Journal of The Korea Society of Computer and Information
Vol. 17 No. 5 , May 2012

2012-17-5-3-1

HTTP Outbound Traffic에 HMM을 적용한 웹 공격의 비정상 행위 탐지 기법

최병하*, 최승교**, 조경산***

Anomaly Detection Scheme of Web-based attacks by applying HMM to HTTP Outbound Traffic

*Byungha Choi, **Sung-kyo Choi, ***Kyungsan Cho

요 약

본 논문은 HTTP Outbound Traffic의 감시를 통해 다양한 웹 공격의 침입 경로에 대응하고, 학습 효율성을 높여 변종 또는 새로운 기법을 이용한 비정상 행위에 대한 오탐을 낮춘 기법을 제안한다. 제안 기법은 HMM(Hidden Markov Model)을 적용하여 HTML 문서속의 태그와 자바스크립트의 학습을 통한 정상 행위 모델을 생성한 후, HTTP Outbound Traffic속의 정보를 정상 행위 모델과 비교하여 웹 공격을 탐지한다. 실제 침입된 환경에서의 검증 분석을 통해, 제안기법이 웹 공격에 대해 0.0001%의 오탐율과 96%의 우수한 탐지능력을 보임을 제시한다.

▸ Keyword : 웹 공격, HTTP Outbound Traffic, 비정상 행위 탐지, HMM(은닉 마코브 모델), HTML 태그, 자바스크립트

Abstract

In this paper we propose an anomaly detection scheme to detect new attack paths or new attack methods without false positives by monitoring HTTP Outbound Traffic after efficient training. Our proposed scheme detects web-based attacks by comparing tags or javascripts of HTTP Outbound Traffic with normal behavioral models which apply HMM(Hidden Markov Model). Through the

• 제1저자 : 최병하 교신저자 : 조경산
• 접수일 : 2011. 12. 16, 심사일 : 2012. 01. 12, 게재확정일 : 2012. 02. 11.
* 단국대학교 일반대학원 컴퓨터학과(Dept. of Computer, Dankook University)
** 강원대학교 컴퓨터공학과 교수(Dept. of Computer Engineering, Kangwon University)
** 단국대학교 소프트웨어학과 교수(Dept. of Software Science, Dankook University)
※본 과제는 정보통신산업진흥원의 SW공학 요소기술 연구개발사업의 결과물임을 밝힙니다.

verification analysis under the real-attacked environment, we show that our scheme has superior detection capability of 0.0001% false positive and 96% detection rate.

▸ Keyword : Web-based Attacks, HTTP Outbound Traffic, Anomaly Detection, HMM(Hidden Markov Model), HTML tag, Javascript

I. 서 론

정당한 접근권한을 도용하거나 초과하여 정보시스템에 침입하는 해킹은 단순한 로컬 시스템의 패스워드를 추측하는 공격에서 네트워크를 이용한 원격 시스템의 공격을 거쳐 최근의 웹 공격(Web-based Attacks)으로 발전하고 있다[1].

웹 공격은 웹 응용 프로그램이나 이의 기능을 공격하는 것이며 응용 계층 공격의 70%가 웹 공격으로 분석된다[2]. 웹 공격은 웹 응용 프로그램과 연결되는 DB 서버, 네트워크 등의 취약점이 노출시키며 이와 접속한 PC들에게도 악성코드 유포 등의 피해로 나타난다.

이의 대응책으로 다양한 방화벽, 침입 탐지 시스템과 보안 응용 프로그램은 외부에서 Inbound Traffic으로 침입되는 웹 공격을 탐지하고 차단할 수 있으나, USB 메모리 악성코드와 관리자 친분을 이용한 공격 등의 다양한 우회 경로와 알려지지 않은 웹 공격 기법은 지속적으로 등장하고 있다[3]. 또한 성공한 웹 공격을 통해 웹 서버의 HTTP Outbound Traffic에서 웹 응용 프로그램의 악성행위 또는 웹 서비스를 제대로 못하는 비정상 행위가 나타나는 것으로 분석된다.

본 논문은 웹 서버의 HTTP Outbound Traffic의 HTML 태그와 자바스크립트를 자연어처리 및 영상 인식 분야에 주로 활용되었던 HMM(Hidden Markov Model)을 적용하고, 학습을 통해 정상 행위 모델을 생성하여, 이를 HTTP Outbound Traffic의 HTML 태그와 자바스크립트를 비교하여 비정상 행위를 신속하고 정확하게 탐지하는 기법을 제안한다.

다양한 우회 경로나 웹 공격 기법을 본 연구의 HTTP Outbound Traffic의 비정상 행위 탐지 기법으로 탐지할 수 있으며 HTTP Outbound Traffic의 HTML 태그와 자바스크립트를 검사하므로 각종 웹서버와 개발언어에 상관없이 탐지 가능하다. 구글의 안전 브라우징처럼 모든 웹 페이지를 탐지하는 것과 달리 본 기법은 탐지할 웹 서버의 웹 페이지만을 학습하여 탐지하므로, 기존 비정상 행위 탐지 기법의 단점인 오탐을 줄일 수 있으며 96% 탐지율의 정확성을 보인다. 본 논문은 2장에서 관련연구로 탐지하기 어려운 웹 공격과 이들을 탐지하는 기존 기법들을 분석한다. 3장은 2장의 분석을 기반으로 개선된 탐지 기법을 제안하고 4장에서 제안 기법을 검증하고, 5장의 결론으로 마무리 한다.

II. 관련 연구

본 장은 탐지하기 어려운 우회 경로를 이용한 웹 공격을 분석하고, 기존의 여러 탐지 기법을 분석한다. 또한, 기존의 침입 탐지 기법과 웹 공격의 문제점을 제시한다.

1. 다양한 우회 경로와 혼합된 웹 공격 기법

OWASP(the Open Web Application Project)와 SANS(SysAdmin, Audit, Network, Security) Institute는 각각 가장 위험한 웹의 취약점을 OWASP Top 10, 또는 SANS TOP 20 의 C1. Web Application으로 제시하였다. 이들 취약점을 TCP/IP의 네트워크 계층의 공격을 방어하는 방화벽과 다양한 계층의 공격을 탐지하는 침입 탐지 시스템, 그리고 HTTP만 집중적으로 공격을 방어하는 웹 방화벽으로 탐지하고 방어하지만, 그림 1과 같이 우회하는 악성코드가 삽입된 스팸메일, USB 메모리의 악성코드, 관리자의 친분, 웹 응용 프로그램의 오류를 이용하는 다양한 우회 경로의 공격은 탐지하기 힘들다[3].

또한 데이터베이스와 파일에 악성 행위 소스코드를 분산시

표 1. 오용 탐지와 비정상 행위 탐지 기법에 대한 비교
Table 1. Comparison of Misuse detection and Anomaly detection

유형	정의	탐지 방법	장점	단점	대표적 기법
오용 탐지	알려진 지식과 일치하는 경우 침입 간주	전문가에 의해 제작된 규칙으로 탐지	높은 탐지율	변형 공격, 새로운 공격 미탐지	전문가시스템, 시그니처분석, 페트리넷, 상태전이분석, 신경망, 유전알고리즘
비정상 행위 탐지	정상 행위을 분석 후 정상 행위와 불일치 시 침입으로 간주	사용자 패턴의 정상 행위 모델 과 입력 패턴의 비교로 정상 행위와 비정상 행위 탐지	변형 및 새로운 공격탐지	높은 오탐 (False Positive)	통계적기법, 전문가시스템, 신경망, 컴퓨터면역학, 기계학습, HMM

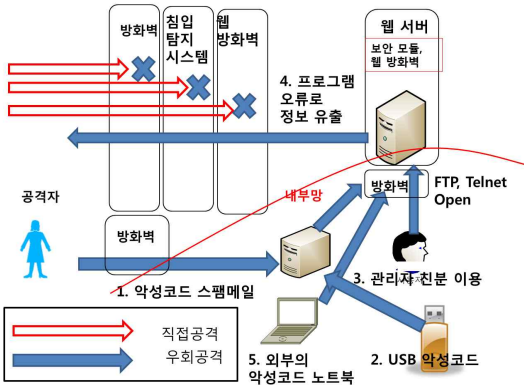

그림 1. 우회 공격기법
Fig. 1. Detoured Attacks

켜 실행시에 결합되어 나타나는 웹 공격이나 ARP(Address Resolution Protocol) Spoofing을 이용하여 웹서버 외부에서 변조된 패킷을 전달하는 기법[3,4] 등도 웹서버에서 이들을 탐지 못해 PC나 스마트폰 등으로 피해가 전이될 수 있다.

2. 웹 공격의 기존 탐지 기법

기존의 탐지 수단인 방화벽, 침입 탐지 시스템, 웹 방화벽에서 사용하는 기법으로는 표 1에 제시된 오용 탐지와 비정상 행위 탐지 기법이 있다[5,6].

또한 웹 서버를 중심으로 적용하는 시점에 따라 웹 공격 탐지 기법은 표 2처럼 분류가 가능하다. 즉, Inbound Traffic과 웹 서버의 침해 여부에 대한 많은 연구와 제품이 존재하지만 Outbound Traffic에 관련된 연구들은 대부분 정보 유출 등의 오용 탐지 기법들에 초점을 맞추고 있으며 변형 또는 새로운 공격에 대한 탐지 기법이 많지 않다. 따라서, 정상 행위의 분석을 기반으로 변형 또는 새로운 공격에 대한 탐지가 가능한 Outbound Traffic 기반의 비정상 행위 탐지 기법이 필요하다.

표 2. 적용시점에 따른 탐지 기법
Table 2. Detecting technologies assorted by the time of application

유형	탐지장비 또는 프로그램	탐지목적
Inbound Traffic의 탐지 및 차단	방화벽, 침입 탐지 시스템, 웹 방화벽	단순침입 시도[7] 유해트래픽, 웹 공격탐지
웹서버의 침해 여부의 탐지	로그관리 프로그램	웹서버와 기타 로그로 판단[8]
	웹 응용 프로그램 보안모듈(캐슬(1))취슬(2), 웰크린(3)	자세하고 세밀한 탐지(1) 웹 서버의 웹쉘 탐지(2),(3)
웹 침해 후 Outbound Traffic 탐지	웹 방화벽(4)	개인정보, 에러메시지, 백업파일 유출차단

(1,2) 한국 인터넷 진흥원 제작, (3) 안철수 보안관제 서비스
(4) WAPPLE(펜타시큐리티) 웹 방화벽

비정상 행위는 정상 행위를 충족시키지 못하는 패턴의 의미로 Anomalies, Outliers 등으로 불리운다[9]. 최근에는 표 1의 비정상 행위 탐지의 대표적 기법을 개선하거나, 여러 기법들을 혼합한 연구가 제안되고 있다. 두 부류를 구분하는 함수를 추정하여 분류하는 기계 학습기반의 SVM(Support Vector Machines)과 오용 탐지를 함께 이용하여 오탐을 최소화시키는 기법, 일반적인 거리 기반의 아웃라이어 클러스터 검출 기법을 대체하여 밀도 함수를 사용한 아웃라리어 클러스터 검출 기법, 유클리언 거리 기반의 K-means 클러스터링 알고리즘과 비균일 이진 분할에 대해 수행 속도가 빠른 LBG (Linde-Buzo-Gray) 알고리즘의 장단점을 혼합한 데이터마이닝 기법, 그리고 FCM(Fuzzy C-Means clustering)과 신경망 알고리즘을 혼합하는 기법 등이 제시되었다[5,10,11,12]. 그러나 이들은 다중 분류에 복잡한 계산과정과 다중의 데이터 차원에서 성능저하가 발생하고 네트워크 계층의 패킷(Packet) 특성이나 헤더만를 이용하므로 다양한 웹의 특성을 모델링하기에는 고려해야 할 사항이 많아 위의 기법들을 사용하기가 쉽지 않다. 웹의 특성을 고려할때 비선형적 관계를 간단하게 표현할 수 있는 신경망 알고리즘을 기반으로 탐지할 수 있으나, 일반적으로 계산량이 많고 입력과 출력간의 관계를 알 수 없어 어떤 결과의 인과관계 파악이 힘들다는 단점이 있다[13].

이러한 단점을 극복하고 웹의 입력 파라미터를 HMM 기반의 정상 행위 모델을 생성하여 실시간으로 탐지하는 기법이 제안되었다[14]. 이는 여러개의 HMM 모델을 다중으로 이용하여 오탐을 줄이고 정확성을 향상시켜, 웹 공격의 탐지 기법에 적용할 수 있음을 보였다.

3. HMM(Hidden Markov Model)

HMM은 숨겨진(Hidden) 모수를 결정하기 위해 관찰이 가능한 기호로 모델링하는 이중의 확률적 과정이다[14]. HMM은 {S, V, Π, A, B}의 매개변수를 가지며 이들의 의미는 표 3과 다음과 같다[15].

표 3. HMM을 구성하는 5가지 매개변수 (S, V, Π, A, B)
table 3. HMM as a 5-tuple (S, V, Π, A, B)

변수	의미
S	N개의 은닉상태 값의 집합 $= \{S_1, S_2, \cdots, S_n\}$
V	M개의 관측가능한 관찰 기호의 집합 $= \{v_1, v_2, \cdots v_m\}$
Π	각 상태의 초기 확률 집합 $= \{\pi_1, \pi_2, \cdots \pi_i\}$
A	상태 천이 확률의 집합 = {aij} aij : 상태 Si 에서 Sj 로 천이할 확률 $= a_{ij} = P(q_{t+1}=S_j \mid q_t = S_i),\ 1 \le i, j \le N$
B	상태에 대한 출구 확률의 집합 = {b(vk)} b : 상태(k) Si에서 관찰기호 v_k를 출력할 확률

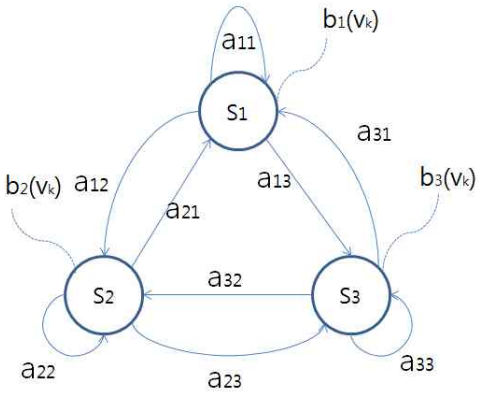

그림 2. Hidden Markov Model
Fig. 2. Hidden Markov Model

이를 이용한 예로 3가지의 상태를 가진 HMM은 그림 2와 같이 도식화가 가능하며 이는 방향성과 확률값의 가중치를 가진 간선과 노드의 그래프 형태가 된다.

HMM은 5가지 구성요소로 특정 모델 λ를 규정할 수 있으며 Π, A, B를 이용하여 λ=(Π,A,B)으로 식으로 나타낼 수 있다.

또한 주어진 목적에 따라 3가지 문제와 그에 따른 해결책이 있는데, 이들은 확률 추정, 최적 상태 순서, 매개 변수 추정이다[15].

확률 추정은 주어진 T 개의 관측열 O=O1O2O3...Ot이 모델 λ에서의 발생할 관측 확률을 의미한다.

최적 상태 순서는 주어진 T개의 관측열 O=O1O2O3...Ot이 이 모델 λ를 만족하는 최적의 상태열 Q=q1q2...qt 를 구하는 것이다.

매개 변수 추정은 모든 관측열에 대해 우도(likewood)를 최대화 시키는 것으로 매개변수 Π, A, B의 최적화된 값을 찾는 것이다. 이는 Baum-Welch 알고리즘으로 최적의 매개변수 값을 찾을 수 있다.

HMM의 장점은 신경망과 달리 관찰기호의 순서가 의미있는 값이 될 수 있으며 출력을 확률로 취급할 수 있다는 것이다. 그러므로 예측하기 힘든 상황을 모델링하여 비정상 행위를 확률로 탐지할 수 있는 장점이 있으며, 단점으로는 학습과정에 시간이 많이 소요되는 단점이 있다[16].

III. 개선된 공격 탐지 기법 제안

본 장은 기존의 탐지 기법의 분석과 HMM을 기반으로 다양한 우회 경로로 침입하는 웹 공격, 그리고 DB에 공격된 태그나 자바스크립트가 삽입되어 웹 페이지 실행시 나타나는 웹 공격, 그리고 신종 웹 공격을 탐지할 수 있는 기법을 제안한다.

1. 제안 기법의 탐지 과정

제안 탐지 기법은 HTTP Outbound Traffic을 이용하여 패킷속에 들어있는 HTML 태그가 정상 또는 비정상인지를 HMM 모델로 판단하는 것으로 그림 3과 같은 6단계의 탐지 과정을 거친다.

① 학습해야할 대상인 HTML 문서를 수집한다.
② 제안 기법에서 사용될 HMM 모델을 초기화하고 수집된 HTML 문서의 태그와 자바스크립트를 HMM으로 적용하기 위한 관측열로 변환한다.
③ 관측열로 변환된 태그와 자바스크립트를 Baum-Welch 알고리즘으로 학습시킨다.
④ 학습의 결과로 정상 행위 모델을 생성한다.
⑤ 생성된 정상 행위 모델을 침입 탐지 시스템에서 사용할 수 있도록 파일로 저장한다.
⑥ 저장된 정상 모델을 침입 탐지 시스템에서 읽어서, HTTP Outbound Traffic의 패킷들을 재조립하여 이들 패킷속의 태그와 자바스크립트의 관측열이 정상 행위 모델과 비교하여 웹 공격인지 판단한다.

2. HMM을 이용한 자바스크립트 모델링

자바스크립트는 JScript, Javascript, ECMA Script 등의 종류가 있으며, 공급자의 문법과 객체 등은 대동소이한 형식을 갖는다 이들은 함수와 변수 등의 조합과 순서에 따라 완전히 다른 행위를 할 수 있는 예측하기 힘든 형태로 HTTP Outbound Traffic에 존재한다.

그림 3의 과정에서 수집된 HTML 문서에서 자바스크립트를 HMM의 초기 모델로 적용하기 위해 그림 4와 표 4처럼 설정한다. 이는 5개의 숨겨진 상태(“연산자”, “예약어”, “객체와 함수명”, “변수의 할당값 또는 함수의 인자 및 반환 값”, “기타”)를 정의하고 주요하게 관측될 관측열의 확률값을 초기화하여 도식화 하였다.

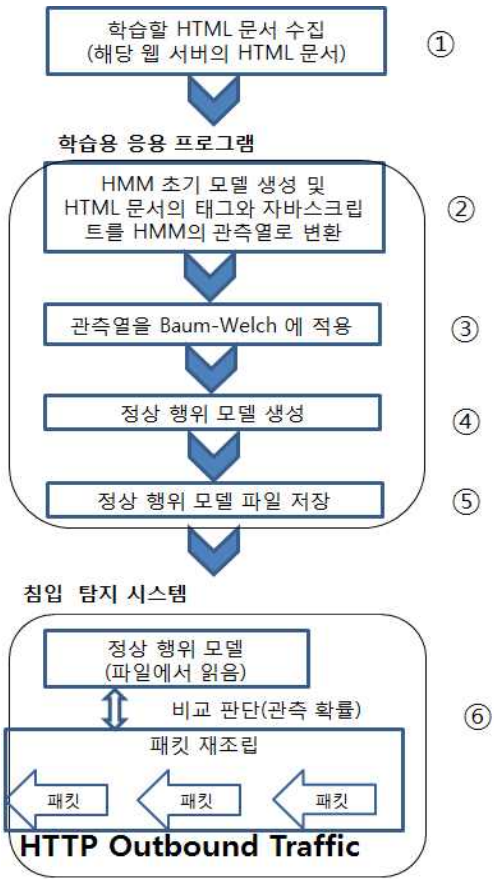

그림 4. 제안 기법 구현 과정
Fig. 3. Implementation Process of the Proposed Scheme

표 4. HMM의 5가지 매개변수 (S, V, Π, A, B)
table 4. 5-tuple of HMM (S, V, Π, A, B)

변수	내용
S	= {S_1 , S_2 , S_3 , S_4 , S_5 }
V	= {연산자 44개, 키워드 28개, 객체와 함수명 500개, "변수의 할당값 또는 함수의 인자 및 반환값" 30개, 기타 88개 }=총 690 개)
Π	{π1, π2, π3. .π4. π5 }={0.2, 0.2, 0.2, 0.2, 0.2}
A	S_1 에서 {aij} : ={1/4, 1/6, 1/6, 1/4, 1/6 } S_2 에서 {aij} : ={1/4, 1/4, 1/6, 1/7, 1/5 } S_3 에서 {aij} : ={1/5, 1/5, 1/5, 1/5, 1/5 } S_4 에서 {aij} : ={1/5, 1/5, 1/5, 1/5, 1/5 } S_5 에서 {aij} : ={1/5, 1/5, 1/5, 1/5, 1/5 } I,j는 1부터 5까지의 자연수
B	S_1에서는 연산자가 관측될 확률 80% S_2에서는 키워드가 관측될 확률 80% S_3에서는 기타가 관측될 확률이 80% S_4에서는 "변수할당값" 등이 관측될 확률이 80% S_5에서는 "객체및함수명" 등이 관측될 확률이 80%

그림 5와 같이 자바스크립트를 HMM의 관측열로 생성할 수 있다. 즉 왼쪽 상단의 자바스크립트를 파서를 통하여 AST(Abstract Syntax Tree)로 생성하고, 이를 다시 AST의 깊이(Depth) 1을 기점으로 삼아 그림 5의 하단 좌측 같은 관측열을 생성한다.

이들 관측열에 HMM의 확률 추정의 알고리즘을 적용하여 각 관측열의 확률 값을 구할 수 있다.

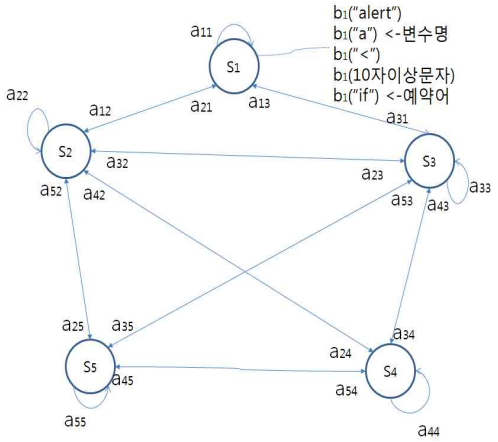

그림 5 자바스크립트의 HMM 모델
Fig. 4. HMM Model of Javascript

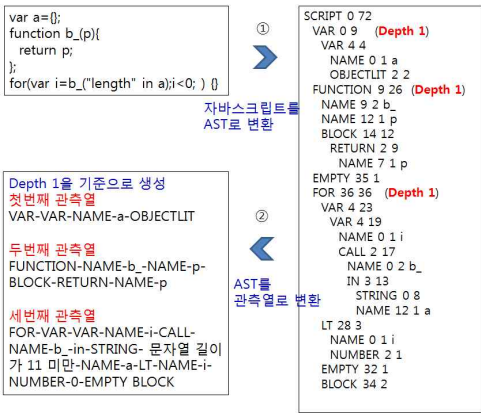

그림 6. 관측열 생성 과정
Fig. 5. process of generating a sequence

3. HTML 태그 모델링

HTML 태그는 태그명, 태그 속성, 태그 내용 등의 상태를 3가지의 은닉된 상태로 하고 태그명 91개, 태그 속성 119개, 태그 내용 24 개의 총 234개의 매개변수 V 가 되도록 하고, 초기화 값(π)은 태그명이 1 그리고 나머지는 0이 되도록 초

기화 한다. 또한 관측열 생성은 HTML 파서를 이용하여 <img src=0>이라는 태그를 img-src-0 으로 생성한다.

4. 학습을 통한 정상 행위 모델 구축

학습을 위한 해당 서버의 HTML 문서를 내려받아 앞장의 3.1과 3.2절의 모델링에 의해 관측열을 생성하여 Baum-Welch 알고리즘에 적용한다. 이것은 앞장에서 설명된 매개 변수 추정이다. 이 알고리즘에 적용하여 출력되는 결과는 HMM의 매개변수 Π, A, B를 최적화 시킨 값으로 변경된다. 이는 해당 웹 서버의 HTML 문서의 학습으로 최적화된 정상 행위 모델을 생성할 수 있다.

5. 침입 탐지

HTTP Outbound Traffic의 패킷을 실시간으로 재조립하여 태그와 자바스크립트를 관측열로 생성한다. 그리고 학습된 HMM의 정상 행위 모델에 대해 그 관측열을 HMM의 확률 추정으로 확률값을 구한다. 즉 정상 행위 모델에 대해 패킷속에 있는 자바스크립트나 태그의 관측열을 확률 추정하였을 때 학습된 패턴은 0 이상이 되며, 학습되지 않은 비정상은 0 이 된다. 따라서 확률값이 0 일 때 웹 공격으로 침해된 상태라고 판단한다.

IV. 제안 탐지 기법의 검증 분석

본 장에서는 실제 사이트에서 수집된 HTML문서와 국내의 보안 기관에서 제공된 웹 공격된 페이지를 이용하여 제안 기법을 검증한다.

1. 검증 환경

표 5. 검증 시스템의 사양
table 5. Details of Verification System

서버	항목	내용
웹 서버	DB	MS - SQL 2000
	웹서버	IIS 5.0
	웹프로그램 언어	ASP
	실행환경(가상머신)	MS Virtual PC
	운영체제	윈도우 2000 서버
	비고	IIS와 같은 컴퓨터에 설치
침입 탐지 시스템	Packet capture lib	Jpcap 0.7
	HTML 파서	Jericho HTML Parser 3.2
	자바스크립트 파서	Rhino 1.7 R3
	HMM 라이브러리	JaHMM
	JDK	Oracle JDK 1.6
	실행환경(가상머신)	MS Virtual Box
	운영체제	윈도우 2000 서버

제안 기법의 검증을 위해 표 5와 그림 6처럼 보안에 취약한 윈도우 2000 서버 기반의 웹 서버와 침입 탐지 시스템을 MS Virtual PC의 가상머신으로 구축하였다. 이들 두 시스템으로 HTTP Outbound Traffic의 웹 공격을 탐지하는 환경을 구축하는데,웹 서버는 실제 사용되었던 쇼핑몰 사이트의 데이터베이스와 웹 응용 프로그램을 동일하게 구축하였다. 침입 탐지 시스템은 Java 기반의 패킷 캡쳐 라이브러리인 JPCap을 이용하여 HTTP Outbound Traffic의 패킷을 실시간으로 재조립하여 학습된 HMM 모델에 적용시켜 탐지하는 응용 프로그램을 개발하여 설치하였다.

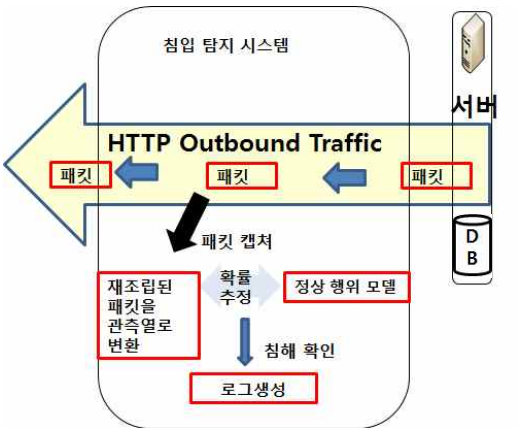

그림 6. 검증 시스템 구성도
Fig 6. Architecture of Verification System

2. False Positive 오탐 검증

초기의 웹 서버는 공격되지 않은 상태이며, 검증을 위해 웹 브라우저로 무작위의 1691 페이지를 요청하였다. 본 연구에서 구현된 HTTP Outbound Traffic의 침입 탐지 시스템으로 이들 페이지 요청에 대한 응답을 검색하여 96245개의 관측열 중에서 잘못 판단된 관측열의 비율이 표6과 같이 0.0001%의 오탐율을 가지며 13의 오탐된 관측열이 각각 다른 13개의 페이지에 나타난 탐지결과를 얻었다.

표 6. False Positive 오탐율
Table 6. Detection rate of False Postive

페이지수	관측열	오탐된 관측열	오탐율	오탐 페이지수
1691	96245	13	0.0001%	13개 페이지

오탐율의 원인은 웹서버의 다양한 오류페이지 중 일부와 학습되지 못한 특수한 HTML 문서에 기인한다.

3. 웹 공격 탐지

국내 보안기관에서 제공받은 웹 공격된 페이지 99개와 실제 운영되는 사이트가 공격되어 DB에 포함된 SQL Injection의 웹페이지 1개를 웹서버와 DB에 삽입하고 그 웹 페이지를 PC에서 요청했을 때, 제안 탐지 시스템이 웹 공격으로 판단하는지를 실험한 결과는 표 7과 같다. 공격된 웹 페이지는 자바스크립트를 파싱 못하는 비율이 3%가 되며 1%의 미탐이 존재하여 96%의 탐지율을 가진다. 미탐된 페이지는 정상적인 자바스크립트와 구분되지 않는 형태로 다른 사이트로 이동시키는 자바스크립트다. 이는 정상적인 자바스크립트와 차이가 없으므로 탐지하지 못하였다. 한편 정상적인 HTML 문서에서 파싱 오류가 발생하지 않으므로 자바스크립트 파싱 오류를 제외하면 97개중 96개를 탐지하므로 99%에 이른다.

표 6과 7에 제시된 결과는 Inbound Traffic를 HMM에 적용하여 탐지하는 기법인 참고문헌[14]의 96 %의 탐지율과 오탐율 1% 이하로 비슷한 성능을 나타낸다. 또한 파싱 오류를 제외한다면 3% 더 향상된 성능을 보여준다.

표 7. 탐지율
Table 7. Detection rate

총 공격 페이지	파싱 오류 페이지	미탐 페이지	탐지율
100	3	1	96%

V. 결 론

본 연구에서는 해당 웹 서버의 HTML 문서로 그 서버에 최적화된 정상적인 HTML과 자바스크립트의 HMM 모델을 생성하여 HTTP Outbound Traffic을 이용하여 웹 공격을 탐지하는 효율적인 비정상 행위 탐지 기법을 제안하였다.

본 제안 기법은 HTTP Outbound Traffic을 이용하므로 다양한 경로의 공격과 DB 속에 공격된 태그나 자바스크립트가 포함되어 실행시 나타나는 혼합된 웹 공격도 탐지할 수 있다. 또한 알려진 패턴 등을 이용하는 오용 탐지와 달리 비정상 행위 기반의 탐지 기법이므로 변종이나 알지 못하는 형태의 기법의 공격도 탐지 가능하다.

본 기법은 0.0001%의 오탐율과 96%의 탐지율을 보여주며 이는 Inbound Traffic을 이용하는 참고문헌[14]의 96 %의 탐지율과 오탐율 1% 이하인 것과 비교하면 파싱 오류로 인한 미탐을 포함할 경우 유사한 성능을 제공하며, 파싱오류를 제외한다면 탐지율이 향상된 성능을 보여준다.

그러나 본 연구의 한계는 일반적으로 서버측의 Inbound Traffic 보다 Outbound Traffic의 양이 많기 때문에 트래픽이 폭주할 경우에 대한 문제와 SSL(Secure Socket Layer) 등의 암호화된 트래픽의 문제를 고려하지 않아 이들로 구성된 트래픽의 탐지에 제한이 있다. 또한 표 7의 미탐은 비정상 행위 탐지 기법보다 오용 탐지 기법으로 해결하는 방법이 더 쉽게 접근할 수 있다.

향후 연구로는 성능 좋은 파서를 이용하여 미탐되는 부분을 탐지하기 위해 오용탐지 기법과 본 연구의 비정상 행위 탐지 기법을 혼용하여 다양한 웹 환경에서 제약없이 탐지 할 수 있는 기법을 제안한다.

참고문헌

[1] Wang Qinquan, Piao ZaiLin, “Research on Network Attack and Detection Methods,” Procs. of Education Technology and Computer Science (ETCS), pp. 630-633, Mar. 2010.

[2] Justin Crist, “Web base Attacks,” SANS Instutute, Jan. 2008.

[3] ByungHa Choi, Kyungsan Cho, "An Efficient Detection Scheme of Web-based Attacks through monitoring HTTP Outbound Traffics," Journal of The Korea Society of Computer and Information, Vol. 16, No 1, pp. 125-132, Jan. 2010.

[4] Ahn LAB, http://core.ahnlab.com/261

[5] Gill-Han Kim, Hyung-Woo Lee, "False Alarm Minimization Technology using SVM in Intrusion Prevention System," Journal of Korea Society for Internet Infromation, Vol .7, No. 3, pp. 119-132, Jun. 2006.

[6] Dong-Jin Shin, Hae-Sool Yang, "Design and Implementation of an Intrusion Detection System based on Outflow Traffic Analysis," The Journal of the Korea Contents Association, Vol. 9, No. 4,pp. 131-141, Apr. 2009.

[7] Karen Scarfone, Paul Hoffman, “Guidelines on Firewalls and Firewall Policy,” National Institute of Standards and Technology, Sep. 2009.

[8] Ha-Na Yoon, Taesu Kim and Hyung-Woo Lee, "Design and Implementation of Web Attack Detection System Based on Audit Data," Procs. of Korea Society for Internet Infromation, Vol. 20, pp. 295-298, Dec. 2009.

[9] Varun chandola, Arindam Banerjee and Vipin Kuma "Anomaly Detection : A Survey," ACM Computing Surveys (CSUR), Vol. 41 Issue 3, Jul. 2009.

[10] Jae-young Chang, Han-joon Kim and Jongmyoung Park, "An Outlier Cluster Detection Technique for Real-time Network Intrusion Detection Systems," Journal of Korean Society for Internet Infromation, Vol. 8, No. 6, pp. 43-53 , Dec. 2007.

[11] Seongchul Park, Juntae Kim, "Improvement of Network Intrusion Detection Rate by Using LBG Algorithm Based Data Mining," Journal of Intelligence and Information Systems Vol. 15, No. 4, pp. 23-35, Dec. 2009.

[12] Muna Mhammad T. Jawhar, Monica Mehrotra, "Design Network Intrusion Detection System using hybrid Fuzzy-Neural Network," International Journal of Computer Science and Security, Vol. 4, Issue 3, Jul. 2010.

[13] Sang-Jun Han, Sung-Bae Cho, "Intrusion Detection Using Multiple Measure Modeling and Integration," Procs. of Korean Information Science Society, Vol. 29, No. 2, pp. 523-525, Dec. 2002.

[14] Igino Corona, Davide Ariu and Giorgio Giacinto "HMM-Web: a framework for the detection of attacks against Web applications," Procs. of ICC '09. IEEE International Conference, pp. 1-6, Jun. 2009.

[15] Hyung-deuck Moon, Ja-young Koo "Recognition of Conducting Motion using HMM," Journal of The Korea Society of Computer and Information, Vol. 9, No 1, pp. 25-30, Jan. 2004.

[16] Sang-Jun Han, Sung-Bae Cho, "Effective Intrusion Detection using Evolutionary Neural Networks," Journal of KIISE : Information Networking Vol. 32, No. 3, pp. 279-432, Jun. 2005.

저 자 소 개

최 병 하

2010~현재 단국대학교 컴퓨터학과(박사과정)

관심분야: 네트워크 보안

Email : notanything@hanmail.net

최 승 교

1982 단국대학교 전기공학과(학사)

1992 단국대학교 대학원 전산통계학과(석사)

2001 단국대학교 대학원 전산통계학과(박사)

1994~현재 삼척/강원대학교 컴퓨터공학과 교수

관심분야 컴퓨터구조, 성능평가, 시뮬레이션

Email : skchoi@kangwon.ac.kr

조 경 산

1979: 서울대학교 전자공학과(학사)

1981: 한국과학원 전기전자공학과(공학석사)

1988: 텍사스 대학교(오스틴) 전기전산공학과(Ph.D.)

1988~1990: 삼성전자 컴퓨터부문 책임연구원, 실장

1990~현재: 단국대학교 컴퓨터학부 교수

관심분야: 네트워크시스템 및 이동통신 보안, 컴퓨터시스템

Email : kscho@dankook.ac.kr