능동형 봇넷 탐지 및 관리를 위한 단계적 마이닝 프로세스

김도훈*, 이택*,신승용*, 인호*, 정현철**
*고려대학교 컴퓨터 전파통신 공학과
**한국정보보호진흥원
e-mail:*{karmy01, comtaek, syshin, hoh_in}, **hcjung@kisa.or.kr

The Gradation Mining Process for Active Botnet Detection and Management

Do-Hoon Kim*, Sung-yong Shin*, Hoh Peter In*, HyunCheol Jeong**
*Dept of Computer Science Engineering, Korea University
**Korea Information Security Agency, Seoul, Korea

요 약

사이버 공간에서 미래 최대 위협 중 하나로 인식되고 있는 봇넷의 공격이 점차 증가함에 따라, 봇넷 공격에 기반한 피해가 증가하고 있으며, 금전적인 피해 유발로 그 심각성이 점차 증대되고 있는 실정이다. 특히, 봇넷은 좀비 PC를 활용하는 측면에서 제 2차, 3차 피해가 우려되고 있다. 따라서 봇넷의 탐지를 1차적으로 끝나는 것이 아니라 지속적인 관찰과 관리를 통해 변종 봇넷을 탐지 하고 이에 기반한 악성행위를 탐지하는 것이 무엇보다도 중요하다. 따라서 본 논문에서는 이러한 봇넷을 능동적으로 탐지하기 위한 능동형 봇넷 탐지 및 관리를 위한 단계적 마이닝 프로세스를 제안하고 기존 탐지 알고리즘과의 비교 평가를 하여 향후 적용을 위한 고려사항들을 논의 하고자 한다.

1. 서론

전 세계적으로 C&C(Command & Control) 서버와 악성 봇이 국제적인 영역에 걸쳐 광범위하게 분포하고 있으며, 국내에도 C&C 서버와 좀비가 미국, 중국 다음으로 밀집되어 있고 봇넷으로 인한 국내의 피해사례가 보고되고 있다. 이러한 봇넷은 다음과 같이 분류가 될 수 있다. 첫 번째로 중앙집중형인 IRC/HTTP 봇넷과 분산형인 P2P 봇넷으로 구분되며, 현재는 IRC를 이용한 봇넷이 널리 퍼져 있으나 최근에는 HTTP 봇들과 P2P봇들이 빠르게 증가하는 추세를 보이고 있다. 다음은 IRC, HTTP, P2P2 봇넷의 비교표이다.

<표 1> IRC, HTTP, P2P 봇넷

구 분	명령/제어	주요 악성 행위	접속 유지 방법
IRC 봇넷	IRC 채팅 패킷	DDoS,Spam 정보 유출	PING-PONG
HTTP 봇넷	Http 변수들 (Get 요청/ 명령어 수신	DDoS, Spam 정보 유출	No
P2P 봇넷	P2P 프로토콜 (search, publish등)	Spam, 정보 유출	P2P 프로토콜

이러한 봇넷은 점차적인 지능화에 따라 기존의 탐지/분석 도구로는 탐지 및 대응이 어려워지고 있다. 특히, 봇넷은 자기복제 및 빠른 전파성(Worm)을 지니고 있으며 DDoS, RootKit, Backdoor, Spam mail, Spyware, Adware 등을 이용하여 봇 마스터 (Bot Master)가 시스템 통제 및 제어를 하는 악성 행위를 가능케 하고 있다. 이처럼 다양한 변종 및 신종 봇넷은 그 악성행위에 따라 다양한 잠재적 관찰 요인을 가지고 있으며 그 요인에 의거 향후 발생하 가능한 변종 봇넷 및 신종 봇넷의 출현을 야기 시킬 수 있다. 이는 행위 기반의 봇넷 탐지 연구가 무엇보다도 중요하며 다양한 요구사항이 필요하다. 따라서, 본 논문에서 봇넷 탐지 및 관리를 위해 효율적인 마이닝 접근방법을 제안하고자 한다. 즉, 봇넷 행위 기반 탐지는 그 특성상 악성행위들의 형태가 다양하게 나타나기 때문에 넷플로우(NetFlow) 정보를 이용하여 그 성격(IRC, HTTP, P2P 형태)을 고려한 분류법이 필요하다. 이렇게 분류된 정보는 봇넷의 구성, 분포, 행동에 대해 분석할 수 있고, 이를 통해 직접적인 봇넷 탐지를 하거나 시공간 정보를 이용, 확률 기반의 학습을 하여 봇넷 탐지를 2차적으로 수행 할 수 있다. 또한 이러한 탐지 정보를 통해 다양한 봇 감염 관련 Blacklist를 만들어 데이터베이스를 구축하고 관계 정보를 이용하여 다양한 검증 알고리즘을 통해 그 유효성을 검증하고자 한다. 이러한 봇 마이닝 시스템은 기존의 행위기반 시스템의 단점을 보완하여 우회 공격에도 탐지가 가능하고 학습 시스템에 의해 효과적으로 신종 및 변종 봇넷에 대응할 수 있으며 검증 시스템에 의해 그 신뢰성을 규명하여 실무에 직접적으로 응용 할 수 있는 프

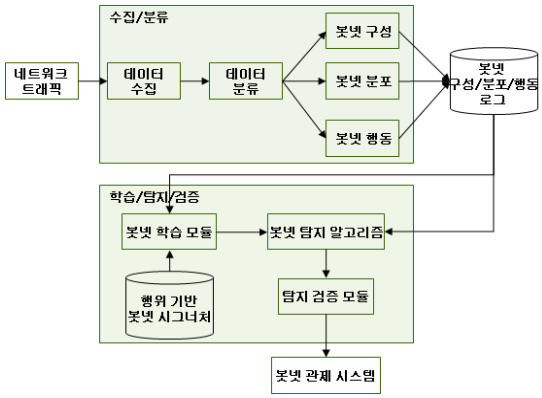
로세스를 제시하고자 한다.

2. 관련연구

전 세계적으로 봇넷 연구는 다양하게 진행되고 있다. 특히 이러한 연구는 봇넷 탐지 기반의 연구로 많은 한계점과 추가 연구에 대한 필요성을 내포하고 있다. 먼저 국외사례로 Georgia 공대의 Dagon[1]이 2005년 봇넷의 도메인네임들의 패턴을 이용해 명령/제어 채널 서버를 찾는 기술을 제안하였으나, 봇마스터가 단순하게 도메인 이름들의 패턴만을 바꿔줌으로써 탐지가 되지 않도록 할 수 있는 한계를 지니고 있다. 그리고, Georgia 공대의 N.F.A.Ramachandran [2]이 스패밍 봇을 탐지하기 위해서 봇마스터가 (DNS-based Blacklist)을 참조하는 방법을 이용하여 탐지하는 기술을 제안하기도 했다. 다음은 독일 RWTH Aachen 공과 대학의 Jan Goebel[3]이 봇과 C&C 서버 사이의 통신 채널을 발견하는 개선된 탐지 기술을 제시하였지만, 기존의 탐지법은 일반적으로 사용하지 않거나 의심이 되는 IRC nickname / server / 특이한 포트 번호의 특징을 이용하여 네트워크 트래픽을 모니터링 하는 수동적인 방법을 사용했는데 False Negative 의 빈도가 매우 높았다. 마지막으로 UCF(University of Central Florida)의 PingWang[4]이 2007년 USENIX HOTBOT 2007에서 기존의 P2P 봇넷의 단점을 고려해 고안된 하이브리드 봇넷을 소개했다. 이 하이브리드 봇넷은 봇 자신이 고정, 공인 IP를 가지고 있으면서 클라이언트와 서버 두 가지 모드로 동작하는 'servent 봇' 들과 사설이면서 유동 IP를 가지고 있는 'client 봇' 들로 이루어지고 봇마스터는 report 명령을 위해 Sensor 호스트라 불리는 봇을 지정하는데 이 Sensor 호스트로 부터 봇넷의 전체 크기를 알아낼 수 있으며, 이 봇넷에 대한 대응책을 찾아 낼 가능성이 있다. 한편 국내 사례로는 안철수 연구소와 국방과학연구소에서는 메모리 감시를 이용하여 봇 서버의 행위를 지속적으로 모니터링하고 자동분석하는 호스트 기반의 탐지 및 대응 봇넷 역추적 기술[5]을 제안하였다. 하지만 논문을 통해 기술을 제안하였지만 아직 구현 및 검증이 수행되지 않았고, P2P 봇넷에 대해서는 언급이 없는 한계점을 지니고 있다. 또한, 국민대학교는 패킷 모니터링을 통하여 IRC를 이용하는 공격성 악성코드들을 분석하고, 이들 악성코드들을 효과적으로 탐색하여 감염 여부를 판단할 수 있는 악성 봇 탐색 시스템 설계하였다. 하지만 암호화 통신, IRC 포트의 변경, 허가된 포트를 사용하는 신종/변종 봇넷들의 탐지 및 IRC 채널 내에 접속된 감염 시스템에 대한 대처 방안이 없는 한계점을 지니고 있다. 이처럼 국내에서는 봇넷에 대하여 전문적으로 대응할 수 있는 솔루션이 전무한 실정이고 이러한 국내 백신업체들은 KISC에서 제공되는 봇 샘플에 대하여 감당할 수 있는 수준을 넘어선 관계로 백신 프로그램에 적용시키기 어려움을 토로하고 있다. 그리고 ISP 업체들 역시 블랙리스트에 기재된 도메인 네임들에 대하여 즉각적인 처리를 하지 못하고 있는 상황이다. 특히, DDoS, Key Logging, 스팸, 피싱등 봇넷으로 인한 다양하고 심각한 위협에 노출돼 있음에도 불구하고 아직 봇넷에 의한 특정한 피해가 발생되지 않았다는 사실로 적극적인 대처가 이루어지지 못하고 있다. 따라서 다양한 봇넷 탐지 연구의 한계성을 극복하기 위해서는 마이닝 기법이 필요하고 이를 통해 우회하는 변종 봇을 탐지해내고 지속적인 관찰과 관리를 통해 2~3차적인 악성행위를 사전에 차단하거나 예측이 가능하겠다.

3. 능동형 봇넷 탐지 및 관리를 위한 단계적 마이닝 프로세스

다음은 능동형 봇넷 탐지 및 관리를 위한 단계적 마이닝 프로세스의 전체 구성도이다.



(그림 1) 전체 구성도

그림 1은 본 논문에서 제시하는 프로세스로서 다음과 같은 단계로 나누어 볼 수 있다.

(1) 데이터 수집

네트워크 트래픽에서 다음과 같은 정보를 수집한다.

- 행위(activity) 로그 수집
  (scan, spam, binary downloading, exploit code등)
- NetFlow 로그 수집 (UDP나 TCP로 전송되는 플로우 데이터)

(2) 데이터 분류

수집된 정보를 통해 다음과 같이 마이닝 기법을 사용하여 특성별로 분석한다.

- 봇넷 구성 분석 (Clustering, Correlation): 봇으로 의심이 되는 행위들을 클러스터링 하여 감염(좀비)PC의 물리적 위치 정보를 분석함. 또한, 클러스터링 된 정보들은 관리 대상 그룹(국가, 지역, 기관)으로 설정, 분류하고 취약지역에 대한 구성 밀도 분석이나 봇 전파에 관한 그룹별 상관관계분석을 시행한다.
- 봇넷 분포 분석 (Pattern, Classification, Prediction)

: 위에서 클러스터링 된 그룹들의 밀도 분포에 대한 변화를 측정하고 이를 봇취약성 지역 선포에 대한 지표로 활용한다. 그리고, 해당 지표의 변화 추이를 분석/예측하여 지속적인 분포변화를 관측하며 클러스터링된 그룹들의 밀도 분포의 변화를 이용하여 보안(봇에 노출된 정도) 취약성에 대한 신뢰도 평가나 위험도(봇 밀도에 의한 피해 추정) 평가로 활용한다.

- 봇넷 행동 분석 (Time-Series, Mutirelation Data Mining) : 클러스터링된 그룹들은 밀도 분포에 의한 양적(Density) 측정뿐만 아니라 질적(Severity) 측정(해당 그룹의 악용성 분석)을 시행함. 그리고 특정 그룹들이 다양한 피해 시나리오(DDoS, 스팸, 개인정보탈취 등)를 생성하여 해당 그룹의 향후 봇넷 행동(성향)을 분석하는데 이용된다.

(3) 봇넷 학습모듈

기본적으로 학습을 위해 기존의 봇넷 행위 시그너처와 함께 다음과 같이 분석한다.

- 시/공간 대별 반복 학습 : 봇넷의 분포(혹은 밀도)를 특정 시/공간대별로 묶어 봇에 의한 오염도 분석과 향후 확산에 대해 분석한다.
- 확률 기반 반복 학습 : 봇 의심 로그 자료에 기반하여 다양한 확률 모델(예: 확률미분방정식, 은닉 마르코프 모델)를 적용하여 그 전이 상태를 고려, 반복 학습 모델링을 통해 향후 봇의 이주 혹은 재접속 가능성을 분석한다.

(4) 봇넷 탐지 알고리즘

- IRC/HTTP 봇넷 C&C 탐지는 기존의 다양한 탐지 기법을 응용하여 DNS쿼리 정보 기반 봇 C&C 및 좀비리스트등을 탐지한다. 또한, DNS쿼리 정보 및 네트워크 트래픽 정보 분석을 통한 봇C&C 및 의심 좀비 리스트를 탐지한다.

(5) 탐지 검증 모듈

- 탐지된 봇넷의 진의 여부를 판별하기 위해 다양한 통계기법(ROC커브, 몬테카를로등)을 이용함. 또한 오탐율을 공시하여 신/변종 봇넷의 리스트를 작성함.

이러한 단계적 마이닝 기법은 다음과 같은 특징을 가진다.

(1) 봇넷의 분포도의 상태 변화를 보고 그 특징을 마이닝하여 탐지 하는 기술이기 때문에 특히 암호화 통신 및 스텔스 스캐닝에 효과적인 성능을 보인다.

(2) DNS 쿼리 정보를 마이닝하여 쿼리의 가부(可否) 여부를 따지고 그룹핑하여 탐지회피를 막고 나아가서 가짜 DNS 쿼리 집단의 향후 감염 여부를 사전에 측정하여 그 확산도를 그려낼 수 있음.

4. 모델링 비교 분석 검증

본 논문에서 제안한 마이닝 방법과 기존의 봇넷 탐지를 위한 알고리즘에 대해서 분석 모듈별로 비교 분석 하였다.

<표 2> 기존 봇넷 탐지 알고리즘 vs. 제안모델 프로세스

	[1]	[2]	[3]	[4]	제안모델
봇넷분류분석	*	-	*	*	*
봇넷구성분석	-	-	-	*	*
봇넷행동분석	*	*	*	*	*
봇넷학습분석	-	-	-	-	*
봇넷탐지분석	*	*	*	*	*
봇넷검증분석	-	-	-	-	*
신종봇넷탐지	-	-	-	*	*
변종봇넷탐지	-	-	-	*	*

표 2에서 보듯이, 기본적인 알고리즘 구성도에 따라 능동형 봇넷 탐지에 대한 성능을 비교 할 수 있다. 이는 현재의 봇넷 탐지 보다 향후 발생 가능한 신종/변종 봇넷의 탐지 측면에서 중요한 비교 분석 결과라 할 수 있다.

5. 결론

본 논문에서 제안한 능동형 봇넷 탐지 및 관리를 위한 단계적 마이닝 프로세스는 다양한 봇넷 탐지 알고리즘에 직/간접적으로 적용이 가능하며, 기존의 봇넷 탐지의 어려움을 해결함으로써 웜, 바이러스등 다양한 사이버 위협에 대해서도 광범위적으로 적용이 가능하다. 또한, 마이닝 기법은 기존의 봇넷 탐지의 한계를 극복할 수 있다. 즉, 마이닝 기법을 통해 신종 및 변종 봇과 같은 우회성 봇의 탐지를 가능하게 한다. 마지막으로 기존의 탐지 알고리즘에 없는 검증 모듈의 도입은 현 봇넷의 실효성을 알 수 있는 중요한 정보가 될 수 있다. 따라서 향후 연구로 최상의 탐지 성능과 정확성을 보여주기 위한 실질적인 알고리즘을 적용하고 검증하고자 한다.

Acknowledgment

본 연구는 지식경제부 및 정보통신연구진흥원의 IT핵심기술개발사업의 일환으로 수행하였음. [2008-S-026-01, 신종 봇넷 능동형 탐지 및 대응 기술]

참고문헌

[1] D. Dagon (at Georgia Tech), "Botnet detection and response", CAIDA DNS-OARC Workshop 05

[2] N. F. A. Ramachandran (at Georgia Tech), "Revealing botnet membership using dnsbl counter-intelligence", USENIX SRUTI 06.

[3] Jan Goebel, "identify bot contaminated hosts by IRC nickname evaluation", USENIX 07.

[4] Ping Wang, Sherri Sparks, and Cliff C. Zou, "An Advanced Hybrid Peer-to-Peer Botnet", USENIX HOTBOT 07.

[5] 박찬호, 강권학, 권영찬, 장희진, 김철호, "메모리 감시를 이용한 허니팟 기반의 봇넷 역추적", 2007 한국컴퓨터종합학술대회 논문집.