은닉마코프모델을 이용한 이상징후 탐지 기법*

이은영, 한찬규, 최형기
성균관대학교 정보통신공학부
{eylee, hedwig, hkchoi}@ece.skku.ac.kr

An Anomaly Detection based on Probabilistic Behavior of Hidden Markov Models

Eunyoung Lee, Chan-Kyu Han, Hyoung-Kee Choi
Dept. of Computer Engineering, Sungkyunkwan University

요 약

인터넷의 이용이 증가함에 따라 네트워크를 통한 다양한 공격 역시 증가 추세에 있다. 따라서 네트워크 이상징후를 사전에 탐지하고 상황에 따라 유연하게 대처할 수 있도록 하기 위한 연구가 절실하다. 본 연구는 은닉마코프모델을 이용해 트래픽에서 이상징후를 탐지하는 기법을 제안한다. 제안하는 기법은 시계열 예측 기법을 이용해 트래픽에서 징후를 추출한다. 징후추출 과정의 결과를 은닉마코프모델을 활용한 징후판단과정을 통해 네트워크 이상징후인지를 판단하고 결정한다. 일련의 과정을 perl 로 구현하고, 실제 공격이 포함된 트래픽을 사용하여 검증한다. 하지만 결과가 확연히 증명되지는 않는데, 이는 학습과정의 부족과 실제에 가까운 트래픽의 사용으로 인해 나타나는 현상으로 연구의 본질을 흐리지는 않는다고 판단된다. 오히려 실제 상황을 가정했을 때 접근이나 적용을 판단함에 관리자의 의견을 반영할 수 있으므로 공격의 탐지와 판단에 유연성을 증대시킬 수 있다. 본 연구는 실시간 네트워크의 상황 파악이나 네트워크에서의 신종 공격 탐지 및 분류에 응용 가능할 것으로 기대된다.

1. 서론

인터넷 수요 증가는 경제, 문화산업에 막대한 가치를 창출시키고 있다. 그러나 인터넷 수요와 함께 네트워크를 통한 시스템 침입, 악성코드 유입, 네트워크 공격 등과 같은 위협 역시 빠른 속도로 증가하고 있다. 이로 인해 네트워크 공격의 탐지, 진단, 분석은 중요한 연구 분야로 각광받고 있으며, 네트워크 공격 탐지와 관련된 연구는 중요한 부분을 차지하고 있다.

네트워크 공격의 탐지는 규칙 기반(Rule-based) 탐지와 이상징후 기반(Anomaly-based) 탐지로 구분한다. 규칙기반은 사후처리에 치중하는데, 규칙 기반 탐지로는 기존 침입형태에서 변형된 형태의 변종 공격이나 새로운 형태의 공격을 예방하기가 어렵다. 반대로 이상징후 기반은 취약점이나 침입을 미리 탐지하는 것에 중점을 둔다. 그러나 아직까지는 오탐율과 미탐율이 매우 높은 실정이다. 오탐율과 미탐율을 줄이는 것이 이상징후 기반 탐지의 최우선과제이다.

따라서 본 논문은 네트워크 상황에 유연하게 대처할 수 있도록 하기 위한 이상징후 탐지기법을 제안하며, 관련한 연구 성과를 기술한다. 2 장에서는 이상징후 탐지와 관련된 기존 연구와 은닉마코프모델(Hidden Markov Model)에 대해서 기술한다. 3 장에서는 네트워크 이상징후 탐지에 기반이 되는 징후 추출과정을 설명하고, 4 장에서는 은닉마코프모델을 이용한 확률 예측 모델에 대하여 기술한다. 5 장에서는 구현 및 데이터에 대해 설명한다. 6 장은 결과 및 분석에 대해 다루고, 마지막으로 7 장에서 전체적인 연구 성과에 대한 논의로 결론을 맺는다.

2. 관련연구

2.1. 이상징후 탐지

대부분의 이상징후 탐지에 관한 연구는 모델링을 통해 이루어진다. 이러한 접근법은 정상 데이터에 대한 일종의 모델 구축을 통해 새로운 데이터를 사용자가 설정한 기준으로 판별하는 방식으로 이루어진다 [1][2][3].

- 연속적인 시퀀스와 서두의 쌍을 사용한 정상 시퀀스 모델링을 위한 접근법 연구.
- 정상 트레이스에서 발생되는 알려진 시퀀스와 침입에서 발생될 예측 시퀀스를 비교한 후 순위를 매기는 통계적인 메소드 연구.
- 가능한 False Alarm Rate 를 줄이는 동안 새로운 공격을 탐지하는 가상 베이지언을 사용한 연구.
- 정상 데이터의 확률 모델 매개변수 구축을 통한 연구.

이상에서 기술한 이상징후 탐지와 관련한 연구는

* 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었음 (IITA-2008-C1090-0801-0028)

이외에도 다양한 형태로 진행되고 있다.

2.2. 은닉마코프모델

은닉마코프모델은 미지(hidden)의 확률론적 과정(stochastic process)을 관찰 가능한 기호(symbol)를 발생시키는 다른 확률론적 과정을 통해 모델링하는 이중 확률론적 과정을 말한다[4][5].

은닉마코프모델은 두 개의 상태 집합과 세 개의 확률 집합으로 구성되는 요소를 가진다. 두 개의 상태 집합은 *은닉상태 집합*(hidden state set)과 *관찰가능상태 집합*(observable state set)을 말한다. *은닉상태 집합*은 관찰 불가능한 상태를 말하고, *관찰가능상태 집합*은 외형적으로 눈에 보이는 전이 상태들의 집합을 말한다.

확률 집합은 *초기값 벡터*(vector), *상태전이 행렬*(state transition), *관찰확률 행렬*(emission distribution)이다.

초기값 *벡터*( )는 특정 은닉상태가 시간 = 일 때 모델의 확률을 말한다. 즉, 의 번째 원소는 수식 (1)과 같이 초기 상태인 이 일 확률을 나타낸다.

$$\Pi:\ \mathcal{N}\times 1,\ \Pi_{q_i}=\mathcal{P}(q_i),\ q_i\in\mathcal{N} \tag{1}$$

상태전이 행렬( )은 이전의 은닉상태에서 현재의 은닉상태로의 전이 확률이다. 수식 (2)는 모든 , 대하여, 시각 에서의 상태가 일 때, 시각 +1번째에서 상태가 일 확률을 나타낸다.

$$\mathcal{S}:\mathcal{N}\times\mathcal{N},$$
$$\mathcal{S}_{ij}=\mathcal{P}\{q_j(t+1)=q_j \mid q_i(t)=q_i\},\ q_i,q_j\in\mathcal{N} \tag{2}$$

관찰확률 행렬( )은 특정 은닉상태에서의 관찰 가능한 상태들의 확률을 나타내는 행렬이다. 수식 (3)과 같다.

$$\mathcal{T}:\mathcal{N}\times\mathcal{M},\mathcal{T}_{ij}=\mathcal{P}\{f_j(t)=f_j \mid q_i(t)=q_i\},\ f_j\in\mathcal{M} \tag{3}$$

은닉마코프모델은 *관찰가능상태 집합*의 관측열에서 상태전이 경로를 계산한다. 이렇게 얻어진 경로를 통해 *은닉상태 집합*의 열을 예측할 수 있게 된다. 예측의 신뢰성을 높이기 위해서는 초기값 벡터, 상태전이 행렬, 관찰확률 행렬에 대한 학습 및 최적화 과정이 필요하다. 자세한 과정은 본 논문에서 생략하며, [4]에 기술되어 있다.

3. 징후추출

징후는 이상징후 과정 또는 그 과정의 결과로 인해 네트워크 트래픽에 나타나는 증상으로 정의한다. 본 논문에서는 징후에 대해 은닉마코프모델을 적용함으로써, 이상징후를 효율적으로 탐지하고자 한다. 그러나 징후를 직접적으로 관측하는 것은 상당히 어려운 작업이다. 이에 본 연구에서는 시계열 트래픽 정보를 이용하여 징후를 추출하고 이를 관측하고자 한다.

시계열 예측법은 시간에 따라 주기적으로 변하는 트래픽의 주기성을 시계열 모델을 통해 예측하고, 예측 범위를 벗어나는 트래픽을 탐지한다. 본 논문에서 사용한 트래픽 정보는 12 가지이다: Destination IP usage, Number of open port, Packet arrival rate, Payload size, Payload size variation per packet, Port number variation, Session time, Payload size per session, Number of session per port, Source, Destination IP ratio, Destination IP distribution, Reflection traffic.

본 논문에서 시계열 예측을 이용해 추출하고자 하는 징후는 다음과 같다.

- Port Scanning: Port Scanner 를 이용, 목적지 system 의 취약한 port 를 찾는 행위를 말한다. 공격자는 scanning 을 통해 침입 가능한 port 가 있는지 검사한다.
- IP Scanning: 공격자들은 Port Scanning 과 마찬가지로 공격대상을 탐색하는 방법으로 이용한다.
- IP Spoofing: 다른 IP 주소에서 전송된 것처럼 보이는 패킷을 생성하는 것을 의미한다. 이러한 방법은 DoS 공격에 주로 사용된다.
- Packet Spoofing: 공격자가 패킷의 Header 나, Payload 를 임의대로 수정한다. 이를 공격대상 시스템에 전송해 Checksum error 를 발생시킨다.
- Sudden Increase: 네트워크에 갑작스럽게 Packet 의 요청이나 응답이 늘어나는 경우 공격을 의심할 수 있다. 이는 공격자가 악의적인 목적으로 이용자의 정보를 수집하고 있을 가능성이 높다는 것을 의미한다.
- Reflection: ICMP Unreachable, TCP RST 등으로 전송한 패킷이 되돌아온다면 공격을 위한 정보 수집을 의심할 수 있다.

4. 은닉마코프모델을 이용한 확률 예측 모델

본 논문에서는 네트워크 공격 확률을 예측하기 위해 은닉마코프모델을 적용한다. 직접 관측이 불가능한 네트워크 공격을 *은닉상태 집합*으로 놓는다. *은닉상태 집합*은 네 개의 원소로 구성되는 *공격집합* ( )으로 수식 (4)과 같이 표시한다.

$$\mathcal{N}=\{\text{DoS/DDoS, Worm, Unknown, Clean}\} \tag{4}$$

징후를 결정하기 위해 본 연구에서는 시계열 예측 기법을 이용하였다. 결정된 징후를 *관찰가능상태 집합* ( )으로 정의한다. 수식 (5)와 같다.

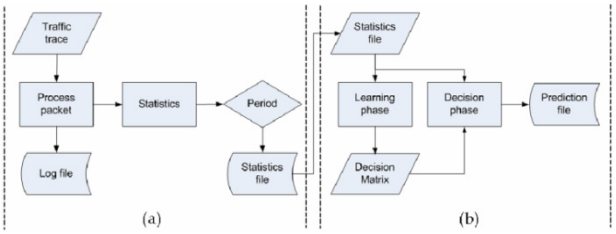
$$\mathcal{M}=\left\{\begin{matrix}\text{Port scanning, IP Scanning,}\\ \text{IP Spoofing, Packet Spoofing,}\\ \text{Sudden Increase, Reflection}\end{matrix}\right\} \tag{5}$$

공격집합( )과 *징후집합* ( )은 예측을 위한 정보로 각각 은닉마코프모델에서의 *상태전이 행렬* ( )와 *관찰확률 행렬* ( )에 대응된다. 정상 트래픽을 수집하여 *상태전이 행렬* ( )와 *관찰확률 행렬* ( ), 그리고 *초기값 벡터* ( )를 결정하였다. 결과적으로 이상징후를 탐지하기 위해서 은닉마코프모델을 적용할 수 있다. 은닉마코프모델을 적용하여 은닉상태 집합의 열을 예측하는 과정은 Viterbi 알고리즘을 사용한다[3].

5. 구현 및 데이터 설명

은닉마코프모델을 활용한 이상징후 탐지 프로그램은 Perl 로 구현하였으며, 구현에는 트래픽 수집과 관련된 Net::Pcap 모듈과 트래픽 분석에 사용되는

Net::Packet 모듈을 사용하였다. 프로그램은 3 장의 징후추출 과정과 4 장에서 설명한 징후판단 과정으로 구분하여 처리했다.



(그림 1) 징후추출 및 판단 과정 (a)징후추출, (b)징후판단

(그림 1)의 (a)징후추출 과정은 트레이스 파일을 읽어 들여 패킷 단위로 처리하여 패킷 로그파일을 생성한다. 패킷별로 각종 시계열 통계자료를 도출하고, 관측시간마다 통계자료를 출력한다. (그림 1)의 (b) 징후판단 과정은 징후추출 과정에서 추출된 시계열 통계자료를 입력값으로 하여, 각 단위시간마다 하나의 징후를 출력한다. 징후열은 관측열로 이용되며, 은닉마코프모델에서 이를 이용해 이상징후 경로를 계산할 수 있다.

제안한 기법의 검증을 위해 네트워크 공격이 포함된 트래픽 데이터 셋을 이용했다. 검증은 다섯 가지의 이상징후가 포함된 트래픽 데이터 셋을 이용했다. 이상 트래픽에 포함된 네트워크 공격은 Slammer worm, Witty worm, Codered worm, DoS(Denial of Service), 그리고 DDoS(Distributed DoS)이다. 이상 트래픽은 모두 사전에 Snort 와 Bro 등의 침입탐지 시스템으로 검사하여 네트워크 공격이 포함되어 있음을 확인하였다.

<표 1> 공격이 포함된 트래픽 데이터 셋 정보

데이터 셋	포함된 공격	수집시기(년)	포맷
SONY MAWI	Slammer Warm	2003	Pcap
CAIDA	Witty Worm	2001	Pcap
MIT Lincoln Lab	DoS, DDoS	1998, 1999, 2000	Pcap
NLANR	Codered Worm	2001	Tsh

<표 1>은 연구를 위해 수집한 트래픽 데이터 셋의 정보를 나타내고 있다. 트래픽 데이터 셋 들은 각각 다음과 같은 공격을 포함하고 있다. SONY MAWI 이 포함하고 있는 Slammer worm 은 1434 UDP 포트로 초당 20Mbps 이상의 패킷을 전송해 DoS 공격을 시도한다. CAIDA 의 Witty worm 은 취약점이 존재하는 방화벽 프로그램의 결함을 통해 UDP 포트 4000 번으로 전파되는 웜으로, 시스템의 데이터를 손상시키고 부팅을 진행할 수 없게 한다. NLANR 의 Codered worm 은 Microsoft 의 Internet Information Services 웹 서버를 공격하는 웜으로 감염될 경우, 웹사이트가 정상적인 웹 페이지를 띄우지 못하도록 DoS 공격을 시도한다.

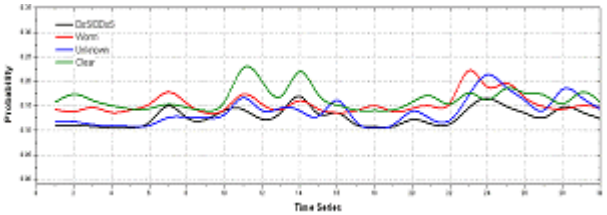
Codered worm 과 Witty worm 이 포함된 데이터 셋은 이상 트래픽만을 포함하고 있어 정확한 검증을 할 수 없었다. 이에 본 논문에서는 정상 트래픽을 배경 트래픽으로 재생시킨 후 Codered worm 과 Witty worm 이 포함된 데이터 셋을 3:1 의 비율로 재생시켜 결합 데이터 셋을 생성하였다.

MIT Lincoln Lab 에서 제공하는 1998 년 1999 년 데이터 셋에는 Smurf, mailbomb, ipsweep 등의 DoS 공격이 포함되어 있다. 그리고 2000 년 데이터 셋에는 Spoofing 된 IP 로 공격대상의 정상적인 서비스를 막는 DDoS 공격이 포함되어 있다.

본 논문에서는 앞서 설명한 징후추출 및 판단 프로그램에 공격을 포함한 데이터 셋을 입력해 이상징후 탐지과정을 수행했다. 다음 장에서는 도출한 탐지 결과를 분석한다.

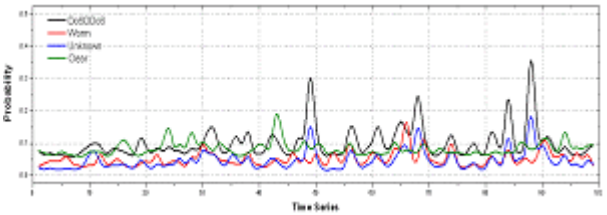
6. 결과분석

5 장에서 각 단위시간 별로 추출한 징후와 은닉마코프모델을 활용하여 각 이상징후(공격)확률 예측결과를 도출할 수 있었다.

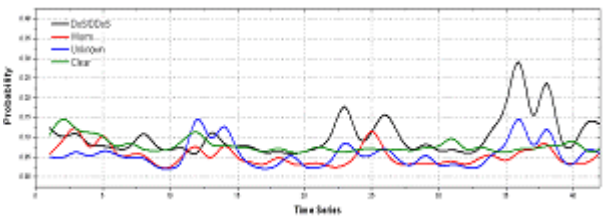


(그림 2) Normal traffic 의 공격확률 예측 결과

(그림 **2**)의 Normal traffic 공격확률은 상태 집합인 {DoS/DDoS, Worm, Unknown, Clear}에 대한 각각의 확률을 시계열로 보여준다. Clear 의 최대 확률 25.5%로 나타나고, 평균은 16.37%이다. Worm 의 발생 확률이 높아지는 시간은 국지적으로 분포한다. Worm 이 발생할 확률은 평균 15.3%이며, DoS/DDoS, Unknown 상태가 발생할 확률은 각각 평균 13.0%와 15.3%로 극히 적다. 전반적으로 Clear(녹색선)의 확률이 높은 상태를 보인다.

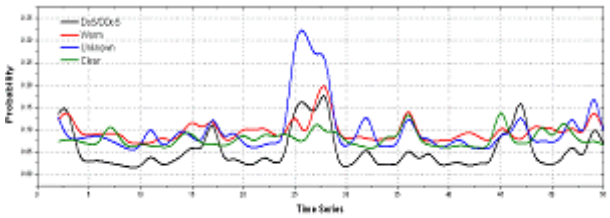


(그림 3) DoS 의 공격확률 예측 결과



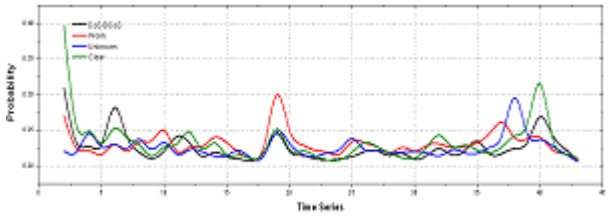
(그림 4) DDoS 의 공격확률 예측 결과

(그림 **3**)은 DoS, (그림 **4**)는 DDoS 의 공격확률 예측 결과를 보여준다. DoS 의 경우 전반적으로 DoS/DDoS 의 확률이 높고, DDoS 의 경우 공격이 출현한 시간에 DoS/DDoS 의 확률이 확연히 높은 것으로 나타났다.



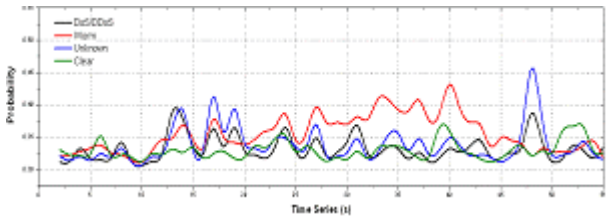
(그림 5) Codered worm 의 공격확률 예측 결과

(그림 **5**)의 Codered worm 공격확률 예측결과를 보면 Normal traffic 과 시계열 특징이 큰 차이를 보이지 않지만 25~30 초 사이의 Unknown 으로 최대 33.02%의 경고를 보인다. 이는 Unknown 으로 학습된 IP Scanning 관측열이 다수 발생하기 때문이다. 25~30 초 사이의 Worm 은 최대 22.49%의 경고를 보인다.



(그림 6) Slammer worm 의 공격확률 예측 결과

(그림 **6**)은 Slammer worm 의 공격확률 예측 결과이다. Worm 의 경우 20 초대에서 최대 33.9%의 경고를 보인다. 초기상태가 불안정한 특징이 있는데, 이는 Normal Traffic 에서 학습된 Reflection 관측열이 연속되어 발생하기 때문으로 분석된다. Worm(적색선)의 확률이 대다수이며, 20 초 정도에서 Worm 경고가 높아진다.



(그림 7) Witty worm 의 공격확률 예측 결과

(그림 7)을 보면, 트래픽 데이터에서 12~42 초 동안 Worm 의 활동 활발하며, 이때 최대 36.5%의 확률로 Worm 경고가 발생한다. 47 초에서는 Unknown 의 확률이 51.3%로 경고를 나타내는데, 이는 Unknown 으로 학습된 IP Scanning 관측열 때문이다.

결과를 보면 전체적인 확률저하 현상을 보이고 있으나 이는 이상징후 트래픽과 정상징후 트래픽의 결합으로 탐지의 난이도가 상승한 것과 다양한 데이터셋을 이용한 알고리즘 학습 부재로 인한 것이다. 하지만 이상징후 트래픽만 포함해 검증하지 않고 정상트래픽을 4~5 배 정도 결합하여 사용하였기 때문에, 보다 Real-trace 에 근접한 환경을 구축했다. 따라서 확률정확도가 저하된 것처럼 보는 것은 문제가 되지 않는다.

7. 결론

본 논문에서는 은닉마코프모델을 이용하여 네트워크 이상징후를 탐지하는 기법을 제안했으며, 제안한 기법의 검증을 위해 징후를 추출하고, 네트워크 공격을 정의하였다. 그리고 은닉마코프모델을 이용한 학습을 통해 네트워크 공격 탐지 확률을 높여 공격확률을 예측한다.

본 논문의 장점은 크게 두 가지로 요약할 수 있다. 첫째는 일기예보 방식처럼 확률적으로 이상징후를 제시하여 관리자가 공격을 판단하도록 하였다. 이는 시스템에 의한 오탐율과 미탐율 감소에 비해 획기적인 방식이라 할 수 있다. 둘째는 이상징후를 직접 관측하기 어렵다는 사실에 착안하여 은닉마코프모델을 적용한 것이다.

본 연구결과는 보안솔루션의 트래픽 데이터 분류나 징후 판단 방법을 정의하는데 이용될 수 있다. 또한 실시간 네트워크 상황을 파악하고 공격을 예측하는데 이용 가능하며, 신종 공격의 탐지와 분류에도 적용 가능하다.

참고문헌

[1] Ghosh, A. Schwartzbard, "A Study in Using Neural Networks for Anomaly and Misuse Detection", In Proceedings of *the 8th USENIX Security Symposium*, Washington, D.C., USA, August 1999.

[2] N. Ye, "A markov chain model of temporal behavior for anomaly detection", in Workshop on *Information Assurance and Security*, West Point, NY, June 2000.

[3] Yang Li, Li Guo, "An Efficient Network Anomaly Detection Scheme Based on TCM-KNN Algorithm and Data Reduction Mechanism", 8th Annual IEEE SMC Information Assurance Workshop (IAW 2007), United States Military Academy, West Point, New York, June 2007.

[4] Rabiner L. R., "A Tutorial on Hidden Markov Models and selected applications in speech recognition", *Proceedings of IEEE*, Vol. 77, No. 2, February 1989.

[5] Leek T. R., "Information Extraction using hidden Markov models", Master's thesis, UC San Diego, 1996.