DoS 공격의 순차의존성 측정에 관한 연구

홍동호°, 유황빈
광운대학교 컴퓨터과학과
dhhong@netlab.kwangwoon.ac.kr, ryou@kwangwoon.ac.kr

A Study on Measuring Sequential Dependencies in Denial of Service Attacks

Dongho Hong°, Hwangbin Ryou
Dept. of Computer Science, Kwangwoon University

요 약

서비스거부공격(DoS)은 공격의 목표가 되는 호스트에 악의적인 패킷을 단지 하나만 보냄으로써 공격이 이루어 질 수 없다. 그렇기 때문에 침입탐지시스템에서는 패킷에 대한 일련의 순차성을 알아냄으로써 연속적인 패킷을 보내는 DoS 공격으로서의 침입을 탐지해 내는 것이 가능하다.

본 연구에서는 네트워크 패킷에 대한 송신지의 주소와 서비스, 목적지의 주소와 서비스를 이용하여 이벤트를 정의 하였으며, 이렇게 정의된 이벤트를 이용하여 정상적인 상황에서의 이벤트에 대한 순차의존성과 DoS 공격 상황에서의 이벤트에 대한 순차의존성을 알아내었다. 이벤트에 대한 순차의존성을 측정하기 위해서 정상 이벤트가 발생하는 확률을 이용하였으며, 더 나아가서는 정보이론 분야에서의 조건부 엔트로피의 사용도 제안하였다.

1. 서 론

침입(Intrusion)이란 권한을 부여받지 않은 사용자가 시스템에서 혹은 시스템의 자원에 접근하여 그것의 기밀성, 무결성, 가용성을 침해하는 단일, 또는 일련의 보안 사건을 의미한다. 이러한 침입을 이루는 방법에는 여러 종류가 있으며, 그것을 탐지하는 것이 침입탐지시스템(Intrusion Detection System)이다.

침입탐지시스템(IDS)은 크게 단일 호스트로부터의 감사 자료를 이용하여 침입을 탐지하는 호스트 기반의 침입탐지 시스템(Host-based Intrusion Detection System)과 네트워크의 트래픽 데이터를 침입 탐지에 이용하는 네트워크 기반 침입탐지시스템(Network-based Intrusion Detection System)으로 나뉜다. 요즘에는 위 두 가지의 장점만을 이용한 하이브리드 타입(Hybrid type) 침입탐지시스템을 사용하는 경우도 있다. 이러한 침입탐지시스템에는 대표적으로 두 가지의 침입탐지 방법이 사용되어지는데, 그 첫 번째가 해당 인프라, 시스템 혹은 응용 프로그램의 취약성을 이용한 잘 정의된 침입을 탐지하는 기법으로 이것을 오용탐지(misuse detection)이라고 한다. 또 다른 한 가지는 시스템의 비정상적인 행위를 사전에 구성한 자료에 근거하여 침입을 탐지하는 비정상행위 탐지(anomaly detection)가 있다.

현재의 NIDS에서는 대부분 규칙(rule)을 기반으로 한 오용탐지 기법을 사용하고 있으나 규칙에 없는 새로운 공격을 탐지하지 못한다는 큰 단점을 가지고 있어, 비정상행위 탐지 방법의 필요성이 대두되고 있는 실정이다.

본 연구에서는 네트워크에서의 DoS공격에 대한 비정상적인 탐지를 하기 위해서 이벤트 발생 확률을 이용하였으며 연구 데이터로는 DARPA에서 이뤄진 오프라인 데이터를 이용하였다[1].

2. 조건부 엔트로피(Conditional Entropy)

정보이론(Information Theory)에서의 엔트로피(Entropy)는 단일 변수에 대하여 우리가 모르고 있는 발생의 정도를 측정할 수 있는 개념이다. 이것은 곧, 각 항목에 대한 발생 불순도(불확실성)를 측정할 수 있는 도구로 활용될 수 있다는 의미이다[2]. 다음은 엔트로피를 구하는 식을 표기한 것이다[3].

$$H(X) = \sum_{x \in C_x} P(x) log \frac{1}{P(x)}$$

그러나 엔트로피는 어떠한 항목이나 이벤트에 대한 시간성이나 순차성을 적용할 수 없고, 어떠한 항목이나 이벤트에 대한 순차의존성을 적용하기 위해서는 조건부 엔트로피(Conditional Entropy)를 사용해야 한다[2].

주어진 Y에 대한 X의 조건부 엔트로피 *(H(X|Y))*는 확률분포 P(x|y)로서,

$$H(X|Y) = \sum_{x,y \in C_x, C_y} P(x,y) log \frac{1}{P(x|y)}$$

와 같이 표기된다[2][3]. 위 정의를 설명하기 위해 다음의 예를 들어보자. X를 $\{e_1, e_2, e_3, \ldots e_{n-1}, e_n\}$과 같이 표시되는 이벤트라 하고, Y를 $\{e_1, e_2, e_3 \ldots e_k\}$와 같이 표시되는 이벤트라고 할 때(k<n), 조건부 엔트로피는 y 이후의 x에서 y와 같은 순차성이 얼마나 많은 불확실성을 가지고 나타날 수 있는가를 말해준다. 예를 들어 X가 동일한 패턴의 연속이라면 그에 대한 조건부 엔트로피는 0이 된다.

3. DoS 공격에서의 순차성

서비스거부(Denial of Service) 공격은 시스템 혹은 네트워크 자원을 모두 고갈시켜 실제 사용자에게 서비스를 방해하거나 무력화하는 공격법이다. 일례로 2000년 초 야후와 아마존이 분산서비스거부공격으로 인해 피해를 입었고, 최근에는 분산서비스거부공격의 다양성과 위험성의 증가되고 있는 추세이다. 이러한 공격들의 가장 큰

표 1. 대표적인 DoS 공격에서의 순차성

공격	순차성
SYN Flooding 공격	TCP 연결 설정에서의 순차성 이용 1. 호스트는 원격지로부터 SYN 패킷을 받음 2. 호스트는 원격지로 SYN + ACK 패킷을 보냄 3. 호스트는 원격지로부터 ACK 패킷을 받지 못함
ping of death 공격	호스트에 데이터의 길이가 큰 ICMP 메시지가 짧은 시간 내에 대량으로 들어옴
Land attack 공격	송신지와 목적지의 주소와 포트 번호를 동일하게 보냄 1. 호스트는 원격지로부터 목적지 주소가 자신인 SYN 패킷을 받음 2. 호스트는 자기 자신에게 패킷을 보냄
IP scan	일련의 시간 간격을 갖고 ICMP 메시지를 이용하여 네트워크 호스트의 IP를 검사 호스트(목적지) IP의 연속성
Port scan	일련의 시간 간격을 갖고 TCP를 이용하여 호스트의 포트를 검사 호스트(목적지) 포트의 연속성 TCP 연결의 설정에서의 순차성 이용

특징은 송신지 IP를 위장하는(Spoofing) 것이다. 이와 같은 특징 때문에 DoS 공격은 특정한 일련의 패턴을 갖게 된다. 예를 들어 같은 내용의 패킷이 되풀이 하여 탐지된다던가, 또는 단 시간 내에 계속 위조된 송신지 IP로 패킷이 들어 올 수도 있다.

위의 표는 지금까지 잘 알려진 DoS공격들에 대한 일련의 순차성을 알아본 것이다.

4. 순차의존성 검사 방법

4.1 탐지 모듈 구조

본 연구에서의 침입 탐지 모델은 DARPA의 1999년도 연구 데이터와 테스트 환경에서 DoS 공격을 가한 호스트의 tcpdump 데이터를 사용하였다.

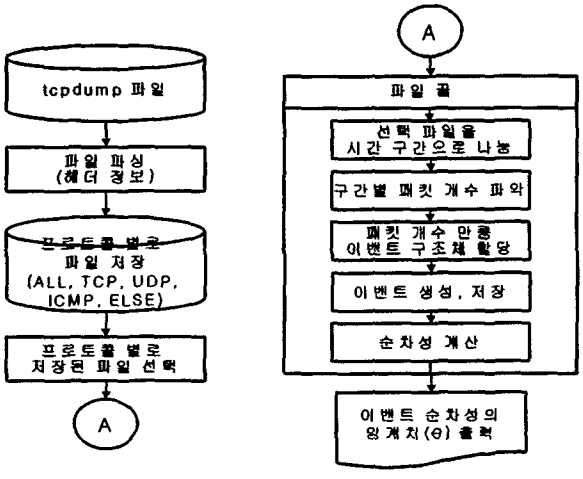

그림 1. 조건부 엔트로피를 이용한 오프라인에서의 트래픽 순차성 탐지 모듈 구조

그림 1은 본 연구에서 트래픽의 순차의존성을 계산하기 위한 탐지 모듈의 구조도이다. tcpdump파일을 처리하여 각 헤더에서 필요한 필드 정보를 추출하여 프로토콜을 기준으로(ALL:모든 프로토콜 포함, TCP : TCP 프로토콜만 처리, UDP : UDP 프로토콜만 처리, ICMP : ICMP 프로토콜만 처리, ELSE : TCP, UDP, ELSE을 제외한 모든 프로토콜 처리) 나눈 후 파일로 저장한다. 그 후 원하는 프로토콜을 선택하여 사용자 입력 시간 별로 구간을 나누고 나서 각각의 패킷을 정의한다. 이렇게 생성된 패킷에 대한 정의는 이벤트로 조합되어 각 프로토콜에 대한 순차의존성을 검사하게 된다[4].

다음의 그림은 TCP의 정상적인 연결에 대한 개수를 파악하기 위한 의사코드이다.

```
if (i 번째 패킷의 flag == "SYN")
 (j = i+1)
 if((i번째 패킷 Source IP == j번째 패킷 Dest IP)
  &&(i번째 패킷 Dest IP == j번째 패킷 Source IP)
  &&(i번째 패킷 Source port == j번째 패킷 Dest port)
  &&(i번째 패킷 Dest port == j번째 패킷 Source port)
  &&(현 패킷 flag == "SYN,ACK"))
  (k = j+1)
   if((i번째 패킷 Source IP == k번째 패킷 Dest IP)
    &&(i번째 패킷 Dest IP == k번째 패킷의 Source IP)
    &&(i번째 패킷 Source port == k번째 패킷 Dest port)
    &&(i번째 패킷 Dest port == k번째 패킷 Source port)
    &&(현 패킷 flag == "SYN,ACK"))
          정상 연결 회수++ ;
```

그림 2. TCP 연결 설정에 대한 순차성 검사 모듈의 의사코드

본 연구에서 트래픽의 순차성을 계산하는 식은 그림 3과 같다.

구간 내의 패킷 수 : n
이벤트 조합에 필요한 패킷 수 : k
구간 내 발생 가능한 이벤트 수 : m
실제 이벤트 발생 수 : realocc
이벤트 발생 가능 확률 : p

$$m = n - (k-1)$$

$$p = \frac{realocc}{m} * 100$$

그림 3. 이벤트 발생 확률 식

TCP 연결에 대한 순차성을 계산하는 식을 예로 들자면

한 구간 내에 존재하는 패킷의 수가 500개라 하자. 그리고 이벤트 조합에 필요한 패킷 수는 3개(SYN, SYN+ACK, ACK) 이므로 구간 내 이벤트 발생 가능 수는 498개가 된다. 실제 발생한 이벤트 숫자를 구간 내에 발생 가능한 이벤트 수로 나누면 실제 발생한 이벤트의 확률을 얻을 수 있다.

4.2 실험 환경

다음 그림은 공격 트래픽을 수집하기 위한 테스트 환경이다. 내부 네트워크 주소를 할당하여 그 중 하나를 공격 대상 호스트로 사용하였고, 외부 네트워크에 공격 호스트를 위치시켰다.

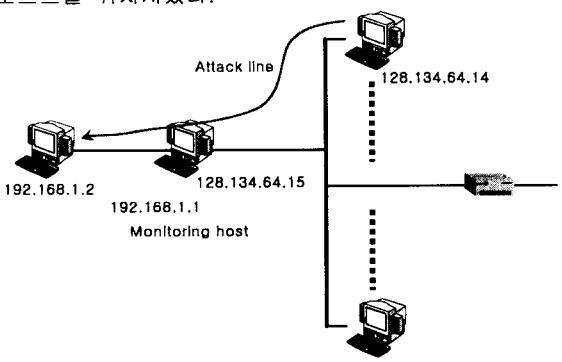

그림 4. 공격 트래픽 수집 환경

모니터링 호스트는 dual-homed gateway로 구성하였으며, 데이터 수집도구로는 tcpdump를 사용하였다.

5. 실험 결과

5.1 정상 트래픽과 공격 트래픽에서의 순차성 비교

정상 트래픽에 대한 오프라인 실험 데이터는 1999년 DARPA에서 만들어진 데이터를 사용하였다[3]. 데이터에서 구해진 순차의존성의 임계치는 바로 온라인상에서 트래픽에 대한 순차의존성과 비교하여 비정상행위에 대한 침입 탐지를 해 낼 수 있다.

표2는 DARPA의 1999년 오프라인 자료의 3월 셋째 주의 공격이 없는 정상 트래픽에 대한 TCP 정상 연결에 대한 정상 이벤트 조합의 발생 확률을 계산한 것이다.

표 2 정상 트래픽에서의 이벤트 발생 확률

	월	화	수	목	금	평균
결과값	5.414	4.55	4.844	5.335	4.859	5.00

표2와 같이 정상적인 트래픽에서는 평균 5.0라는 결과값이 나왔으나, 본 실험에서 teardrop, SYN Flooding, Ping of Death, Port scan 공격을 직접 시도한 경우에 대해서는 SYN Flooding과 Port scan에 대해서 상이한 결과 값이 나왔다.

표 3 공격 트래픽에서의 이벤트 발생 확률

공격	공격 탐지 시간	순차값 평균
SYN Flooding	05:13:14 ~ 05:16:16	0.082
	05:45:22 ~ 05:47:11	0.014
	06:23:25 ~ 06:28:13	0.090
	06:45:14 ~ 06:49:13	0.048
	07:06:14 ~ 07:09:13	0.002
평균		0.050
Port Scan	05:42:22 ~ 05:43:13	0.035
	05:58:14 ~ 05:59:13	0.166
	06:11:14 ~ 06:12:13	0.022
	06:53:15 ~ 06:54:09	0.011
평균		0.060

표2와 표3에서 본 보와 같이 정상 트래픽과 공격 트래픽(SYN Flooding, Port Scan)에서의 순차성이 확연히 차이가 나는 것을 알 수 있다. 즉, 네트워크 패킷의 순차성을 파악하여 그 이상 유무를 알아 낼 수 있다.

6. 결론 및 향후 연구

본 논문에서는 네트워크 트래픽에서의 정상적인 순차성과 비정상적인 순차성의 차이를 보여 그것이 공격을 판단하는 기준이 될 수 있음을 보였다. 정상적인 순차에 대한 확률을 이용하여 그 기준을 보였지만 그 확률을 토대로 조건부 엔트로피를 이용하면 더 명확한 결과를 얻어 낼 수 있다. 그리고 본 연구에서는 단지 TCP에 대해서만 그 분야를 한정하여 실험하였는데 SYN Flooding과 Port Scan 공격은 그 순차성에 대해서 민감한 반응을 보였다. 다른 프로토콜에서도 충분한 양의 정상 데이터를 이용한 트레이닝을 거쳐서 그 순차의존성을 파악할 수 있을 것이다.

앞으로의 연구에서는 패킷 이벤트를 정의하는데 있어서 패킷의 각 헤더 필드를 좀더 유기적으로 조합하고, 정상 트랙픽에서 좀더 많은 트레이닝을 한 임계치를 구해내야 할 것이다. 그리고 같은 방식으로 여러 번의 실험을 통해 얻은 결과를 통계적으로 표현해야 할 것이다. 현재 정상 트래픽은 DARPA 데이터만 이용하고 있으나, 이것으로서는 정상 트래픽의 신뢰적인 임계치를 구하는데 그 양이 부족하므로, 소량의 데이터라는 단점을 극복하기 위해서는 Cross-validation을 이용해야 한다.

참고 문헌

[1] 'MIT Lincoln Laboratory - DARPA Intrusion Detection Evaluation' URL: http://www.ll.mit.edu/IST/indeval/index.html

[2] Wenke Lee, Dong Xiang "Information-Theoretic Measure for Anomaly Detection", Proceedings of the IEEE Symposium on Security and Privacy, May, 2001

[3] THOMAS M. COVER, JOY A. THOMAS, "Element of Information Theory"

[4] Protocol Anomaly Detection in Network-based IDSs URL: http://erwan.lemonnier.free.fr/exjobb/report/protocol_anomaly_detection.pdf