봇넷의 악성행위 탐지를 위한 임계치 측정 방법론

김도훈*, 1)인호*, 정현철**
*고려대학교 컴퓨터·전파통신 공학과
**한국인터넷진흥원
e-mail:*{karmy01, hoh_in}@korea.ac.kr, **hcjung@kisa.or.kr

Threshold estimation methodology for detection of Botnet malicious activity

Do-Hoon Kim*, Hoh Peter In*, Hyun-Cheol Jeong**
*Dept of Computer and Radio Communications Engineering, Korea University
**Korea Internet & Security Agency

요 약

최근 7.7 DDoS 대란과 폭발적인 SPAM 발송과 같은 다양한 봇넷의 악성행위는 정보 시스템에 막대한 악영향을 미친다. 특히, 봇넷의 구조적 특징인 좀비PC의 제어는 네트워크 환경에서는 다양한 악성행위를 유발한다. 때문에, 봇넷 탐지와 관련한 다양한 연구가 시도되었지만, 탐지의 한계점을 지니고 있다. 즉, 기존의 봇넷 탐지 방법은 임의의 임계값을 설정하고, 그 값을 벗어나는 시점에서 경고를 보내어 탐지하게 된다. 하지만, 전문가에 의한 임계값 설정은 자칫 오탐율과 미탐율을 야기할 수 있다. 따라서, 본 논문에서는 봇넷 탐지를 보다 능동적으로 하기 위하여 특정 타임 윈도우 구간동안의 봇넷이 유발하는 네트워크 트래픽을 분석하고 마르코프 체인을 응용한 학습을 하여 능동적으로 적용 가능한 임계값을 측정 방법론에 대하여 고찰하고자 한다.

1. 서론

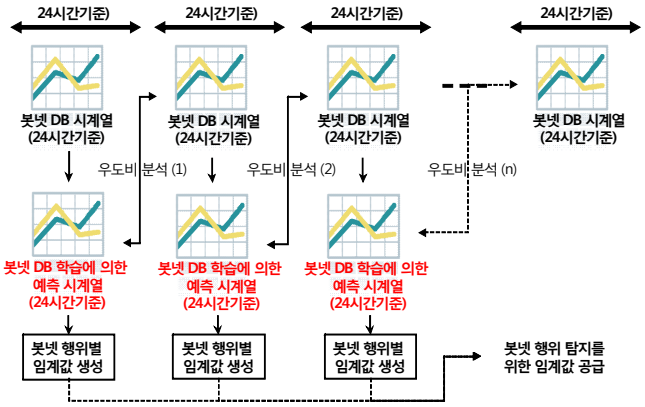
최근 지능적으로 진화하는 봇넷은 7.7 DDoS 대란을 야기 시켰다. 7.7 DDoS 대란의 공격유형은 대부분이 HTTP Get과CC Attack이었으며, UDP, ICMP, SynFlooding 공격도 병행 되어졌다. 결국, 이러한 공격으로 인하여 기관 및 업체 서버의 네트워크 자원 고갈을 유발 시켰으며, 네트워크 장비 장애로 정상적인 서비스를 불가능하게 하였다. 이처럼 봇넷의 탐지도 중요하지만, 봇넷이 유발하는 악성행위는 그 피해 파급이 막강함으로 지속적인 관찰과 분석이 필요하다.

따라서, 본 논문에서는 탐지된 봇넷 그룹의 악성행위 요인(DDoS, SPAM, Propagation)을 탐지하기 위해 관찰 요인을 분석하고, 능동적인 악성행위를 결정하기 위한 학습 프로세스를 제안하고자 한다. 즉, 탐지된 봇넷 그룹이 발생시키는 네트워크 트래픽을 특정 타임 윈도우 동안 관찰 및 분석하여 학습을 하고 이를 기반으로 탐지를 위한 동적 임계값을 산정하여 봇넷의 악성행위를 탐지하고자 한다. 기본적으로 악성코드의 임계값 측정을 위한 관련연구는 다음과 같다. 관련연구[1]은 좀비들이 유발하는 DNS쿼리의 분석을 통하여 그룹행위를 인지하여 봇넷을 탐지하는 기법으로 경험적 임계값 설정에 의한 탐지를 수행한다. 또한, 관련연구[2]에서는 TRW(Threshold RandomWalk) 알고리즘을 응용하여 봇넷의 밀도를 측정하고 탐지하는 척도로 사용하였다. 마지막으로 관련연구[3]에서는 호스트로부터 관찰되어지는 비정상 행위를 통하여 로그 사이즈의 변화를 임의의 임계값으로 탐지하여 봇넷을 탐지하는 알고리즘을 계발하였다.

하지만, 동적으로 변화하는 네트워크 환경에서는 오탐 및 미탐이 발생할 수 있다. 따라서, 본 논문에서는 특정 타임 윈도우 기간을 설정하고 관찰 및 학습을 하여 임계값의 추이를 분석하여 능동적으로 봇넷 악성행위 탐지에 사용하고자 한다.

2. 봇넷 임계값 측정 방법

본 논문에서는 효과적인 봇넷의 임계값을 산정하기 위해서 그림1. 과 같은 개념도를 그릴 수 있다.

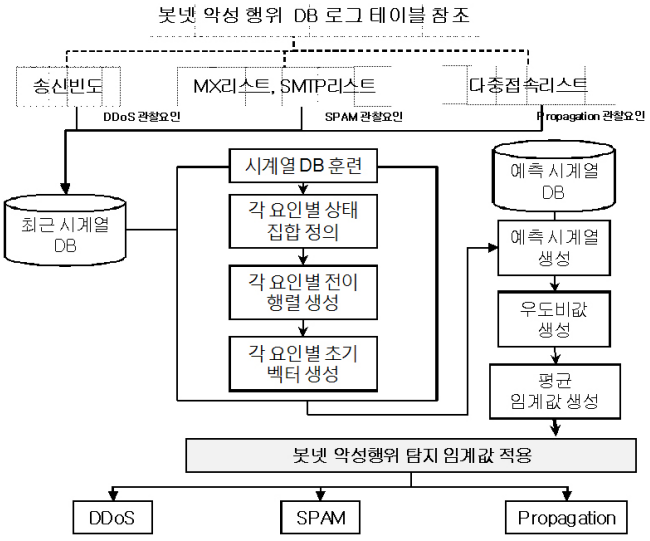


(그림 1.) 봇넷 임계값 예측 개념도

1) 교신저자

그림1.과 같이 24시간 기준으로 관찰되어진 봇넷 DB 시계열을 마르코프 기반의 학습을 통하여 향후 24시간의 예측 시계열을 생성한다. 이렇게 생성된 정보는 다시 24시간 이후의 관찰 24시간 데이터와 우도비 분석을 하여 적합한 평균 임계값을 생성하게 된다. 이러한 분석 방법을 통하여 충분히 훈련을 거쳐서 최적의 봇넷 악성 행위 탐지를 위한 임계값을 생성하게 된다.

다음은 구체적인 봇넷 임계값 산정 프로세스이다.



(그림 2.) 봇넷 임계값 측정 및 적용 프로세스

즉, 그림2.을 통하여 다음과 같이 프로세싱을 한다.

① 현 기준 24시간에 대한 해당 탐지된 봇넷 그룹의 트래픽을 관찰 (송신빈도 / MX·SMTP리스트 / 다중접속리스트)

② 이전 24시간 시계열 DB를 참조하여 시계열 DB 훈련한다. 이때 시계열 표현은 백분율로 한다. 아래는 훈련시에 필요한 요인들에 대한 정의이다

*** 요인별 상태 집합 정의**

송신빈도: $S_{sendFreq} = \{S_1, S_2, S_3, S_4\}$

MX리스트: $S_{mxList} = \{S_1, S_2, S_3, S_4\}$

SMTP리스트: $S_{smtpList} = \{S_1, S_2, S_3, S_4\}$

다중접속리스트: $S_{multiList} = \{S_1, S_2, S_3, S_4\}$

*** 각 요인별 전이 행렬 생성**

$$T = \begin{array}{c|cccc} & S_1 & S_2 & S_3 & S_4 \\ \hline S_1 & & & & \\ S_2 & & \frac{\text{전이횟수}}{\text{각열의총전의횟수}} = \frac{a}{a+b+c+d} & & \\ S_3 & & & & \\ S_4 & & & & \end{array}$$

*** 초기확률 벡터(π) 생성**

$$\pi = \left(\frac{S_1}{S_1+S_2+S_3+S_4}, \frac{S_2}{S_1+S_2+S_3+S_4}, \frac{S_3}{S_1+S_2+S_3+S_4}, \frac{S_4}{S_1+S_2+S_3+S_4}\right)$$

③ 현시점으로부터 다음 24시간 예측시계열 생성 및 저장

* 생성된 전이 행렬(T)와 초기벡터(π)을 곱으로 나타낸다. (단, 생성된 벡터들 중에 제일 큰 값을 P로 정한다.) ==> **P = π * T**

* 예상 발생 25시간 일때 시계열 값은 다음과 같다.
==> **FT = P * 24시간 일때의 시계열 값**

* 이런식으로 매시간 시계열 값을 예측한다

④ 예측시계열 DB를 참조 실제 시계열과의 우도비 분석을 하여 정확도 측정 및 평균 임계값 생성

$$* \ -2LL = -2\sum_{i=1}^{n} Y_i \ln P_i + (1-Y_i)\ln(1-P_i)$$

Y_i : 관측값, P_i : 예측값

* 우도가 0에 접근할수록 적합도는 높다.

⑤ 평균 임계값을 봇넷의 행위 탐지 임계값으로 적용

⑥ ③~⑤ 반복 시행 (경험적으로 1개월 기준)

⑦ 1개월 후에 최적의 임계값 설정 적용에 사용

3. 결론

본 논문에서는 기존의 경험에 의한 임계값 설정 방법 대신, 특정 타임 윈도우 동안에 관찰되어지는 해당 봇넷의 네트워크 트래픽을 분석하여 임계값을 산정하는 분석 알고리즘을 제안하였다. 특히, 학습에 의한 임계값 산정 방법은 동적으로 변화하는 네트워크 환경에 맞게 능동적으로 적용 가능하겠다.

향후 연구로는, 학습시에 발생하는 관찰요인의 증가에 따른 행렬 폭발 문제의 해결과 검증에 필요한 실험 그리고, 학습에 따른 오탐과 미탐에 대하여 기존의 연구와 비교 분석 절차가 요구된다.

Acknowledgment

"본 연구는 지식경제부 및 한국산업기술평가관리원의 IT 산업원천기술개발사업의 일환으로 수행하였음.
[2008-S-026-02, 신종 봇넷 능동형 탐지 및 대응 기술]"

참고문헌

[1] H. Choi, H. Lee, H. Lee, H. Kim, "Botnet Detection by Monitoring Group Activities in DNS Traffic", IEEE CIT, Oct. 2007.

[2] Gu, G., Zhang, J., Lee, W.: BotSniffer: Detecting Command and Control Channels in Network Traffic. In Proceedings of the 15th Annual Network and Distributed System Security Symposium (NDSS'08), San Diego, CA, February 2008

[3] Al-Hammadi, Y., Aickelin, U.: Detecting Botnets Through Log Correlation. In Proceedings of the Workshop on Monitoring, Attack Detection and Mitigation (MonAM 2006), Tuebingen, Germany, 2006