Edit Distance를 이용한 오용탐지 시스템의 침입유형 판별

구자민 ○ 조성배
연세대학교 컴퓨터과학과
icicle@candy.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

Intrusion Types Identification for HMM-based Anomaly Detection System Using Edit Distance

Ja-Min Koo and Sung-Bae Cho
Dept. of Computer Science, Yonsei University

요 약

전산 시스템에 대한 침입에 대응하기 위하여 시스템 호출 감사자료 척도를 사용하여 은닉 마르코프 모델(HMM)에 적용하는 비정상행위 기반 침입탐지 시스템에 대한 연구가 활발하다. 하지만, 이는 일정한 임계값 이하의 비정상행위만을 감지할 뿐, 어떠한 유형의 침입인지를 판별하지 못한다. 이에 Viterbi 알고리즘을 이용하여 상태 시퀀스를 분석하고, 공격 유형별 표준 상태시퀀스와의 유사성을 측정하여 유형을 판별할 수 있는데, 외부 혹은 내부 환경에 따라 상태 시퀀스가 항상 규칙적으로 추출될 수 없기 때문에, 단순 매칭으로 침입 유형을 판별하기가 어렵다. 본 논문에서는 이러한 문제를 해결하기 위하여 시퀀스의 변형을 효과적으로 고려하는 편집거리(Edit distance)를 이용하여 어떠한 유형의 침입이 발생하였는지를 판별하는 방법을 제안한다. 본 논문에서는 루트권한을 취득하기 위한 대표적인 침입유형으로 가장 널리 쓰이는 버퍼오버플로우 공격에 대해 실험하였는데, 그 결과 세부적인 침입 유형을 잘 판별할 수 있음을 확인하였다.

1. 서론

전산 시스템에 대한 침입의 피해를 최소화 하기 위하여 연구가 활발하게 이루어지고 있으며, 그 중 가장 대표적인 것이 침입탐지 시스템(Intrusion Detection System: IDS)이다. 하지만, 대부분의 침입탐지 시스템은 높은 침입 탐지율에 중점을 두어 개발되고 있기 때문에, 침입의 원인과 경로를 추적하는데는 기술적인 어려움이 있다. 특히 오용 침입탐지 시스템은 알려진 침입에 대해서는 거의 정확한 탐지를 할 수 있지만, 감사 정보에 대한 의존도가 높고 새로운 침입을 탐지할 수 없는 단점이 존재한다. 이로 인하여, 새로운 침입이 시도될 경우, 그 대응책을 강구하기 힘들어 진다.

이러한 단점을 보완하기 위한 방법으로 Viterbi 알고리즘을 이용하여 상태 시퀀스를 분석한 후, 각 공격 유형별 상태 시퀀스와의 유사성을 측정하여 유형을 판별할 수 있다[1]. 하지만, 상태 시퀀스가 항상 일정하지 않고 외부 혹은 내부적 요인에 따라 삽입, 삭제 혹은 치환된 결과의 시퀀스가 추출될 수 있다. 이로 인하여, 판단할 수 있는 침입임에도 불구하고 침입의 유형을 파악하는데 어려움이 생기게 된다. 본 논문에서는 유사성 판단 척도 중 이러한 문제에 잘 적용할 수 있는 방법인 편집거리를 이용하여 침입유형을 판별하는 방법을 제안하고, 다른 유사성 척도와 성능을 비교하여 제안한 방법의 유용성을 보이고자 한다.

2. 편집거리

편집거리란 동적 프로그래밍 (Dynamic Programming)을 기반으로, 두 개의 문자열을 비교할 때, 삽입, 삭제 또는 치환 등의 연산을 통해 같게 만드는 최소한의 연산수를 말한다. 이때, 모든 연산은 한번 행해질 때마다 동일하게 1씩 증가하며 계산 방법 및 수식은 다음과 같이 정의된다[2].

- $\delta(\varepsilon, a)$: 문자 a가 문자열 ε에 삽입 됨
- $\delta(a, \varepsilon)$: 문자 a가 문자열 ε에서 삭제 됨
- $\delta(a, b)$: 문자 a에서 b로 치환 됨

행렬 $C_{0..|x|,0..|y|}$에서, $C_{i,j}$는 $x_{1...i}$와 $y_{1...j}$를 매칭하기 위한 최소한의 연산수를 나타낸다. 두 문자열 x, y의 편집거리를 $ed(x,y)$라고 할 때, $C_{i,j}=ed(x_{1...i}, y_{1...j})$이며 다음과 같이 계산된다.

$C_{i,0} = i$

$C_{0,j} = j$

$C_{i,j} = \min(C_{i-1,j-1} + \delta(x_i, y_j),\ C_{i-1,j}+1,\ C_{i,j-1}+1)$

만일 $\delta(a, b)$가 같다면 0이 되고, 그렇지 않다면 1이 되며, 최종적으로 $C_{|x|,|y|}=ed(x,y)$를 계산하게 된다. 표 1은 ed("survey", "surgery")의 계산과정을 보여준다.

표 1. "survey"와 "surgery"의 편집거리 계산

		s	u	r	g	e	r	y
	0	1	2	3	4	5	6	7
s	1	**0**	1	2	3	4	5	6
u	2	1	**0**	1	2	3	4	5
r	3	2	1	**0**	1	2	3	4
v	4	3	2	1	**1**	2	3	4
e	5	4	3	2	2	**1**	**2**	3
y	6	5	4	3	3	2	2	**2**

표 1에서 보는 바와 같이, 진한 색으로 표시된 부분이 각 행과 열이 매칭되는 지점을 가리킨다. 예를 들

어, "surgery"의 's'와 "survey"의 모든 문자를 매칭하며, 그 중 's'와 매칭될 때 두 문자가 일치하므로 그 값은 0이 되고 매칭에 성공했다고 한다. 다음으로, 각각 'u', 'r'역시 매칭에 성공하지만, 4번째 문자 'g'에서 일치하는 문자가 없으므로, 치환이 발생하여 그 패널티로 1이 증가하게 된다. 또한 'survey'에서 'e'와 'y' 사이에 'r'이 빠져 있으므로, 삭제 연산이 발생하여 역시 1이 증가하게 된다. 이런 과정을 문자의 끝까지 비교하면 둘 사이의 편집거리는 2가 된다.

3. 제안하는 방법

본 논문에서는 사용된 침입 유형별로 기준이 되는 상태 시퀀스 값을 Viterbi 알고리즘을 이용하여 구한다. 이것을 기준으로 침입이 발생되었을 때 분석된 상태 시퀀스와의 유사성을 비교하여 침입 유형을 판별한다. 이때 유사성을 판별하기 위해서 편집거리를 이용하는데, 제안하는 시스템의 구조는 그림 1과 같다.

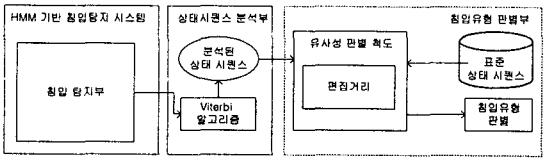

그림 1. 제안하는 시스템 구조

3.1 HMM기반 침입탐지 시스템

솔라리스에서 제공하는 BSM을 이용한 감사데이터는 운영체제에서 발생한 시스템 호출 이벤트들과 그에 관련된 사용자 및 프로세스 정보를 포함하는 자료로 침입탐지 시스템에서 가장 많이 쓰이는 척도이다. 본 논문에서는 기존의 음성인식이나 영상인식 분야에서 널리 쓰이는 HMM을 기반으로 한 침입탐지 시스템을 사용하였는데 이는 HMM이 실제적인 생성 경위를 알기 힘든 이벤트 시퀀스를 잘 모델링할 수 있는 방법으로 시스템 호출 감사자료를 이용하는 데에 유용하기 때문이다[3].

HMM은 상태 전이 확률 분포 A, 관측기호 확률 분포 B, 그리고 초기상태 확률 분포 π로 구성되며, 일반적인 기호로는 $\lambda=(A,B,\pi)$로 나타낸다[4].

HMM 을 기반으로 한 침입 탐지 시스템에서는 생성된 정상모델을 기반으로 새로 입력된 시스템호출 감사자료를 비교하여, 정상행위 모델에서 입력된 행위가 나올 확률을 계산한다. 이 확률값이 정상행위 모델링에서 구한 임계값보다 낮으면 침입으로 판정한다.

3.2 Viterbi 알고리즘을 이용한 상태 시퀀스 분석

앞선 침입탐지 모듈에서 침입이라고 판정되면, 그 시점의 시스템호출 이벤트의 상태 시퀀스를 알아내기 위해 상태 시퀀스 분석부에 Viterbi 알고리즘을 이용한다. Viterbi 알고리즘은 주어진 상태 시퀀스에서 가장 높은 확률을 가진 상태전이 경로를 찾아주는 것으로서[5], 본 논문에서는 HMM 기반 정상 행위 모델에 시스템 호출 시퀀스를 입력으로 넣어, 각 정상행위에서 현재 행위가 생성되었을 확률이 가장 높은 상태 시퀀스를 구하기 위해 사용한다. Viterbi 알고리즘은 시간 t일 때 상태 i에 있을 확률인 $\delta_t(i)$, 상태 j로 전이될 가장 높은 확률 상태를 가리키는 $\psi_t(j)$, 가장 높은 확률의 상태시퀀스를 가진 s_t^*의 변수로 구성된다. 주어진 상태 시퀀스 $O=O_1,O_2,\cdots,O_T$를 시스템 호출 이벤트로 매칭한다.

3.3 편집거리를 이용한 침입유형 판별

앞선 단계에서 구해진 상태 시퀀스와, 경험적으로 얻어진 공격 유형별 상태 시퀀스와의 유사성을 측정하기 위해 편집거리를 이용한다.

예를 들어, 각 유형별로 표 2와 같은 기준 시퀀스와 입력값이 있을 경우, 편집거리의 결과는 다음과 같다.

표 2. 유형별 상태 시퀀스 및 입력 상태 시퀀스

A	1	2	3	4	5
B	1	2	3	2	4
C	1	2	3	5	5
입력	1	3	3	5	

각 A, B, C와 입력 시퀀스 사이에는 한번의 치환과 한번의 삭제가 일어났다고 볼 수 있으며, 그의 편집거리는 모두 2가 된다. 따라서, 입력된 시퀀스는 어떤 유형인지를 판단할 수 없는 약점을 지니고 있다. 따라서, 본 논문에서는 이러한 약점을 보완하기 위하여 기존의 편집거리를 응용한 방법을 이용한다. 기본 틀은 같으나, 각 연산자 마다 가중치를 달리 하고, 두 개의 시퀀스를 매칭할 때 그 차이만큼을 가중치와 곱하여 편집거리를 계산하며 각 연산자 별 가중치는 삽입의 경우 1, 삭제는 2, 그리고 치환은 3으로 둔다.

4. 실험결과

정상행위를 모델링하는 데에 6명의 사용자가 보름동안 16,470개의 명령어를 입력하여 발생한 160,448개의 이벤트가 수집된 13MB의 감사자료를 사용하였으며, 탐지 대상 감사 자료에 쓰인 공격유형은 버퍼오버플로우 공격을 사용하였다. 버퍼오버플로우 공격은 시스템의 취약점을 이용하여 루트 권한을 획득하는 방법으로, 가장 많이 쓰이는 침입 유형이므로, 이 공격을 대상으로 본 논문에서는 침입유형을 판별하는 실험을 하였다. 각 침입 유형과 침입 횟수는 표 3과 같다.

표 3. 침입 유형 및 침입 횟수

유형	침입형태	횟수
Buffer Overflow	Open View xlock Heap Overflow	30
	Lpset - r Buffer Overflow Vulnerability	30
	Kcms_configure KCMS_PROFILES	30

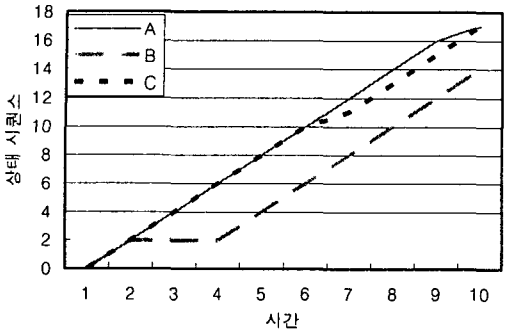

그림 2. 공격 유형별 기준 상태 시퀀스 변화

그림 2는 각 공격 유형별 기준이 되는 상태시퀀스 값으로 침입 유형별로 30회의 침입을 시도하고, Viterbi 알고리즘을 이용하여 상태 시퀀스를 추출한 결과이다. A는 xlock, B는 lpset, C는 kcms_sparc를 가리킨다.

HMM의 상태수와 한번에 입력될 시스템호출 이벤트 길이를 여러가지로 변화시켜 HMM 구성을 바꾸며 실험을 반복하였다. 상태수 5, 10, 15일 경우에는 모든 침입 유형의 상태 시퀀스가 동일하게 나타났다. 따라서 본 논문에서는 상태수 20의 환경에서 100%의 침입 탐지율과 최소의 탐지 오류율을 내는 HMM 임계값 -20.83에서 침입 유형 판별 실험을 하였다. 침입 유형을 판별하기 위한 척도인 편집거리 외에 유클리드 거리를 이용하여 그 결과를 비교하였으며 그 결과는 표 4, 표 5와 같다.

표 4. 유클리드 거리를 이용한 결과

	Trial	Correct	Rate
Xlock	30	24	80%
Lpset	30	26	86%
kcms	30	27	90%

표 5. 편집거리를 이용한 결과

	Trial	Correct	Rate
Xlock	30	28	93
Lpset	30	29	96
kcms	30	29	96

표 6. 공격 유형별 상태 시퀀스 및 입력 시퀀스

Xlock	0	2	4	6	8	10	12	14	16	17
Lpset	0	2	2	2	4	6	8	10	12	14
Kcms	0	2	4	6	8	10	11	13	15	17
Input	0	2	3	5	7	9	11	13	17	18

표 6은 실제 kcms 공격의 상태 시퀀스와 각각의 표준 상태 시퀀스를 보여주고 있다. 이를 기초로, 유클리드 거리를 사용하여 침입 유형을 판별하면, 입력 시퀀스는 xlock과 가장 큰 유사성(2.8284)을 가지게 되므로 xlock이라고 판별하게 된다. 하지만, 편집거리를 사용하여 실험하면, xlock과의 편집거리, 즉 시퀀스가 변화된 수는 8, kcms와의 편집거리는 7로서 kcms라고 정확하게 판별하게 된다. 즉, 유클리드 거리를 사용하였을 때 시퀀스의 변형으로 인하여 오판하는 경우를 편집거리를 사용하게 되면 적절한 연산을 수행함으로 오판률을 감소시켜 주게 되며, 그 결과 표 4, 표 5에서 보는 바와 같이 편집거리를 이용하는 것이 더 좋은 성능을 보임을 알 수 있었다.

5. 결론

본 논문에서 HMM 기반의 비정상행위 침입탐지 시스템에서 침입유형을 판별하기 위한 방법을 제안하고, 그 가능성을 밝히기 위해 상태수를 5, 10, 15, 20 으로 각각 변화해가며 실험을 하였다. 실험 결과, 버퍼오버플로우 공격의 침입 유형을 판별하는데에 유클리드 거리 함수를 사용하는 것보다 편집거리를 사용하는 것이 더 성능이 좋음을 보였다. 이로써 상태 시퀀스가 내외적 환경의 영향으로 변경되어 추출될 수 있음을 알 수 있었다.

하지만, 침입유형을 판별하기 위해서는 최소 20개 이상의 상태수가 필요하였는데, 상태수가 많아짐에 따라 결과를 알아내는 데에 걸리는 시간이 많이 소요되므로, 적은 상태수에서 침입 유형을 알아낼 수 있는 방법에 대한 연구와 지금은 버퍼오버플로우 공격에만 국한된 결과를 다른 침입 종류를 판별할 수 있도록 확장하는 연구가 필요하다.

감사의 글

본 연구는 대학 IT 연구센터 육성/지원 사업의 연구 결과로 수행되었음

참고문헌

[1] J.-M. Koo and S.-B. Cho, "Viterbi algorithm for intrusion type identification in anomaly detection system," *Proc. of WISA2003*, vol. 4, pp. 115-128, 2003

[2] G. Navarro, "A guided tour to approximate string matching," *ACM Computing Surveys*, vol. 33, no. 1, pp. 31-88, 2001.

[3] S.-B. Cho and H.-J. Park, "Efficient anomaly detection by modeling privilege flows using hidden Markov model," *Computers & Security*, vol. 22, no. 1, pp. 45-55, 2003.

[4] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. of IEEE*, vol. 77, no. 2, pp. 257-286, February 1989.

[5] G. D. Forney Jr. "Maximum-likelihood sequence detection in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. 18, no. 30, pp. 363-378, May 1972.