침입탐지시스템의 유사 패턴 매칭을 위한 알고리즘

정유석, 홍만표
아주대학교 정보통신공학과
j8508, mphong@madang.ajou.ac.kr

Algorithm for Similar Pattern Matching on Intrusion Detecting System

Yoo-Suk Jung, Man-Pyo Hong
Dept. of Information & Community, Ajou University

요 약

최근 정보통신 관련 시장의 양적 팽창과 함께 발생된 수많은 시스템 침입 사건들은 컴퓨터 보안 문제와 이를 해결하기 위한 보안 시스템에 대한 관심을 증가시키고 있다. 침입탐지시스템은 보안 시스템의 핵심 요소로, 그 중 대부분이 패턴 매칭을 이용한 침입탐지방식을 채택하고 있다. 그러나 현재의 패턴 매칭을 위한 알고리즘은 유연하지 못하기 때문에, 다양한 특성을 지닌 공격들에 대한 대처 능력이 부족하다. 이를 해결하기 위해 본 논문에서는 패턴들의 특성에 따라 유연하게 대처할 수 있는 세 가지의 유사 패턴 매칭 알고리즘을 제안한다.

1. 서론

최근 정보통신 관련 시장의 양적 팽창과 함께 발생하는 수많은 시스템 침입 사건들은 컴퓨터 보안 문제에 대한 관심을 고조시키고 있다.

불법적인 침입으로부터 컴퓨터를 보호하기 위해 침입을 탐지하고 이에 대한 적절한 조치를 취하는 역할을 수행하는 침입탐지시스템(Intrusion Detection System : IDS)은 방화벽(Firewall)과 함께 보안 시스템의 중요한 요소로 인식되고 있으며[1, 2, 3], 현재 이에 대한 연구가 활발히 진행되고 있다. 침입 탐지 시스템은 크게 데이터의 소스를 기반으로 하는 분류 방법과 침입의 모델을 기반으로 하는 분류 방법으로 나눌 수 있으며, 데이터 소스를 기반으로 하는 분류 방법은 호스트로부터 생성되고 모아진 감사(audit) 데이터를 침입 탐지에 사용하는 호스트 기반(host based)과, 네트워크의 패킷 데이터를 모아 침입을 탐지하는데 사용하는 네트워크 기반(network based)으로 구분할 수 있다. 또한 침입 모델을 기반으로 하는 분류방법은 정상적인 시스템 사용에 관한 프로파일과 시스템 상태를 유지하고 있는 동안 이 프로파일에서 방어하는 행위들을 탐지하는 비정상적인 행위탐지(anomaly detection) 방법과, 시스템의 알려진 취약점들을 이용한 공격 행위들에 대한 공격 특징 정보를 통해 침입을 탐지하는 오용 침입탐지(misuse detection) 방법으로 분류할 수 있다[7].

비정상적인 행위탐지 방법과 오용 침입탐지 방법은 여러 가지 방식으로 침입 여부를 판단 하는데, 그 중 패턴 매칭(Pattern Matching)은 선언적인 기술(Declarative Specification), 이식성(Portability), 효과적인 실시간 탐지율, 특정 공격 특징 선별 가능 등 많은 이점을 가지고 있다. 따라서 오용침입 탐지 방법에 많이 사용 되어 왔으며, 최근 면역 기반 침입탐지 방법과 관련한 비정상적인 행위탐지 방법에도 이용되고 있다.[4, 8, 9]

그런데 침입탐지시스템에 이용되는 패턴 매칭 방법은 다양한 침입 패턴에 대응 하기 위해서 완전 패턴 매칭(Perfect Pattern Matching)이 아닌 유사 패턴 매칭(Similar Pattern Matching)이 되어야 하는데, 현재까지 이용되고 있는 패턴 매칭 방법들은 대부분 패턴 특성의 변화에 유연하게 대응하기가 힘들다.

이를 개선하기 위해 본 논문에서는 패턴 매칭 방법을 사용하는 침입탐지시스템을 위한 유사도(Similar Value : SV) 측정 알고리즘을 제안하여 침입탐지시스템의 정확성과 유연성을 높이려고 한다.

2. 유사도 측정과 관련된 기존 연구

패턴 매칭을 사용하는 침입탐지시스템의 경우 침입과 관련된 모든 패턴을 갖는다는 것은 불가능하기 때문에 기존에 존재하는 각 패턴들을 이용하여 그와 유사한 여러 패턴까지 인식하는 것은 매우 중요한 일이다.

그러나 현재까지 이와 관련된 연구는 거의 없는 실정이며, 대부분의 패턴 매칭 침입탐지 시스템은 패턴간의 단순 이벤트 비교를 사용하거나, 은닉마코프모델(Hidden Markov Model)등의 신경망을 사용하여 유사도를 측정하여왔다[5, 6].

그러나 신경망을 통한 유사도 측정은 그 결과가 학습을 통한 우연성에 기인하므로 유사도에 대한 확실한 보장을 할 수가 없을 뿐더러, 많은 연산량을 요구하는 기술이므로 침입탐지 시스템에 적용하기가 어렵다.[9]

DTW(Dynamic Time Warping) 알고리즘은 대표적인 유사 패턴 인식을 위한 방법으로, 고립 단어에 대한 인식율이 높다는 장점이 있는 반면 계산량이 상당히 방대하고 융통성이 부족해 다양한 패턴에 대한 적응력이 충분하지 못하다.

3. 패턴 매칭 기법을 사용하는 침입탐지시스템의 유사도 측정을 위한 알고리즘

3.1 정의

[정의 1] 기준 행위 패턴(Standard Behavior Pattern : SBP)

기준 행위 패턴은 침입 탐지의 기준이 되는 일정 길이의 이벤트 패턴을 의미하며 다음과 같이 표현한다.

$$SBP = Ev_1, Ev_2, Ev_3, \cdots Ev_n$$

이때 Ev는 개별 이벤트를, n은 단위 이벤트 길이를 의미한다.

[정의 2] 비교 행위 패턴(Compare Behavior Pattern: CBP)

비교 행위 패턴은 비교하려는 대상 이벤트 패턴을 의미하며 다음과 같이 표현한다.

$$CBP = Ev_1, Ev_2, Ev_3, \cdots Ev_n$$

이때 Ev는 개별 이벤트를, n은 단위 이벤트 길이를 의미한다.

[정의 3] 대응 사건(Matching Event : ME)

대응 사건은 임의의 사건에 대응되는 사건을 가리키며 다음과 같이 표현한다.

$$ME_{pi} = Ev_{qj}$$

이때 p, q는 대응되는 패턴을 의미하며 i, j는 각 패턴에서의 사건 위치를 의미한다.

[정의 4] 대응요소거리(Matching Elements Distance:MED)

대응 요소 거리는 기준 행위 패턴의 요소와 이에 대응 되는 비교 대상 행위 패턴의 요소와의 거리를 의미하며 다음과 같이 표현한다.

$$MED_i = dist.\ from\ Ev_i\ to\ ME_i$$

이때 i는 요소의 위치를 의미한다.

[정의 5] 요소 거리 함수(Element Distance Function:EDF)

요소 거리 함수는 행위 패턴간의 거리를 결정하기 위한 함수로 파라매터로 결정되며 다음과 같이 표현한다.

$$EDF(d)$$

이때 d는 요소간의 거리이다.

[정의 6] 유사도(Similar Value : SV)

두 패턴간의 유사도는 기준 행위 패턴에 대해 비교 행위 패턴이 얼만큼 유사한가를 나타내는 값이며 다음과 같이 표현된다.

$$SV = 1 - \frac{\sum_{i=0}^{n} EDF(MED_i)}{n \times EDF(n)}$$

이때 n은 행위 패턴의 길이이다.

3.2 새로운 알고리즘

새로운 알고리즘들은 비교 행위 패턴의 각 요소(사건)가 정상 행위 패턴의 동일 요소와 얼마나 떨어져 있는가를 계산함으로 유사도를 측정하게 된다. 각 알고리즘들은 유사도를 측정하기 위한 계산 우선 순위에 따라 다음과 같이 분류된다.

3.2.1 중복 허용 유사도 측정 알고리즘(Overlapable Similar Pattern Matching Algorithm)

중복 허용 유사도 측정 알고리즘은 기준 행위 패턴의 각 요소 관점에서 비교 행위 패턴의 대응 요소를 찾고 요소 거리 값을 결정한다. 만약 기준 행위 패턴의 두개 이상의 요소가 비교 행위 패턴의 동일 요소를 대응 요소로 갖는 경우 해당 되는 기준 행위 요소의 대응 요소 모두를 인정하게 된다.

요소 거리 값은 요소 거리 함수를 통해 계산 되는데, 이 함수는 필요에 따라 변경시킬 수 있다. 본 논문에서는 요소 거리 함수에 독립적으로 수행될 수 있는 알고리즘을 소개하며, 요소 거리 함수 자체에 대한 언급은 하지 않는다.

예를 들어 요소 거리 함수는 $EDF(d) = x$, 두 패턴이 $NBP = \{abccade\}$, $CBP = \{bcccaed\}$로 정의 되었을 때, NBP, CBP 간의 유사도는 다음과 같다.

$$SV = 1 - \frac{\sum_{i=1}^{n} EDF(MED_i)}{n \times EDF(n)} = 1 - \frac{\sum_{i=1}^{n} MED_i}{n^2}$$
$$= 1 - \frac{4+1+0+0+0+1+1}{49} = 1 - \frac{7}{49} \cong 0.8571$$

다음은 중복 허용 유사도 측정을 위한 알고리즘이다.

Algorithm *OverlapableSimilarPatternMatching(SBP, CBP, EDF)*
Input : SBP (Standard Behavior Pattern)
CBP (Compare Behavior Pattern)
EDF (Element Distance Function)
Output : tDistance (a distance of NBP and CBP)

```
begin
    for i = 1 to |SBP|
        insert position of closest same value on CBP with
        SBP[i] to e
        tDistance = tDistance + EDF(|i - e|)
end
```

3.2.2 중복 불허용 거리 우선 유사도 측정 알고리즘 (Non-Overlapped Distance First Similar Pattern Matching Algorithm)

중복 불허용 거리 우선 유사도 측정 알고리즘은 중복 허용 유사도 측정 알고리즘과 마찬가지로 정상 행위 패턴의 각 요소 관점에서 비교 행위 패턴의 대응 요소를 찾고 요소 거리 값을 결정한다. 이때 요소 거리 값은 정상 행위 패턴의 관점에서 결정되며, 대응 요소는 비교 행위 패턴의 관점에서 결정된다.

비교 행위 패턴 내에 존재하는 요소들의 대응 요소가 동일한 경우 가까운 요소 거리 값을 갖는 요소만이 인정 되며, 그 외의 요소들은 새로운 대응 요소를 찾게 된다.

중복 허용 유사도 측정 알고리즘과 마찬가지로 각 요소들의 요소 거리 값은 일정 함수를 통해 결정 되는데, 이 함수는 환경변수로 결정 된다.

예를 들어 요소 거리 함수는 $EDF(d) = x$, 두 패턴이 $NBP = \{abccade\}$, $CBP = \{bcccaed\}$로 정의 되었을 때, NBP, CBP 간의 유사도 SV는 다음과 같다.

$$SV = 1 - \frac{\sum_{i=1}^{n} EDF(EDV_{Ai})}{n^2} = 1 - \frac{\sum_{i=1}^{n} EDV_{Ai}}{n^2}$$
$$= 1 - \frac{7+1+0+0+0+1+1}{49} = 1 - \frac{10}{49} \cong 0.7959$$

다음은 중복 불허용 거리 우선 유사도 측정을 위한 알고리즘이다.

Algorithm *DistanceFirstSimilarPatternMatching(SBP, CBP, EDF)*
Input : SBP (Standard Behavior Pattern)
CBP (Compare Behavior Pattern)
EDF (Element Distance Function)
Output : tDistance (a distance value of NBP and CBP)

```
begin
    for i = 1 to length of SBP
        SearchCloestEvent(i, 0, SBP, CBP)
    for i = 0 to length of SBP
        tDistance = tDistance + EDF(MEDi)
end

procedure SearchClosestEvent(i, initCmpEvent, SBP, CBP)
Input : SBP (Standard Behavior Pattern)
        CBP (Compare Behavior Pattern)
        i (Event Position of SBP)
        initCmpEvent (Start Position of Searching)

begin
    insert a event, which is a closest same value among
    farther than initCmpEvent on CBP with SBP[i], to
    Candidate
    CandidateMED = distance from Evi to Candidate
    if CandidateMED < MED_CandidateMED
        ME_CandidateMED = ith event of SBP
        MED_CandidateMED = CandidateMed
        if ME_Candidate is not a end of CBP
            SearchClosestEvent(Candidate, ME_Candidate,
            SBP, CBP)
    else
        if Candidate is not a end of CBP
            Candidate=SearchClosestEvent(i,Candidate,
                    SBP, CBP)
end
```

3.2.3 중복 불허용 방향 우선 유사도 측정 알고리즘 (Non-Overlapped Direction First Similar Pattern Matching Algorithm)

중복 불허용 방향 우선 유사도 측정 알고리즘은 중복 허용 유사도 측정 알고리즘과 마찬가지로 정상 행위 패턴의 각 요소 관점에서 비교 행위 패턴의 대응 요소를 찾고 요소 거리 값을 결정한다. 이때 요소 거리 값은 정상 행위 패턴의 관점에서 결정되며, 대응 요소는 비교 행위 패턴의 관점에서 결정된다.

비교 행위 패턴 내에 존재하는 요소들의 대응 요소가 동일한 경우 가까운 요소 거리 값을 갖는 요소만이 인정 되며, 그 외의 요소들은 새로운 대응 요소를 찾게 된다. 이런 경우 새로 찾은 대응 요소가 이전에 찾았던 요소와 같은 방향에 존재할 경우 그 요소는 무시된다.

중복 허용 유사도 측정 알고리즘과 마찬가지로 각 요소들의 요소 거리 값은 일정 함수를 통해 결정 되는데, 이 함수는 환경변수로 결정 된다.

예를 들어 요소 거리 함수는 $EDF(d) = x$, 두 패턴이 $NBP = \{abccade\}$, $CBP = \{bcccaed\}$로 정의 되었을 때, NBP, CBP 간의 유사도 SV는 다음과 같다.

$$SV = 1 - \frac{\sum_{i=1}^{n} EDF(EDV_{Ai})}{n^2} = 1 - \frac{\sum_{i=1}^{n} EDV_{Ai}}{n^2}$$
$$= 1 - \frac{7+1+0+0+0+1+0}{49} = 1 - \frac{9}{49} \cong 0.8163$$

<알고리즘 3>는 중복 불허용 방향 우선 유사도 측정을 위한 알고리즘이다.

Algorithm *DirectionFirstSimilarPatternMatching(SBP, CBP, EDF)*
Input : SBP (Standard Behavior Pattern)
CBP (Compare Behavior Pattern)
EDF (Element Distance Function)
Output : tDistance (a distance value of NBP and CBP)

```
begin
    for i = 1 to length of SBP
        SearchCloestEvent(i, 0, SBP, CBP)
    for i = 0 to length of SBP
        tDistance = tDistance + EDF(MEDi)
end

procedure SearchClosestEvent(i, initCmpEvent, SBP, CBP)
Input : SBP (Standard Behavior Pattern)
        CBP (Compare Behavior Pattern)
        i (Event Position of SBP)
        initCmpEvent (Start Position of Searching)

begin
    insert a event, which is a closest same value among
    farther than initCmpEvent on CBP with SBP[i], to
    Candidate
    if Direction of Candidate is not same to initCmpEvent
        CandidateMED = distance from Evi to Candidate
        if CandidateMED < MED_CandidateMED
            ME_CandidateMED = ith event of SBP
            MED_CandidateMED = CandidateMed
            if ME_Candidate is not a end of CBP
                SearchClosestEvent(Candidate,
                ME_Candidate, SBP, CBP)
        else
            if Candidate is not a end of CBP
                Candidate = SearchClosestEvent(i,
                Candidate, SBP, CBP)
end
```

<알고리즘 3>

4. 결론 및 향후과제

시스템에 대한 침입 행위들은 그 패턴의 변화가 다양하여, 이에 대한 감지 방식이 해당 패턴의 특성에 따라 유연하게 이루어져야만 한다. 이를 위해 본 논문에서는 무속성, 거리우선, 위치우선 등의 유사성 관련 속성과 연관하여 유사 패턴 매칭 알고리즘을 제안하였으며, 또 알고리즘에 대해 유사한 정도를 필요에 따라 요소 거리 함수로 제공할 수 있게 하여 기존의 유사성 비교를 위한 알고리즘에 부족했던 유연성을 부여하였다.

중복 허용 유사도 측정 알고리즘은 패턴의 각 요소의 거리나 위치에 관계없이 유사성을 계산한다. 또한 중복 불허용 거리 우선 유사도 측정 알고리즘은 패턴 내의 요소에 대한 독립성을 강화시켜 유사도를 계산하며, 중복 불허용 위치 우선 유사도 측정 알고리즘은 패턴 내의 대응되는 요소간의 방향성을 강화시켜 유사도를 계산한다.

본 연구에서는 무속성, 거리우선, 위치우선에 대한 속성과 관련된 알고리즘들을 제안하였지만, 계속적으로 새로운 속성과 관련된 알고리즘을 개발하고, 제안된 알고리즘들을 통합에 관련된 연구를 하고자 한다.

5. 참고문헌

[1] James Cannady, Jay Harrell, "A Comparative Analysis of Current Intrusion Detection Technologies," http://iw.gtri.gatech.edu/Papers/ids_rev.html, 1998. 2.

[2] Mansour Esmaili, Rei Safavi-Naini, "Case-Based Reasoning for Intrusion Detection," Computer Security Applications Conference pp.214-222

[3] Jai Sundar B. Spafford E, "Software Agents for Intrusion Detection," Technical Report, Purdue University, Department of Computer Science, 1997.

[4] Crosbie M, Spafford E, |applying Genetic Programming to Intrusion Detection" Technical Report, Purdue University, Department of Computer Science, 1996

[5] H. Debar, M. Dacier M. Nassehi and A. Wespi, "Fixed vs, Variable-Length Patterns for Detecting Suspicious Process Behavior," Research Report, RZ 3012 IBM Zurich Research Laboratory, 1998

[6] Christina Warrender, Stephanie Forrest and Barak Pearlmutter, "Detecting Intrusions Using System Calls:Alternative Data Models", 1999 IEEE Symposium, 1999. 5.

[7] 은유진, 박정호, "침입탐지 기술 분류 및 기술적 구성요소", 정보보호센터 정보보호 뉴스 1998. 7. 통권 13 호

[8] 전문석, "침입탐지 모델분석 및 설계," 정보보호센터 최종보고서, 1996, 12.

[9] 이종성, 채수환, "컴퓨터 면역 시스템을 기반으로 한 지능형 침입탐지시스템", 한국정보처리학회 논문지 제 6 권 제 12 호 1999. 12.