정상 스크립트 패턴에 기반한 비정상 스크립트 탐지

백성규[0] 위규범
아주대학교 정보통신전문대학원
{shalom34[0], kbwee}@ajou.ac.kr

Abnormal Scripts Detection based on Normal Scripts Patterns

Sung-Kye Baek[0] Kyu-Bum Wee
Graduate School of Information and Communication, Ajou University

요 약

본 논문은 악성 스크립트를 탐지하는 새로운 방법을 제안한다. 정보검색 기법을 이용하여 정상 스크립트들을 기능별로 구분하여 정상 행위를 정의함으로써, 정상 행위에서 벗어나는 경우에 악성이라고 판정한다. 소스 기반의 빠른 검색이 가능하며, 실시간 모니터링을 통한 비정상 스크립트의 탐지가 가능하다. 또한 새로운 악성 스크립트가 생성되는 경우에도 탐지가 가능하다는 장점을 가지고 있다.

1. 서 론

악성 스크립트는 스크립트 언어로 제작된 언어를 말하며, 1994년 12월에 최초로 발견되었다. 악성 스크립트를 구현하기 위해 가장 많이 사용되는 스크립트 언어는 VBS(Visual Basic Script)이다. VBS는 독자적으로 실행이 가능하여 자기복제, 레지스트리 변경 그리고 파일 수정등의 기능들을 가지고 있어서 많은 피해를 입히고 있는 실정이다.

그러나 현재 나와 있는 기법은 대부분 오용탐지(misuse detection)에 관련된 방법들로써, 새로운 악성 스크립트를 탐지할 수 없다.

따라서 본 논문에서는 새로운 악성 스크립트를 탐지할 수 있는 방법을 제안하고자 한다.

제 2장에서는 기존 대응 기법에 대해서 설명한다. 제 3장에서는 시스템 개요와 구조에 대해서 설명한다. 제 4장에서는 정보 검색(Information Retrieval) 기법과 침입 탐지 시스템에 K-Nearest Neighbor의 적용에 대해서 설명한다. 제 5장에서는 실험 결과 및 향후 연구 방향에 대해서 설명한다.

2. 기존의 대응 기법

2.1 패턴 비교 방법

안티 바이러스 프로그램들이 바이러스를 탐지하기 위해 이용하는 방법으로 악성 스크립트의 시그니처(Signature)를 이용한 대응 기법이다.

2.2 면역 시스템을 이용한 방법

인체 면역 원리를 응용한 디지털 면역 시스템 대응 기법이다.

2.3 신경망

악성 스크립트의 시그니처(Signature)를 신경망을 이용하여 새로운 시그니처(Signature) 패턴을 생성하여 패턴 비교하는 대응 기법이다.

2.4 기타 악성 코드 탐지 기법

2.4.1 메일에 첨부되는 파일명을 분석하는 방법

메일에 첨부되는 파일명을 메일서버에서 분석하여 악성여부를 판정하는 방법이다. 대부분의 악성 코드의 파일 이름이 복잡한 형태로 되어 있는 데에 착안한 방법이다.

2.4.2 네트워크 흐름을 통해 분석하는 방법

악성 스크립트가 네트워크에서 전파될 때 네트워크에 사용량이 악성 스크립트가 없는 네트워크의 경우보다 증가 된다. 이러한 네트워크의 흐름을 관찰하여 악성 호스트, 악성 네트워크를 판정하는 기법이다[7].

3. 시스템의 개요와 구조

3.1 시스템의 개요

3.1.1 실시간 스크립트 호스트의 모니터링을 통해 스크립트가 실행되기 전에 해당 스크립트의 악성 여부를 판단하여 실행을 차단하는 것을 목적으로 한다.

3.1.2 정상 스크립트와 대상 스크립트의 중요도를 계산하기 전에 주석 제거와 스크립트들의 기능별로 나누는 전처리를 하게 된다. 전처리 후에는 각 function들이 하나의 스크립트와 같이 처리한다.

3.1.3 악성 여부의 판단은 보유하고 있는 정상 스크립트의 샘플과 탐지 대상이 되는 스크립트의 유사 정도를 근거로 한다. 단 여기서 정상 스크립트와 악성 스크립트를 그대로 사용하는 것이 아니라, 기능별로 나누어서 악성 여부를 결정하게 된다.

3.1.4 탐지 대상은 암호화된 스크립트를 제외한 스크립트를 실험 대상으로 한다.

3.2 시스템의 구조

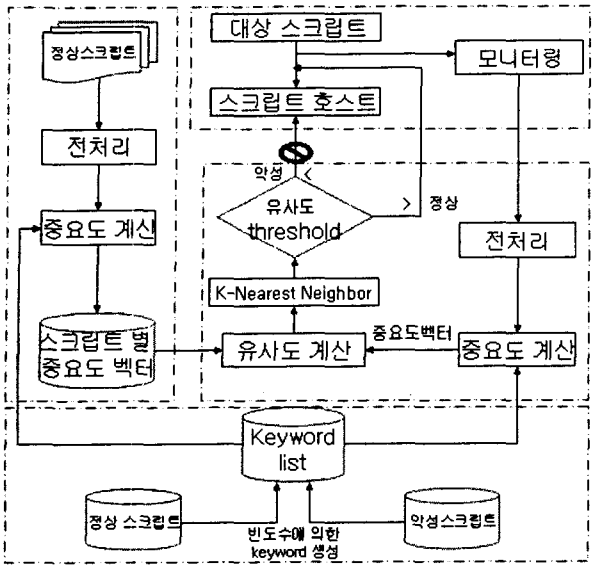


그림1 시스템의 구조

본 시스템은 그 기능에 근거하여 네 부분으로 구성된다.

3.2.1 아랫부분은 중요도 계산을 하기 위해 keyword list를 생성하는 과정이다. 수집되어 있는 정상 스크립트와 악성 스크립트들에서 정상 스크립트에는 많이 나오고 악성 스크립트에는 적게 나오는 keyword들과, 악성 스크립트에 많이 나오며 정상 스크립트에는 적게 나오는 keyword들을 keyword list로 정의하게 된다. 정상 스크립트와 악성 스크립트들을 좀더 정확히 구분하기 위함이다. 본 논문에서는 100개의 keyword를 선택하였다.

3.2.2. 왼쪽 윗부분은 미리 수집된 정상 스크립트로부터 중요도 벡터를 계산하는 부분이다. 정상 스크립트의 불필요한 주석을 제거하고, function별로 나누어 별도의 스크립트로 취급하는 전처리 과정을 거치게 된다. 그 후에 키워드 list에 있는 키워드별로 중요도를 계산하여 저장한다.

3.2.3 오른쪽 윗부분은 스크립트 호스트를 모니터하는 부분으로 대상 스크립트가 스크립트 호스트에 의해 실행되기 전에 가로채어 오른쪽 아랫부분으로 전달하게 되고, 유사도를 계산하여 정상 판단 여부에 따라 실행 여부를 결정하게 된다.

3.2.4 오른쪽 중간부분은 대상 스크립트의 정상 여부를 판단하는 부분으로 불필요한 주석을 제거하고, function별로 나누어 별도의 스크립트로 취급하는 전처리 과정을 거치게 된다. 대상 스크립트의 function 별로 각 키워드의 중요도를 구하고, 미리 저장된 정상 스크립트별 중요도 벡터를 이용하여 유사도를 계산한다.

정상 스크립트인 경우 대상 스크립트가 실행되도록 스크립트 호스트에게 스크립트를 넘겨주고, 악성인 경우 사용자에게 보고하고 스크립트의 실행을 막는다.

4. 정보 검색 기법의 침입 탐지 시스템 적용

4.1 정보 검색 기법

4.1.1 중요도

$$weight\ (i,j) = \begin{cases} (tf_{i,j}) \log(\frac{N}{idf_i}) & \text{if } tf_{i,j} \geq 1 \\ 0 & \text{if } tf_{i,j} = 0 \end{cases}$$

수식1 중요도 계산식

i : 단어의 인덱스
j : 문서의 인덱스
N: 전체 문서의 개수
$tf_{i,j}$: i번째 단어가 j번째 문서에 나타난 횟수
idf_i: i번째 단어가 N개의 문서 중에 나타난 문서의 개수

키워드 리스트의 키워드를 이용하여 정상 스크립트와 호스트에서 실행되어지기 전 대상 스크립트의 중요도를 계산하게 된다.

4.1.2 유사도

$$\cos(q,d) = \frac{\sum_{i=1}^{n} q_i d_i}{\sqrt{\sum_{i=1}^{n} q_i^2}\sqrt{\sum_{i=1}^{n} d_i^2}}$$

수식2 유사도

q : 질의어
d : 문서
n : 매칭 단어 수
q_i : 질의어에서 매칭된 단어의 중요도
d_i : 소스 문서에서 매칭된 단어의 중요도

정상 스크립트의 중요도를 기반으로한 값과 감시된 스크립트의 유사도를 위의 식을 이용하여 구한다[5,6].

4.2 K-Nearest Neighbor 기법의 적용

```
build the training normal data set D;
for each script X in the test data do
    for each script Dj in training data do
        calculate cos(X, Dj);
        if cos(X, Dj) equals 1.0 then
            X is normal ; exit;
        find k biggest scores of cos(X, D);
        calculate sim_avg for k-nearest neighbors;
        if sim_avg is greater than threshold then
            X is normal;
        else
            X is abnormal;
```

그림2 Pseudo code for detection of abnormal Scripts

감시대상 스크립트와 정상 스크립트간의 유사도를 모두 구한 후에 유사도 값이 큰 순서대로 K개의 유사도 값의 평균값을 기준값으로 사용한다. 그 평균값이 threshold 값 보다 크면 normal로 작으면 abnormal로 판정하게 된다[3].

전처리 후에 실험을 하였으며, 하나의 스크립트에서 나누어진 여러 스크립트들 중에서 유사도 값이 가장 적은 것을 기준으로 하였다. 기능별로 분리된 스크립트중 하나라도 비정상이라고 판단이 되면 비정상으로 분류 하였다.

5. 실험 결과 및 향후 연구

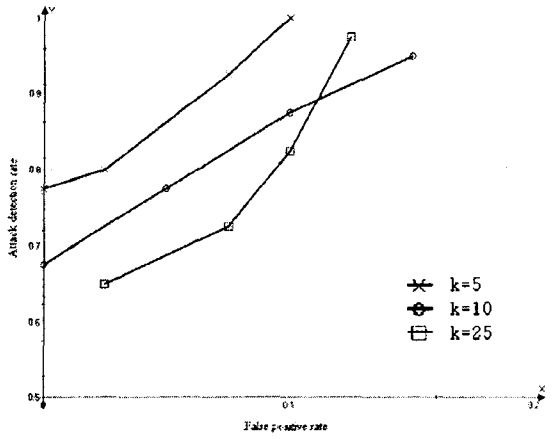
5.1 데이터

종 류		스크립트의 개수	전처리 후 스크립트의 개수
profiling	정상 VBS	352	743
test	비정상 VBS	40	52
	정상 VBS	40	82

표1 dataset

dataset의 종류는 크게 profiling과 test로 나뉘며, profiling에 사용된 것과 test에 사용된 것은 서로 다른 것들이다. 정상과 비정상을 나누는 기준은 기존의 백신 프로그램을 사용하였다[1,2].

5.2 실험결과



[그림3] Performance of kNN classifier method expressed in ROC curves

5.3 결과 분석

위의 실험에서는 threshold값을 네 개를 이용하여서 실험하였고,

그림3의 결과로 볼 때, 정상 스크립트의 패턴을 이용하여 비정상 스크립트 탐지는 충분히 가능하다고 볼 수 있고, K값의 선정이 중요한 요소임을 알 수 있다.

그러나 K-Nearest Neighbor 알고리즘을 적용하였을 경우에 K 값이 증가함에 따라 , 오히려 결과가 나빠지는 것을 볼 수 있는데, 이것은 정상 스크립트에 대해서 실험에 사용된 정상 스크립트가 충분하지 않다는 것을 보여주고 있다. 따라서 이 방법을 실제 침입 탐지 시스템에 적용하기 위해서는 훨씬 더 많은 양은 정상 스크립트를 가지고 정상 행위를 정의해야 될 것이다.

5.2 향후 연구

정상 스크립트를 k-means 방법을 통해 clustering 한다면 키워드를 기반으로 한 유사한 스크립트들이 합치지게 되므로 False positive는 감소하고 detection rate는 향상 될 수 있을 것이다.

6. 참고문헌

[1] 안철수 연구소, http://www.ahnlab.com
[2] 하우리, http://hauri.co.kr
[3] Y. Liao, V. R. Vemuri, "Using Text Categorization Techniques for Intrusion Detection", 11th USENIX Security Symposium, San Francisco, August, 2002.
[4] C. D. Manninig, H. Schutze, "Foundations of Statistical Natural Language Processing", The MIT Press, pp. 495-528, 2001.
[5] B. Yates, R. Neto, "Modern Information Retrieval", Addison-Wesley, 1999.
[6] 문재근, "IR 기반의 유사도 알고리즘을 이용한 침입 탐지 기법 연구", 2002.
[7] 배병우, "정적 분석 기법을 이용한 악성 스크립트 탐지", 2002.