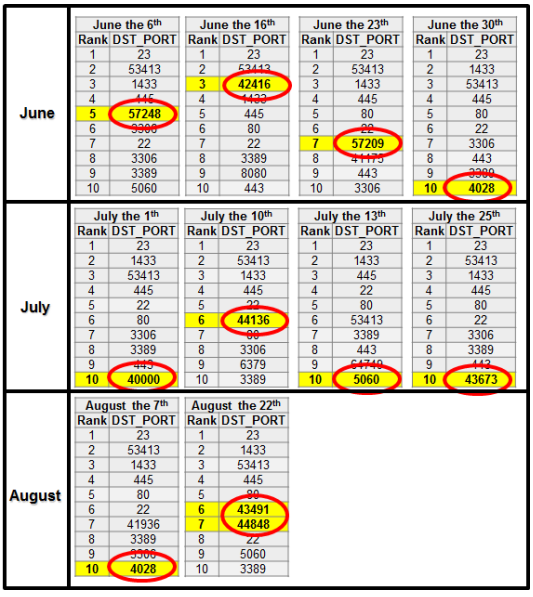


Fig. 2. Experimental results of algorithm for detecting black IPs

소(59.*.*.2)에서 집중된 패킷이 발생된 것을 확인하였고 해당 IP 주소는 Antivirus solutions에서 1회 악성 첨부 파일의 링크 이력이 있었다. 7월 25일의 43673번 포트의 경우, 총 6개의 출발지 IP 주소에서 발생되었고 이중 특정 IP 주소(94.*.*.132)에서 집중된 패킷이 발생된 것을 확인하였고 해당 IP 주소는 Antivirus solutions에서 9회 악성 URL 판단 이력과 1회 악성 첨부 파일의 링크 이력이 있었다. 8월 22일의 43491번 포트의 경우, 총 3개의 출발지 IP 주소에서 발생되었고 이중 특정 IP 주소(37.*.*.18)에서 집중된 패킷이 발생된 것을 확인하였고 해당 IP 주소는 Antivirus solutions에서 10회 악성 URL 판단 이력과 7회 악성 첨부 파일의 링크 이력이 있었다.

따라서 제시한 알고리즘을 통한 악성 IP 탐지율을 Table 5.에서 확인할 수 있다. 의심 IP 주소의 경우 추후 악성 IP로 판단될 가능성이 존재하기 때문에 지속적인 관리가 필요하다.

Table 5. Experimental results list from June to August, 2016

	June	July	August
Number of New Destination Port	32	49	34
Number of Detected Source IPs	19	21	17
Number of Malicious IPs	4	4	3
Number of Suspicious IPs	15	17	14
Ratio of Malicious IPs	21.1%	19%	17.6%

V. 결 론

본 연구는 잠재적인 위협을 유발할 수 있는 블랙 IP 주소를 찾기 위한 알고리즘을 제안하였다. 국내 미사용 IP 주소 8,192개(C클래스 32개) 다크넷 IP 주소 집합을 활용하여 수집한 실제 데이터를 목적지의 포트별로 통계 데이터를 산출하고 이상행위 탐지를 위한 알고리즘을 통해 기존 및 신규 블랙 IP 주소를 찾았다. 기존 알려진 위협인 악성으로 판단된 IP 주소는 6월 4개, 7월 4개, 8월 3개이고 잠재적인 위협을 유발할 수 있는 의심으로 판단된 IP 주소는 6월 15개, 7월 17개, 8월 14개였다.

이처럼 다크넷 패킷의 다양한 정보 중 포트 정보를 활용하여 잠재적인 위협과 알려진 위협을 사전에 차단할 수 있는 정보를 이끌어 낼 수 있다. 기존 방법들에 비해 다크넷 패킷의 대부분이 위협 트래픽이기 때문에 효율적인 분석이 가능하다.

다크넷으로 들어온 트래픽의 대부분은 오설정으로 인해 발생된 PC일 가능성이 있기 때문에 실제 서버인지 악성 웹사이트인지 추가적인 확인이 필요하여, 트래픽을 분류하는 연구가 활발히 진행되고 있다[15]. 이상행위 탐지 알고리즘을 통해 찾아낸 기존 블랙 IP 주소의 경우, 현재는 악성으로 판단되었지만 향후 정상 IP 주소가 될 수 있다. 또한 신규 블랙 IP 주소의 경우, 현재는 정상으로 판단되었지만 향후 악성 IP 주소가 될 수 있기 때문에 지속적인 관리가 필요하다. 향후 의심 IP 주소가 악성 IP 주소로 판단된다면 이상행위 탐지 알고리즘이 효과적으로 잠재적인 위협을 사전 탐지 할 수 있다는 것을 알 수 있다.

References

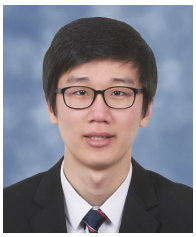
[1] Eto, M., Inoue. D., Song, J., Nakazato, J., Ohtaka, K., and Nakao, K., "nicter : A Large-Scale Network Incident Analysis System," Proc. of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security(BADGERS '11), pp. 37-45, Apr. 2011.

[2] 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.

[3] 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.

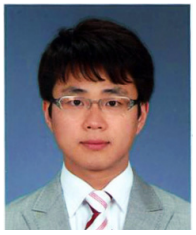
2005.
[4] Fachkha, C., and Debbabi, M., "Darknet as a Source of Cyber Intelligence Survey, Taxonomy and Characterization" IEEE Communications Surveys&Tutorials, pp. 1197-1227, Nov. 2015.
[5] Spitzner, L., "The Honeynet Project: trapping the hackers," Magazine of Security & Privacy, IEEE pp.15-23, Mar. 2003.
[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] Park, J., Choi, J., and Song, J., "How to Design Practical Client Honeypots Based on Virtual Environment" Asia Joint Conference on Information Security(AsiaJCIS), pp.67-73, Aug. 2016.
[9] 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.
[10] 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.
[11] 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.
[12] 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.
[13] 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.
[14] 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.
[15] Liu, J., and Fukuda, K., "Towards a Taxonomy of Darknet Traffic" International Wireless Communications and Mobile Computing Conference(IWCMC), pp. 37-43, Aug. 2014.
[16] Ban, T., Eto, M., Guo, S., Inoue, D., Nakao, K., and Huang, R., "A Study on Association Rule Mining of Darknet Big Data" International Joint Conference on Neural networks(IJCNN), pp. 1-7, Jul. 2015.
[17] S, Mukkamala., K, Yendrapalli., and R, Basnet., "Detection of Virtual Environments and Low Interaction Honeypots," Information Assurance and Security Workshop, 2007, June. 2007.
[18] Ayeni O.A, Alese B.K, and Omotosho L.O., "Design and Implementation of a Medium Interaction Honeypot," International Journal of Computer Applications, May. 2013.
[19] Supinder, K., and Harpreet, K., "Client Honeypot Based Malware Program Detection Embedded Into Web Pages " Supinder Kaur et al Int. Journal of Engineering Research and Applications, pp. 849-854, Dec. 2013.

〈저자소개〉



박 진 학 (Jin-hak Park) 정회원
2013년 2월: 국민대학교 수학과 졸업
2016년 2월: 국민대학교 금융정보보안학과 석사
2016년 2월~현재: 한국과학기술정보연구원 첨단연구망정보보호실 연구원
〈관심분야〉 정보보호, 네트워크 보안, 악성코드 분석

권 태 웅 (Tae-woong Kwon) 정회원
2012년 2월: 숭실대학교 컴퓨터학부 졸업
2014년 8월: 고려대학교 정보보호대학원 정보보호학과 석사
2014년 12월~현재: 한국과학기술정보연구원 첨단연구망정보보호실 연구원
〈관심분야〉 정보보호, 보안관제, 네트워크 보안, 네트워크 가시화



이 윤 수 (Youn-su Lee) 정회원
2007년 2월: 전남대학교 산업공학과 졸업
2010년 8월: 충남대학교 컴퓨터공학과 석사
2017년~현재: 고려대학교 컴퓨터·전파통신공학과 박사과정
2012년~현재: 한국과학기술정보연구원 첨단연구망정보보호실 선임연구원
〈관심분야〉 정보보호, 보안관제, 침해사고대응, 네트워크 보안, 보안이벤트 상관분석

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

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 보안연구소 전문연구원
2010년 10월~2011년 9월: 일본정보통신연구원 네트워크 보안연구소 선임연구원
2011년 10월~현재: 한국과학기술정보연구원 첨단연구망정보보호실 선임연구원
2012 9월~현재: 과학기술연합대학원대학교 그리드 및 슈퍼컴퓨팅 부교수
〈관심분야〉 보안관제, 침해사고대응, 악성코드 분석, 네트워크 보안