행위기반의 봇넷 탐지 기술

고등학생의 장난에서 시작한 바이러스는 놀라운 속도로 발달해 웜, 스파이웨어, 봇 등 다양한 공격기술로 진화, 공격자의 경제적인 이익을 위해 배포되고 있다. 안전한 사이버 공간을 위해서는 이미 나타난 위협 뿐 아니라 새롭게 발표되는 기술에 대한 공격, 아직 알려지지 않은 공격에 대해 예측하고 대응할 수 있어야 한다. 본지는 고려대학교 컴퓨터보안연구실과 함께 6회에 걸쳐 새로운 사이버 공격을 탐지하는 기술에 대해 살펴본다. 이번호에서는 새로운 보안위협으로 질의는 봇넷에 대해 알아본다.

글 · 최현상 고려대 컴퓨터보안연구실 연구원, 이희조 고려대 교수

연재순서

1. *행위기반 봇넷 탐지기술*
2. *알려지지 않은 스파이웨어 탐지기술*
3. *VoIP공격 탐지기술*
4. *알려지지 않은 신종 인터넷 웜 탐지기술*
5. *악성코드 사전 탐지기술*
6. *IP 스푸핑 탐지기술*

그동안 우리 사회는 인터넷 기술의 발달과 이용인구가 증가하면서 인터넷을 구성하는 기반시설이 급격히 설치돼 인터넷 인프라에서 보안적인 요소를 치밀하게 준비하지 못했다. 기술이 발달할수록 더 많이 취약점이 노출되었고, 이를 악용하려는 공격자가 등장해 불특정 다수를 공격하는 사이버 테러가 빈번히 발생하고 있다. 사이버 공격은 경제적 이득을 얻기 위해 전문적이고 조직적인 네트워크를 형성해 나가고 있다.

사이버 공격자들은 불법행위를 통해 금전적 이득을 취득하기 위해 자신의 위치가 노출되지 않으면서 자신이 통제하고 제어하는 컴퓨터 네트워크를 구성하는 기술을 고안하게 됐다. 이러한 목적으로 등장한 것이 '봇넷(Botnet)'이다. 봇넷의 공격자는 수천에서 수만 대의 컴퓨터에 봇을 설치하고 네트워크를 통해 이들을 동시에 제어해 악성행위를 대행하게 한다.

봇넷의 규모가 커지면서 이를 이용한 공격종류와 방법도 다양하게 나타나고 있으며 그에 따른 피해도 심각해지고 있다. 보넷은 최근 제로데이 공격과 더불어 사이버 보안의 최대의 이슈로 대두되고 있다. 최근 봇넷에 대응하기 위해 봇을 탐지하는 여러 기술이 제안이 되고 있으나 지금까지 제안된 기술은 원론적인 수준에 머물고 있다. 봇넷을 탐지하는 여러

기술에 대해 살펴보고 행위기반으로 봇넷을 탐지하는 기술을 소개한다.

사용자, 봇넷 감염여부 확인하기도 어려워

DDoS 공격이나 스팸, 파밍 등 악성공격의 대부분이 봇넷에 의해 행해지고 있다. 봇은 말 그대로 로봇을 의미한다. 악의를 가진 해커가 PC를 감염시켜 자신이 마음대로 조종할 수 있는 봇으로 만들고, 이렇게 감염된 PC 수천, 수만 대가 네트워크를 통해 연결돼 조종, 통제하는 권한을 가진 봇 마스터(Bot master)에 의해 원격조종된다. 이러한 형태를 봇넷이라고 한다.

여러 가지 경로를 통해서 PC가 봇에 감염되는데 스팸메일에 실행코드를 클릭해 감염되는 경우, 웜에 감염코드를 싣고 취약성을 가진 PC를 감염시키는 경우, 메신저를 통해 감염되는 경우 등 그 수법이 다양하다. 루트킷을 이용해 감염되는 경우 PC 사용자가 감염여부를 쉽게 확인하기 어렵다.

봇넷의 라이프 사이클을 살펴보면 봇 마스터는 자신과 봇넷 사이의 명령을 전달하고 제어하기 위해 IRC(Internet Relay Chat) 채널을 이용한다. 봇에 의해 감염된 PC는 자동으로 최근의 봇 코드를 받아 자신을 업데이트하고 명령/제어 서버로 접속한다. 봇넷에서 주로 TCP 6665~6669 포트를 사용하며, DDoS 공격과 피싱, 파밍, 스패밍 등 사이버 상의 공격을 수행한다. 대부분의 DDoS 공격과 스팸메일은 봇넷을 통해 발생하고 있다고 보고되고 있으며, 이를 통해 심각한 피해가 발생하고 있다.

시만텍의 통계에 따르면 2006년 후반기에만 600만대의 새로운 봇이 생성된 것으로 나타난다. 이는 2006년 상반기에 비해 약 29% 증가한 수치이다. 이에 반해 명령/제어 서버는 약 25% 감소했다. 봇마스터가 자신의 봇넷을 결합하거나 새로운 봇 감염 PC를 추가해 봇넷의 크기를 증가시키고 있다는 것을 의미한다. 실제로 봇넷이 14만대 이상의 봇으로 연결돼 있는 사례도 발견이 됐다. 이런 봇넷은 보통 OC3 이상의 DDoS 공격 트래픽을 발생시킬 수 있다.

봇넷, 패턴이용 탐지기술은 쉽게 우회

클라이언트 기반의 봇넷 탐지 기술은 크게 시그니처 탐지와 이상행위 탐지 기법으로 나뉜다. 대다수의 안티바이러스 솔루션에서 제공하는 것이 봇의 감염코드를 이용한 시그니처 탐지기법이다. 이 기법은 변종 봇이나 신종 봇에 대한 탐지가 느리고 소프트웨어적인 패킹기법을 통해 간단하게 회피할 수 있다는 단점이 있다. 클라이언트 기반의 이상행위 탐지기술의 경우, 스탠포드대학의 엘리자베스 스틴슨

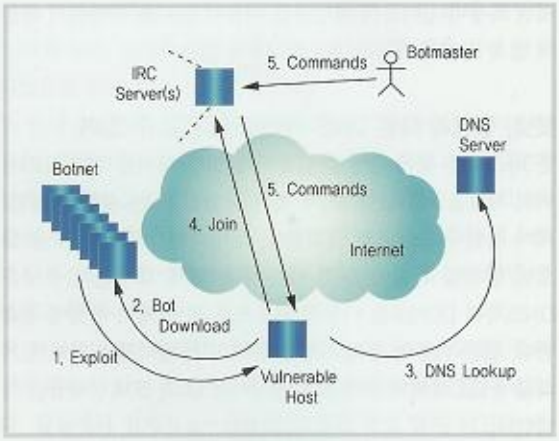


그림1. 봇넷의 라이프 사이클

(Elizabeth Stinson)이 제안한 시스템 콜의 이상행위를 이용해 탐지하는 기술이 제안 됐으나 오탐율이 높다.

네트워크 기반의 봇 탐지에 대한 연구는 주로 봇넷이 사용하는 IRC 포트인 6667번 포트의 모니터링을 통해 이루어진다. 하지만 6667번이 아닌 다른 포트를 사용하는 봇이 증가하면서 다른 포트에서도 트래픽의 패턴 매칭을 통해 봇넷의 IRC 트래픽을 탐지하는 기술이 연구되고 있다. 그러나 이런 기술은 모니터링하는 네트워크 링크의 트래픽 양이 많으면 데이터 처리가 힘들고 악성 봇에 감염된 시스템과 봇넷 명령/제어 서버 간 통신에 암호화(SSL) 통신을 하면 패킷 모니터링을 통해 탐지하기 어렵다. 현재 많은 봇이 통신 암호화를 사용하고 있으며 포트 또한 알려지지 않은 포트를 사용하고 있기 때문에 네트워크 기반의 시그니처 탐지 기술 역시 한계를 갖는다.

최근 제안되는 봇넷 탐지기술은 행위를 기반으로 한 것이다. 조지아공대의 데이빗 데이곤(David Dagon)은 2006년 ORAC 워크숍에서 봇넷의 도메인 네임 패턴을 이용해 명령/제어 채널서버를 찾는 기술을 발표했다. 올해 USENIX Hotbot에서 얀 괴벨(Jan Goebel)은 봇마다 사용하는 고유의 닉네임 특성을 이용한 탐지 기법을 발표했다. 그러나 이들은 봇마스터가 이 기법을 알 경우 봇넷의 구조나 감염기법의 변화 없이 단순하게 이름의 패턴만을 바꿔 탐지되지 않도록 할 수 있다는 한계를 갖는다.

이외에도 봇넷을 통해 스팸메일을 보내기 전에 자신의 봇넷이 블랙리스트에 추가돼 있는지 알아보는 조사쿼리를 탐지해 봇넷을 탐지하는 기술이 SRUTI에서 발표된 바 있으나 마찬가지로 봇넷의 구조적/기술적 변화 없이 간단하게 탐지를 회피할 수 있다. 이처럼 대부분의 봇넷 탐지 기술들은 봇넷의 단편

적인 특징이나 이름의 패턴만을 이용해 탐지를 수행하기 때문에 쉽게 회피가 가능하다는 한계점을 갖는다.

봇넷, 정상과 다른 DNS 쿼리로 탐지할 수 있어

봇 마스터는 봇에 감염된 컴퓨터와 명령/제어를 위해 IRC서버의 채널을 이용한다. 이때 봇에 감염된 PC는 봇넷의 명령/제어 서버에 자동으로 집결한다. 봇마스터가 감염된 봇을 찾으면 발각될 확률이 높기 때문이다. 현재 대부분의 봇넷은 DNS(특히 DDNS)를 이용해 자동으로 명령/제어 서버에 접속한다. 봇은 역공학 등의 기법을 통해 IP가 발각될 위험이 있기 때문에 IRC서버의 IP주소를 인지하지 않고, 도메인네임만 기억한다.

보통 하나의 봇넷은 다수의 명령/제어 서버를 이용해 주기적으로 서버를 이동하는데, 이동 중에도 DNS를 이용해 명령/제어 서버에 접속한다. 처음 봇에 감염된 후 IRC 서버에 접속할 때와 명령/제어 서버에서 이주할 때, IRC 서버의 IP주소가 변경된 경우, 그리고 특정 명령수행을 할 때 봇은 IRC서버에 접속하기 위해 스스로 DNS 쿼리를 전송한다. DDNS를 이용하는 IRC서버는 빈번하게 IP주소가 변경되며 DNS 레코드의 TTL 값이 작게 설정돼 있기 때문에 봇에서 DNS 쿼리가 자주 발생한다. 물론 로컬캐시에 저장된 데이터를 이용하는 경우에는 DNS 쿼리가 외부에 전송되지 않기 때문에 보이지 않는 경우도 있다.

봇의 단체로 행동을 하는 기본적인 특성 때문에 봇넷에서 발생하는 DNS 쿼리는 정상적인 DNS 쿼리와 구분되는 특징을 가지며 이를 통해 봇넷의 탐지를 수행할 수 있다. DNS 쿼리를 보내는 소스가 정상 DNS 쿼리라면 불특정 유저로 추정된다. 봇넷 DNS 쿼리는 크기가 일정한 그룹(즉, 봇넷)이다. DNS 쿼리를 전송하는 시간을 살펴보면, 정상일 때는 지속적이지만, 봇넷이라면 일시적으로 나타났다가 사라지는 등 규칙적이지 않다. 정상적인 DNS 쿼리는 DDNS를 사용하는 경우가 상대적으로 적은데 반해, 봇넷은 대부분 DDNS를 사용하기 때문에 이 부분에서 차이점을 갖는다.

시간에 따른 DNS 쿼리발생 패턴 통해 봇넷 탐지

봇넷의 C&C 과정에서 발생하는 DNS 쿼리의 특징을 이용해서 정상적인 DNS 쿼리와 봇넷의 DNS 쿼리를 구분할 수 있는

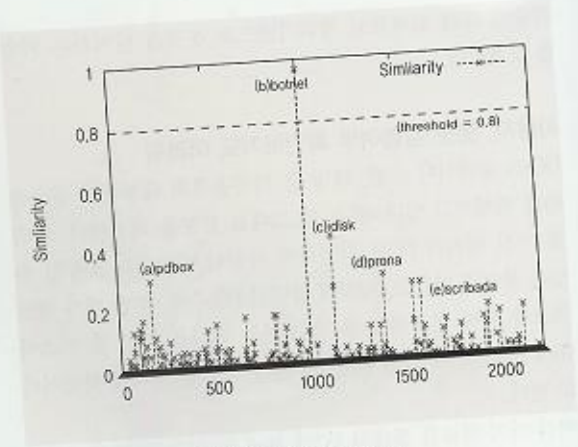


그림2. Domanin Name

봇넷 탐지 알고리즘을 개발하였다. 알고리즘은 DNS 쿼리를 일정 시간단위로 저장해 각각 시간에 따라 준비된 자료구조에 입력과 삭제 과정을 하는 부분과 봇넷의 DNS 쿼리 탐지를 하는 부분으로 나누어져 있다.

특정 타임슬롯 t1 동안 DNS 쿼리를 모아서 준비된 데이터베이스에 저장한다. 이때 이미 존재하는(t1의 이전에) 도메인네임에 대해서는 그 도메인에 대한 IP 리스트 안에 쿼리를 보낸 IP주소가 존재하는지 확인한다. 이렇게 모든 쿼리를 도메인네임을 기준으로 IP주소 그룹을 만든다. 삭제부분에서 도메인네임이 화이트리스트에 있는지, 그룹으로 묶인 IP주소 리스트의 사이즈가 미리 정해진 임계치 T를 넘는지 확인하고 하나라도 참일 경우, 데이터베이스에서 도메인네임과 IP리스트 모두 삭제한다.

임계치를 넘거나 화이트리스트에 존재하지 않으면 블랙리스트와 유사도(Similarity)를 계산한다. 유사도란 시간에 따라 한 도메인에 대한 쿼리의 IP리스트 사이즈가 이전 시간에 같은 도메인에 대한 쿼리의 IP리스트와 얼마나 일치하는가를 나타낸 것이다. 서로 다른 타임슬롯에서 서로 다른 시간대에 블랙리스트에 오른 같은 도메인네임을 갖는 IP리스트를 서로 비교해 얼마나 많은 수가 일치하는지 계산해 유사도임계치 α를 넘으면, 봇에서 이용되는 도메인네임으로 간주한다.

이 방법은 오프라인의 DNS 트래픽 데이터를 분석하는 시스템으로 이용 하는 것뿐만 아니라, 실시간으로 네트워크 DNS

표1. 정상 DNS와 구분되는 봇넷 DNS의 특징

	Source IPs accessed to domain name	Activity and Appearance Patterns	DNS Type
Botnet DNS	Fixed size Group(Botnet members)	Group activity Intermittently appeared(Specific situation)	Usually DDNS
Legitimate DNS	Anonymous(Legitimate users)	Non-group activity Randomly and continuously appeared(Usually)	Usually DNS

트래픽을 모니터링해 봇넷의 존재 여부와 감염 PC를 탐지하는 시스템으로 동작할 수 있다. 시스템의 위치는 이상적인 상황을 가정하면 Root DNS 서버에서 보내주는 DNS 트래픽을 이용해 오프라인/온라인으로 동작하는 경우이고 DNS 트래픽을 제공하는 DNS 서버의 규모가 작으면 작을수록 임계치의 조정이 필요하다. 트래픽 수집의 타임슬롯 t값의 조절도 필요하다.

봇넷이 이동하는 경로를 탐지하기 위해서는 같은 이름의 도메인네임에 대한 IP리스트 유사도를 측정하는 부분에서 서로 다른 도메인네임이면서 비슷한 크기를 갖는 IP리스트를 비교한다.

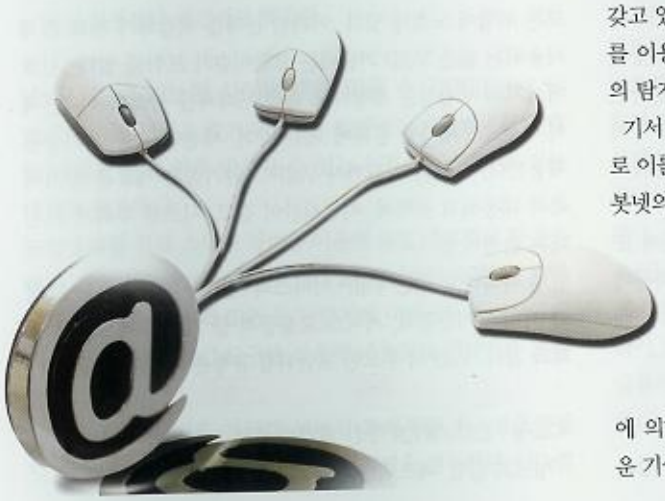
알고리즘을 통해 실제 망에서 봇넷을 탐지하는 것이 가능한지 확인하기 위해서 허니넷(Honeynet)망에 봇넷을 설치해 보았다. 명령/제어 과정을 거쳐 네트워크에서 트래픽을 수집한 뒤 알고리즘을 테스트를 했다. 허니넷 망은 1Gbps 규모이고 허니넷에서 봇넷으로 이용된 PC는 50대, 테스트 수행 시간은 약 10시간 이다. 쿼리 데이터의 입력과 삭제 및 탐지 알고리즘에서 사용한 타임슬롯 t는 1시간이고 도메인 네임에 대한 IP 리스트 크기의 임계치는 5, 유사도(Similarity) 임계치 α는 0.8이었다.

실제 테스트를 한 결과 대부분의 쿼리는 같은 도메인네임을 갖는 IP리스트 크기가 임계치 T를 넘지 못했다. 네트워크의 규모와 타임슬롯 t의 사이즈 설정에 따른 차이는 있겠지만 대부분의 경우 타임슬롯 t동안 같은 도메인네임에 대한 쿼리를 보내는 서로 다른 PC의 수가 많지 않다는 것을 알 수 있다.

테스트에서 그룹크기 임계치 T를 넘은 경우 대체로 세 가지 패턴을 보인다. 그림2는 유사도를 세 가지 도메인에 대해 시간에 따라 어떻게 변하는지 살펴본 것이다. 첫 번째 봇넷은 (b)처럼 높은 유사도 값을 갖고 있으며 임계치 값을 넘어 탐지되었음을 알 수 있다.

두 번째 (a), (c)처럼 대용량 파일전송을 지원해주는 사이트의 경우, 재접속률이 높고 지속적인 접속과 큰 파일 전송 때문에 높은 유사도 값을 갖는 것을 알 수 있다. 대체로 0.2~0.4의 값을 갖는데 이는 특정 시간 t2에 도메인 A에 접속한 호스트가 t1시간에도 30~50%의 확률로 도메인 A에 접속을 했던 것이라고 할 수 있다.

세 번째로 널리 알려진 사이트가 아닌 대다수의 도메인들의 경우 유사도 값이 0.2 이하의 낮은 값을 갖게 됨을 알 수 있었다. 실험에서 임계치 $\alpha = 0.8$ 로 설정해 항상 탐지 되었다. 한 사이트에 많은 수의 접속자가 몰리게 되는(Flash crowd) 상황과 같은 특별한 경우 오탐이 발생할 수 있으나 실제 테스트 안에서도 그러한 오탐이 확인이 되지 않았다.



다양한 단계별 모니터링, 지속적인 관찰로 봇넷 탐지

봇넷 DNS의 특징을 이용한 탐지의 경우 스푸핑 된 IP를 이용한 가짜 DNS 쿼리를 이용한 탐지회피가 가능하다. 이런 경우 TCP 3-way handshaking 체크를 통해 DNS 쿼리만 스푸핑 된 IP를 통해 발생한 것을 소거해 탐지 회피를 막을 수 있다.

최근에는 단편적인 봇넷의 특징만 탐지하는 기술에서 벗어나 다양한 단계를 갖고 있는 봇넷을 단계별 모니터링과 지속적인 관찰을 통해 탐지하는 기법이 개발되고 있다. 이 방법은 높은 탐지율을 보이고 있는 것으로 확인되고 있다. 실제로 IDS, IPS 등에서 보내주는 로그데이터들을 기반으로 봇넷의 단계별 탐지를 수행하는 기술이 발표되고 있다.

봇넷탐지를 위한 여러가지 연구가 진행되고 있지만, 알려진 기법 대부분 봇 마스터가 쉽게 회피할 수 있다는 등의 한계를 갖고 있다. 봇넷에서 명령/제어 과정에서 발생되는 DNS 쿼리를 이용해 봇넷을 탐지하는 알고리즘을 설명했고, 실제 봇넷의 탐지가 가능하다는 것을 실험을 통해 확인했다. 그러나 여기서 제안한 기법도 회피할 수 있는 기술이 나올 수 있으므로 이를 막을 수 있는 기술적인 보완이 필요하다.

봇넷의 탐지와 봇넷 기술의 발전은 마치 끝없는 '군비확대 경쟁(Arms Race)'과 같다. 하나의 탐지 기술이 개발되면 그것을 회피하는 기술이 개발된다. 봇넷의 피해는 사이버 기술이 발달하는 것과 비례하여 앞으로는 더욱더 커다란 위협으로 우리에게 다가올 것이다. 이러한 막강한 위협으로 다가오는 봇넷에 의한 피해를 줄이기 위해 봇넷을 탐지하고 방어하는 새로운 기술의 개발을 수행해야 한다.