봇넷 탐지 연구 조사

허선동, 이민수, 윤현수

한국과학기술원

sdheo@nslab.kaist.ac.kr, mslee@nslab.kaist.ac.kr, hyoon@nslab.kaist.ac.kr

A Survey of Botnet Detection

Seondong Heo, Minsoo Lee, Hyunsoo Yoon

Korea Advanced Institute of Science and Technology

요 약

네트워크가 발전함에 따라 다양한 종류의 악성 소프트웨어들이 현대의 보안을 위협하고 있다. 최근 이러한 악성 소프트웨어 중 하나인 봇의 경우 스스로 네트워크를 구축하고, 이를 이용한 다양한 형태의 공격들에 활용되고 있다. 본 논문에서는 최근 위협이 되고 있는 봇넷 탐지의 연구 방향성을 제안한다. 이를 위해 봇넷의 특성을 분석하고, 기존의 탐지 방식들에 대해서 기술한다.

1. 서 론

봇넷이란, 봇에 감염된 컴퓨터끼리 하나의 그룹을 형성하여 악의적인 용도로 사용되는 네트워크를 의미한다[1]. 이러한 봇넷에 활용되는 감염 PC들의 수가 점점 늘어나고 있는데, *USATODAY*의 조사 에 따르면 인터넷에 접속하고 있는 8억 대의 컴퓨터 중에 40%의 컴퓨터가 봇에 감염된 상태로 파악되었다 [2].

봇은 로봇에서 비롯된 용어로 미리 정의된 방식으로 동작하는 자동화된 소프트웨어를 의미한다. 이러한 봇들에는 사용자에게 악의적인 행위를 하는 악성 봇들이 존재하는데, 본 논문에서는 이러한 악성 봇을 대상으로 연구를 진행하였다

봇넷이 악의적으로 활용되는 사례들을 살펴보면 대표적으로 다음과 같이 분류할 수 있다[3].

1) Distributed Denial of Service(DDoS) 공격: 시스템을 악의적으로 공격해 해당 시스템의 자원을 부족하게 하여 원래 제공하고 있는 서비스를 제공하지 못하게 하는 공격이다[19]. 봇에 감염된 컴퓨터들이 많아지면, 공격자는 이를 통해서 DDoS 공격을 할 수 있다.

2) 개인 정보 불법 수집: 봇넷은 감염된 PC들로부터 개인정보나 비밀번호, 금융 정보 등을 불법적으로 수집할 수 있다. 이는 스팸 메일에 이용하거나, 금전적인 이익을 위해 스패머들에게 제공된다.

3) 애드웨어 설치: 감염된 PC들로 하여 자동으로 광고 프로그램을 다운로드하고 설치하도록 해서, 강제적으로 사용자에게 광고 창을 띄운다거나, 상황에 따라서 다른 광고를 띄울 수 있다.

4) 스팸 메일 발송: 회사나 제품의 광고를 위해 봇넷을 이용하여 스팸 메일을 발송할 수 있다. 2006년의 *SiSCO*사의 조사에 따르면, 스팸 메일의 80 퍼센트가 봇넷에 의해 발송되고, 매년 30%씩 증가하고 있는 추세이다.

5) 클릭 수 위장 사기: 봇넷을 이용하여 사용자 클릭 수를 부당하게 늘리는 것으로, 감염된 PC들로 하여금 클릭을 하게 하여 이득을 취한다. 가령 클릭 수마다 작은 액수를 지급하는 배너 광고의 경우, 봇에 감염된 PC들을 이용하여 배너를 클릭하게 하면, 금액이 작더라도 봇넷의 크기가 커질 경우 매우 큰 이득을 취할 수 있으며, 각 감염된 PC들의 IP 주소가 다르기 때문에 부당한 이득을 취하고 있는지 알기가 어렵다.

봇넷이 이와 같은 행동을 하기 위해서는 네트워크를 통한 C&C(Control and Command) 채널의 유지가 필수적인 요소이다. 이러한 특성은 봇 네트워크를 탐지하기 위해서 중요한 요소가 된다. 따라서 본 논문에서는 봇 넷의 C&C 채널의 특성을 분석하고, 현재의 연구들을 탐지에 사용한 방법에 따라 분류한 후, 향후 봇넷 탐지를 위한 방향성을 제시하고자 한다.

이를 위해 2장에서는 봇넷의 특성을 분류하고, 3장에서는 기존의 봇넷 탐지 연구에 대해서 기술한다. 마지막으로 4장에서는 결론과 봇넷 탐지 연구의 방향성을 제시한다.

2. 봇과 봇넷

본 장에서는 C&C 채널의 특성을 기반으로 봇넷을 분류하였다. C&C 채널이란 공격자가 봇넷에 명령을 전달하기 위한 네트워크 연결을 의미한다.

2.1 IRC 기반 봇넷

C&C 채널이 IRC 프로토콜 기반인 봇넷을 IRC 기반 봇넷이라고 한다. IRC(Internet Relay Chatting) 프로토콜은 텍스트 메시지를 전달하기 위해 사용되는 프로토콜로써 그룹 통신이나 클라이언트끼리의 1:1 통신에 사용되고 있다.

IRC는 많은 사용자들이 이미 사용하고 있고, IRC 프로토콜은 수 천대의 클라이언트들을 동시에 관리하기 용이하다[5]. 이러한 특성은 봇넷에 쉽게 이식되어 현재 가장 많은 봇넷의 C&C 채널로 활용되고 있다[4].

IRC 기반 봇넷에서는 그림1과 같이 봇마스터가 C&C 서버에 명령을 입력하면, 채널에 접속해 있는 봇들은 실시간으로 커맨드를 받아서 실행하고, 그 결과를 채널을 통해 봇마스터에게 보내준다. 봇마스터란 봇을 유포하고 봇넷을 유지, 이용하는 악의적인 공격자를 뜻한다. 즉, IRC 기반 봇넷은 봇마스터에 의해 실시간으로 작동한다.

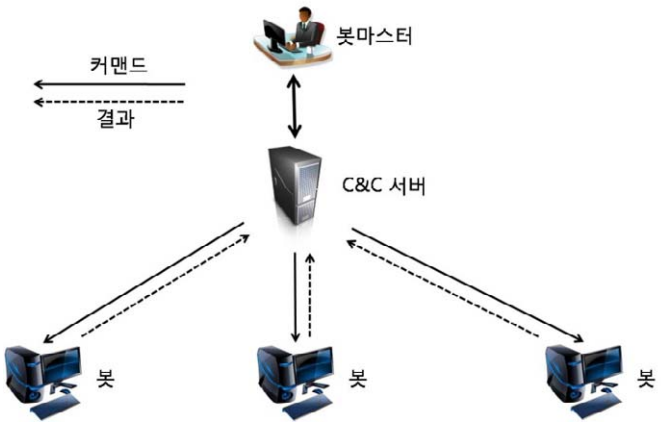

그림 1. IRC 기반 봇넷의 구조

2.2 HTTP 기반 봇넷

C&C 채널이 HTTP 프로토콜 기반인 봇넷을 HTTP 기반 봇넷이라고 한다. HTTP 프로토콜은 대부분의 ISP에서 허용하는 프로토콜이고, 많은 정보가 오고 가기 때문에 탐지가 어렵다.

HTTP 기반 봇넷의 경우, HTTP프로토콜의 특성 상 주기적인 특성을 가진다. 즉, 봇에 감염된 각 호스트들은 주기적으로 명령을 받기 위해 중앙의 서버에 접속을 시도하는 특성을 가진다.

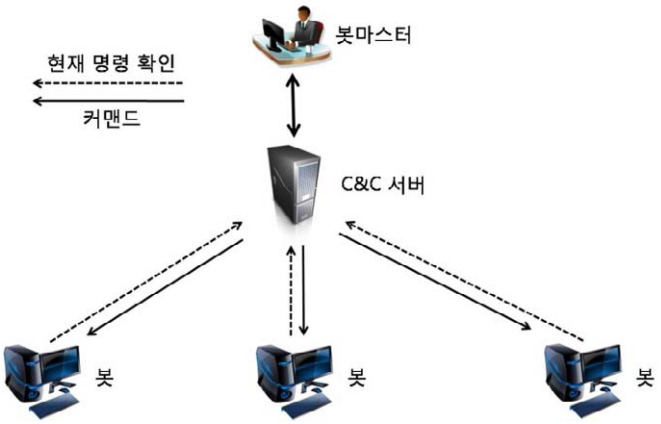

그림 2. HTTP 기반 봇넷의 구조

2.3 P2P 기반 봇넷

C&C 채널이 Peer-to-Peer 구조인 봇넷을 P2P 기반 봇넷이라고 한다.

P2P 기반 봇넷에는 하나의 C&C 서버가 존재하는 것이 아니라, 여러 개의 C&C 서버가 존재하며, 서버와 클라이언트 역할을 모두 수행할 수 있다. 즉, C&C 서버끼리 통신을 하여 명령을 주고 받을 수 있다.

IRC 기반 봇넷이나, HTTP 기반 봇넷은 봇마스터나 C&C 서버가 탐지되어 제거되었을 경우, 봇넷 전체가 기능을 잃게 된다 [8].

하지만 P2P 기반 봇넷의 경우는 그림 3과 같이 한 서버가 기능을 못한다 하더라도 여러 개의 서버가 존재하고 있기 때문에 이러한 약점이 사라진다.

이러한 P2P 기반 봇넷이 많지 않고, 존재하는 P2P 기반 봇넷들은 탐지 및 방어가 용이하지만, *Ping Wang* 의 연구에서 제안한 것과 같은 P2P 기반 봇넷이 나온다면 기존의 연구들로는 탐지 및 방어가 어렵다[9].

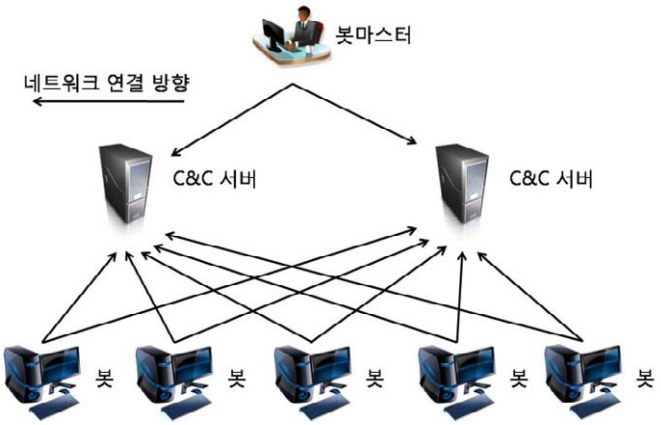

그림 3. P2P기반 봇넷의 구조

3. 봇넷 탐지 연구

본 장에서는 봇넷 탐지 연구들을 탐지에 사용한 방법을 기반으로 분류하였다.

3.1 서명 기반 탐지

봇넷의 서명(signature)과 행동을 축적해서, 축적된 지식을 기반으로 봇넷을 탐지하는 방법이다. 대표적인 예로 Snort가 있다[10].

이러한 서명 기반 탐지 방법은, 기존에 발견 되었던 봇넷이나 악성 코드에 대해서는 높은 탐지율을 보이며 오탐율도 낮은 장점이 있다. 그러나 새로운 봇넷이 공격할 경우, 이에 대한 시그너쳐가 없기 때문에 탐지가 불가능하다.

이와 비슷한 방법으로, Rishi에서는 봇의 IRC nickname으로 봇을 판단한다[11]. 봇들의 IRC nickname이 일반 사용자들의 nickname과 많은 차이점을 보인다는 점에 착안한 것이다.

하지만 이러한 방법은 극히 제한적인 방법이다. 우선 IRC nickname을 일반 사용자들과 비슷하게 만들면 탐지가 어렵다. 또한, IRC 기반 봇넷이 아닌 경우는 탐지가 불가능하다. 최근처럼 봇넷의 통신이 암호화되는 시점에, 이러한 방법은 더 제한을 많이 받게 될 것이다.

3.2 허니넷을 이용한 탐지

이 방법은, 허니넷을 이용한 방법으로, 허니팟을 이용하여 봇을 잡아낸 후, 그 봇을 분석한다[3]. 분석한 결과를 기반으로 봇으로 위장한 소프트웨어를 만들어낸 후, 이 소프트웨어와 통신을 주고 받는 트래픽을 분석하여 봇마스터를 찾거나 봇넷을 찾아내는 방법이다.

오탐율이 낮고 기존 지식이 없더라도 봇넷을 탐지해 낼 수 있다는 장점이 있다. 그러나 이러한 탐지는 몇 가지 어려움에 직면하고 있다[12].

우선, 이 방법은 scalable하지 않다. 한 봇에 하나의 허니팟을 배당해주어야 하는데, 봇의 종류가 많아지면 그 개수만큼 허니팟의 개수가 필요해진다. 결국 네트워크 트래픽이 커지면 커질수록 허니팟이 많이 필요하게 되고, 이는 일반적인 사용자가 사용할만한 방법이 아니다. 또한, 한 봇이나 악성 소프트웨어가 모든 허니넷을 감염시키는 일도 발생할 수 있다. 이럴 경우 허니팟 하나를 제외하고는 모두 자원의 낭비를 하게 된다. 또한, 허니넷을 운영하는 컴퓨터가 봇에 감염될 수 있는 위험도 감수해야 한다.

두 번째로, 봇과 봇마스터의 탐지 차단 기술이 점점 발달하고 있다는 점이다. *Paul Barford*의 연구에서 볼 수 있듯이, 이미 허니팟을 인식하고 작동을 멈추는 봇이 등장했다[6]. 이러한 봇이 아니더라도, 최근 봇들의 통신이 암호화되고 있기 때문에, 블랙박스 테스트로 봇을 흉내내는 일이 점점 더 어려워지고 있다. 봇을 흉내내는 것이 가능하다 해도, 봇마스터의 탐지 차단 기술이 탐지 기술이 발달함과 함께 같이 발전하고 있다. 다른 봇들과 행동이 다르면 그 봇과 연결을 끊는다거나, 봇이 하지 않는 행동을 찾아내서 연결을 끊는 기술 등은 더 이상 특별한 기술이 아니다.

3.3 이상(anomaly) 기반 탐지

이상 기반 탐지는 네트워크 트래픽의 이상한 점, 가령 집중된 트래픽이나 평소 사용하지 않던 포트를 사용하는 등 일반적인 사용자와 다른 행동을 보이는 것들을 봇으로 의심하고 탐지하는 방법이다.

이상 기반 탐지는 새로운 봇넷이라 하더라도 탐지가 가능하기 때문에, 서명 기반 탐지가 가진 단점은 보이지 않는다. 그러나 일반적인 사용자와 다른 행동을 보여야 탐지가 가능하기 때문에, 아무 행동을 하지 않으면 탐지가 불가능하다. *J.R.Binkley*의 연구에서는 이 점을 해결하고, 봇 뿐만 아니라 봇 서버까지 알아낼 수 있는 방법을 제시하였다[13]. 그러나 이 연구는 C&C 채널에서 암호화된 통신을 사용할 경우 탐지가 어렵다는 문제점이 있다.

*W.Timothy Strayer*의 연구에서는 일반적인 트래픽을 모아서 필터링해서 데이터를 모은 다음, 데이터를 기반으로 봇넷을 탐지한다[15]. 이 방법은 사전에 데이터를 모아야 하지만, 새로운 봇넷이 공격하더라도 탐지가 가능하다. 반면 IRC 기반 봇넷만을 탐지 가능하다는 단점이 있다.

*Anestis Karasaridis*의 연구에서는, 트랜스포트 계층의 플로우 데이터를 분석하여 봇넷을 탐지하는 방법을 제시하였다[5]. 플로우 데이터 중 봇의 데이터로 의심되는 데이터를 따로 추출한 후, 이 데이터를 가지고 점수를 계산하여, 이 점수가 임계치를 넘어서면 봇으로 판단한다. 이 방법은 봇넷의 통신이 암호화되어 있다 하더라도 탐지가 가능하고, 오탐율이 낮다. 또한 scalable하며, 봇넷의 크기를 측정하는 데도 도움을 줄 수 있다. 반면 IRC 기반 봇넷만 탐지가 가능하다는 점이 이 방법의 단점이라고 할 수 있다.

*BotSniffer*은 기존에 IRC 기반 봇넷만 대상으로 하던 연구들과는 달리, HTTP 기반 봇넷도 탐지가 가능하다[16]. *BotSniffer*에서 사용하는 두 가지 알고리즘은, Response-Crowd-Density-Check(RCD) 알고리즘과 Response-Crowd-Homogeneity-Check (RCH) 알고리즘이다.

RCD 알고리즘은, 일반적인 사용자와 달리 봇넷의 봇들의 반응은 굉장히 집단적이라는 점에 착안한 알고리즘이다. 집단적으로 비슷한 시간에 반응을 보이는 컴퓨터들은 봇에 감염되었다고 판단하는 것이다. RCH 알고리즘은 집단 내의 메시지들이 서로 얼마나 비슷한지에 대한 검사를 한다. 이 두 알고리즘을 결합하여, 미리 정의된 임계치를 기준으로 봇넷 여부를 판단하게 된다. *BotSniffer*의 경우 기존의 연구와는 달리 HTTP 기반 봇넷도 탐지할 수 있으며, 이를 확장할 수 있다는 장점이 있다. 또한 암호화된 통신을 사용하는 봇넷도 탐지가 가능하다. 반면 통신의 딜레이가 큰

봇넷의 경우는 탐지율이 떨어지며, 오탐율이 높아진다는 단점이 존재한다.

*BotMiner*에서는, *BotSniffer*보다 한 단계 더 나아가서 C&C 채널의 구조나 프로토콜에 상관없이 봇넷을 탐지해낼 수 있는 방법을 제안했다[17]. *BotMiner*는 A-plane과 C-plane을 조합하여 봇넷을 찾아낸다. A-plane에서는 로그 정보를 모니터링하며, '누가 무엇을 하고 있는가'에 대해 분석한다. C-plane은 네트워크 플로우와 기록된 정보를 가지고 '누가 누구와 통신을 하고 있는가'에 대해 분석한다. *BotMiner*는 현재까지 나온 봇넷 탐지 관련 연구에 비해 참신한 방향을 제시했으며, 탐지율이 99.6%로 매우 높고, 거의 0%에 가까운 오탐율을 보였다. 또한 C&C 채널의 구조나 프로토콜에 관계없는 탐지 방법을 사용하기 때문에, 기존 연구들이 가지는 단점들, 즉 사전지식이 필요하거나 특정 프로토콜을 사용하지 않으면 탐지가 불가능하거나, 암호화된 통신을 사용할 경우 탐지가 불가능한 등의 단점을 가지고 있지 않다.

이상 기반 탐지는 봇넷에 대한 사전지식 없이도, 일반적인 데이터나 패턴과 비교해서 봇넷을 탐지하기 때문에, 새로운 봇넷이 출현해도 탐지가 가능한 반면, 일반적인 사용자를 봇이라고 판단하는 오탐이 확률적으로 발생한다. 대부분의 이상 기반 탐지 방법을 사용하는 연구에서는, 오탐율을 줄이기 위해 노력한다. 가령 *W.Timothy Strayer*의 연구에서는 블랙/화이트 리스트를 이용해서 오탐율을 줄이는 것을 볼 수 있다[15]. 이 외에도 분석 데이터량을 늘리거나 좀더 정확한 추측 방식을 사용하는 것도 오탐율을 낮추는 방법이다.

4. 결론 및 향후 연구

지금까지 봇넷의 종류와 봇넷 탐지 방법에 대해 나열해 보았다. 서론에서 언급했듯이, 봇넷은 점점 더 커지고 있으며 이에 따라 봇넷으로 인해 발생할 피해도 점점 커지고 있다. 우리나라도 2009년 7월 7일 주요 정부기관, 포털 사이트, 은행 사이트 등을 DDoS 공격에 당한 적이 있다[20]. 이러한 봇넷은 인터넷 사용자가 점점 더 늘어나면 늘어날수록 크기가 늘어날 것이며, 따라서 앞으로 위험은 더 높아질 것이다.

*Ping Wang*의 연구에서 보여주었듯이, 봇마스터들의 탐지 방어 기술의 발전에 따라서, 기존의 연구들은 점점 더 이러한 봇넷들은 탐지하기가 어려워지고 있다[9].

이러한 봇넷을 탐지하고 방어하기 위해서는, 기본적으로 *BotMiner*와 같은 C&C 채널의 구조나 기반 프로토콜에 관계없이 봇넷을 탐지할 수 있는 연구들이 계속 되어야 한다[17]. 또한, C&C 채널의 구조나 기반 프로토콜에 관한 연구도 병행하여 프로토콜의 특수성이나 프로토콜을 사용하는 사람들의 경향 등을 분석하여 이를 이용하면, 오탐율은 낮고 탐지율은 더 높은 연구가 가능할 것이다.

참고 문헌

[1] Moheeb Abu Rajab, Jay Zarfoss, Fabian Monorose, Andreas Terzis, "A Multifaceted Approach to Understanding the Botnet Phenomenon", In Proc. 6th ACM SIGCOMM Conference on Internet Measurement (IMC'06), pp.41-52, 2006

[2] www.usatoday.com/tech/new/computersecurity/2008-03-16-computer-botnets_N.htm, 2008

[3] The Honeynet Project & Research Alliance, "Know your enemy: Tracking botnets", 2005

[4] K.J.Houle, G.M.Weaver, "Trends in denial of service attack technology", CERT, 2001

[5] Anestis Karasaridis, Brian Rexroad, David Hoeflin, "Wide-scale Botnet Detection and Characterization", USENIX HotBots'07, 2007

[6] Paul Barford, Vinod Yegneswaran, "An Inside Look at Botnets", In Special Workshop on Malware Detection Advances in Information Security, Springer Verlag, 2006

[7] Jose Nazario, "BlackEnergy DDoS Bot Analysis", Technical report, Arbor Networks, 2007

[8] J.B.Grizzard, V.Sharma, C.Nunnery, B.B.Kang, D.Dagon, "Peer-topeer botnets: Overview and case study", USENIX HotBots'07, 2007

[9] Ping Wang, Sherri Sparks, Cliff C.Zou, "An Advanced Hybrid Peer-to-Peer Botnet", USENIX HotBots'07, 2007

[10] Martic Roesch, "SNORT-LIGHTWEIGHT INTRUSION DETECTION FOR NETWORKS", USENIX LISA'99, 1999

[11] Jan Goebel, Thorsten Holz, "Rishi: Identify Bot Contaminated Hosts by IRC Nickname Evaluation",USENIX HotBots'07, 2007

[12] Jose Nazario, "Botnet tracking: Tools, techniques, and lesson learned, Technical report, Arbor Networks, 2007

[13] J.R.Binkley, S.singh, "An algorithm for anomaly-based botnet detection", SRUTI'06, pp.43-48, 2006

[14] Evan Cooke, Farnam Jahanian, Danny Pherson, "The Zombie Round up: Understanding, Detecting, and Disrupting Botnets", USENIX SRUTI'05, 2005

[15] W.Timothy Strayer, David Lapsely, Robert Walsh, Carl Livadas, "Botnet Detection Based on Network Behavior", Proceedings 2006 31st IEEE Conference on Local Computer Networks, pp.195-202, 2006

[16] Guofei Gu, Junjie Zhang, Wenke Lee, “BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic”, NDSS Symposium 2008, 2008
[17] Guofei Gu, Roberto Perdisci, Junjie Zhang, and Wenke Lee, “BotMiner: Clustering Analysis of Network Traffic for Protocol-and Structure-Independent Botnet Detection”, USENIX Security ’08, 2008
[18] Z.zhu, G.Lu, Y.Chen, Z.J.Fu, P.Roberts, K. Han,”Botnet research servey”, 32nd Annual IEEE International Computer Software and Applications Conference(COMPSAC’08), 2008
[19] http://en.wikipedia.org/wiki/Denial-of-service_attack
[20] http://ko.wikipedia.org/wiki/7%C2%B77_DDoS_%EA%B3%B5%EA%B2%A9