컴퓨터 면역시스템의 부정 및 긍정선택과 가중치를 이용한 알려지지 않은 공격탐지 연구

정일안° 김민수 노봉남
전남대학교 정보보호 협동과정
{mir°, phoenix}@lsrc.jnu.ac.kr, bbong@jnu.ac.kr

A Study of Unknown Attack Detection using Weight and Negative/Positive Selection of Computer Immune System

Ilahn Cheong° Minsoo Kim Bongnam Noh
Interdisciplinary program of Information Security, Chonnam National University

요 약

기존의 오용 기반 침입탐지 시스템에서는 변형되거나 새로운 해킹 방법에 대한 지속적인 탐지패턴을 지원해 주어야 하는 단점이 있다. 이러한 변형되거나 알려지지 않은 공격에 대한 탐지는 비정상행위 탐지 방법으로 본 논문에서는 컴퓨터 면역시스템의 부정 및 긍정선택 방법과 가중치의 특성을 이용하였다. 즉, 알려진 공격으로부터 특성을 추출하여 알려지지 않은 공격에 대응할 수 있도록 특성을 변경하는 방법을 사용하였다. 이러한 방법으로 공격 특성을 추출하고 특성 추출에 사용하지 않은 다른 공격에 대한 탐지를 실험한 결과 u2r 공격인 buffer overflow 공격과 race condition 공격에 대하여 정확한 탐지가 이루어짐을 보였다.

1. 서 론

최근 인터넷이 급속히 발전함에 따라 컴퓨터 시스템에 대한 해킹 방법도 다양해지고 있으며, 새로운 공격 기법들에 의해 중요한 시스템이 더욱 위협받고 있다. 이러한 공격들을 탐지하거나 대응하기 위해 침입탐지 시스템(Intrusion Detection System)의 발전을 가져왔다.

침입탐지 시스템은 크게 알려진 공격을 탐지하는 오용(misuse) 탐지와 비정상적인(anomaly) 행위를 탐지하는 방법으로 분류한다[1]. 기존의 오용 기반 침입탐지 시스템에서는 알려진 공격에 대해서는 탐지율이 높은 반면, 변형되는 해킹 방법에 대해서는 탐지율이 낮고 지속적으로 탐지패턴을 지원해 주어야 한다. 이에 대해 변형되거나 알려지지 않는 공격에 대한 탐지는 비정상행위 탐지방법으로 통계적인 방법, 예측 가능한 패턴 생성, 신경망, 데이터마이닝 등을 이용한 방법들이 연구되었으며[8], 또한 생물계의 면역시스템(Immune System)[2]과 유사한 컴퓨터 면역시스템을 이용한 방법 등도 연구되고 있다.

생물계의 면역시스템은 외부에서 침입하여 생체 내에 피해를 주는 항원에 의해 변이된 형태를 스스로 자기세포와 구별하여 인식하고 제거하도록 정교하고 복잡하게 구조화되어 있다. 이러한 특성들을 모델링하여 컴퓨터 시스템에 적용한 것이 컴퓨터 면역시스템(Computer Immune System)이다[3].

본 논문에서는 변형되는 공격에 대해 지속적인 탐지패턴을 지원해 주어야 하는 기존 오용 기반 탐지 방법을 보완하고자 컴퓨터 면역시스템의 특성을 이용하여 변형되거나 알려지지 않은 공격에 대한 탐지 방법을 연구하였다. 실험 결과, 탐지 패턴에 없는 공격에 대해서도 정확한 탐지가 이루어짐을 알 수 있었다.

2. 면역시스템

2.1 면역시스템의 구성

면역시스템은 복잡한 면역 반응을 하는 많은 면역 세포들이 유기적으로 결합하여 하나의 시스템을 구성한다. 여러 구성 요소들 중에서 면역 반응에 관계하는 요소들로 항원(Antigen), 항체(Antibody), T세포(T-cell), B세포(B-cell) 등이 있다.

면역 세포가 자기(self) 세포를 인지하는 방법으로 면역시스템은 주조직 적합성 복합체 인식부와 다양성을 내포하고 있는 항원수용체(Antigen Receptor)의 정상적인 동작여부를 확인하면서 면역 세포를 생성하는데, 부정선택(Negative Selection)과 긍정선택(Positive Selection) 방법이 있다.

2.2 부정 및 긍정선택

항원을 인식하는데 있어 자기를 항원으로 인식하는 것을 배제하기 위한 방법으로 면역세포에 주조직 적합성 복합체를 결합시켜 긍정적인 선택이 되는 세포들로 구성한다. 이 방법을 모델링한 사례로 D. Dasgupta와 S. Forrest의 Anomaly Detection Algorithm이 있다[5, 6, 7]. 이 알고리즘에서 부정선택을 이용하여 Anomaly detector를 구성하고 자기 인식 알고리즘에 적용하였다.

긍정선택은 각 면역세포의 주조직 적합성 복합체 인식기능을 확인하는 방법으로 갓 생성된 면역세포에 주조직 적합성 복합체를 결합시켜 긍정적인 선택이 되는 세포들로만 구성한다. 이 방법을 모델링하여 부정선택 알고리즘을 보완하고, 자기 공간 변경 방법에 따른 자기 인식률의 특성을 가지도록 하였다[9]. 이 두 가지 선택방법을 거친 면역세포는 주조직 적합성 복합체를 자신으로 인식하면서 이를 항원으로 인식하지 못하도록 구성하고 면역물질을 만들어내어 정상적인 면역시스템을 유지한다.

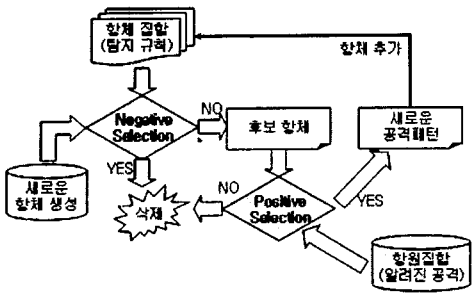
3. 부정 및 긍정선택 결합을 이용한 공격특성 추출

3.1 부정 및 긍정선택 결합 시스템의 전체적인 구성

본 논문에서 제시한 부정선택과 긍정선택을 결합하여 자동으

이 논문은 한국전자통신연구원의 관리로 수행되었음

로 새로운 공격 특성을 추출하는 시스템의 전체적인 구성은 <그림 1>과 같다.

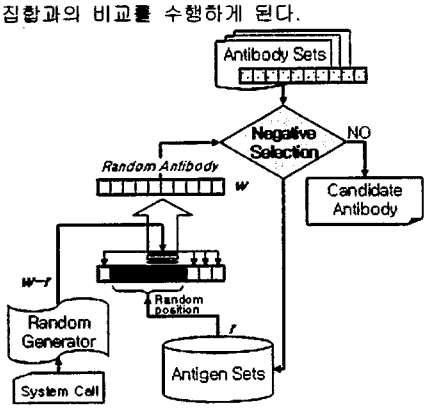


<그림 1> 부정 및 긍정선택 결합 시스템의 구성

먼저, 임의(random)로 새로운 항체를 생성시켜 항체 집합(탐지 규칙들)과 비교를 한다. 만약 기존에 존재하는 항체 집합이면 삭제하고(부정선택) 다시 새로운 항체를 생성하도록 한다. 기존에 존재하지 않는 새로운 항체라면 후보 항체로 선택한다. 다음 단계로, 알려진 공격에 대한 항원 집합에서 후보 항체와 비교하여 후보 항체가 존재하는지를 비교한다. 만약 매핑되지 않으면 후보항체를 삭제하고, 매핑되면(긍정선택) 후보항체를 새로운 공격패턴으로 분류하여 기존 항체 집합에 추가한다. 위와 같은 과정을 반복적으로 수행하여 자동으로 추가/삭제되도록 하여 항체집합을 형성하게 된다.

3.2 공격특성 추출방법

<그림 2>는 공격특성을 추출하는 방법을 나타낸 것으로, 후보 항체는 선택된 시스템 호출 번호를 대상으로 임의로 생성시켜 항원 집합과의 비교를 수행하게 된다.



<그림 2> 공격특성 추출방법

S. Forrest의 부정선택 알고리즘에서 랜덤하게 생성시킨 패턴과 자기로 구성한 것과 매칭시키는 방법에서는 스트링을 0과 1의 조합으로 변환하여 적용하였으나, 본 논문에서 척도(measure)로 사용한 시스템 호출은 중요도가 높은 69개를 사용하였기 때문에 같은 방법을 적용하기에 부적절하다. 또한, 본 논문에서는 69개의 시스템 호출로 패턴을 생성시킬 때에도 그 조합의 수가 매우 많아 지정된 개수(w ; 윈도우 사이즈)만큼 연속적으로 매칭될 확률이 극히 적고, 시간이 오래 걸리기 때문에 <그림 2>와 같이 항원 집합에서 일부를 선택하여 후보 항체를 생성하는 방법을 사용하였다. 즉, w 개가 연속적으로 맞춰질 때까지 수행되는 시간에 비해 w 개가 맞춰진 후 남은 위치에 임의적으로 선택하여 추가하는 시간이 훨씬 적게 걸리면서 같은 효과를 얻을 수 있기 때문이다. 그리고, 이러한 추출 방법은 변형된 공격의 패턴에 적용하여 효과적으로 탐지할 수 있다는 장점을 가지게 된다.

```
Input: r, w, Antigen Sets
Output: Candidate Antibody
begin
  Select r systemcall from Antigen Sets
  Add r into random position of 0 from (w-r)
  repeat
    Random generate one systemcall
    Add to position except filled positions
  until
end
```

3.3 탐지 방법

본 논문의 공격 특성추출 시스템에서 생성된 항체 집합은 다음과 같이 테스트 데이터와 비교하고, 유사도(Similarity)를 계산하여 자기인지(공격 특성과 유사한지)를 판단하게 된다. 실제 공격을 탐지하기 위한 패턴 매칭 방법은 아래와 같다.

```
Input: x_i, y_i, r, w
Output: Count number of matched pattern
begin
  initialize i
  repeat
    if x_i is equal to y_i
      then increasing matched cell's count
        if matched cell's count is equal to r
          then increasing matched pattern's count
      else continue
  until i≤(w-r)
end
```

이러한 패턴 매칭 방법과 가중치를 고려한 결과를 기반으로 유사도(%)를 계산한다.

$$Similarity = \frac{N_{match}}{N_{msets}} \times 100(\%) + weight$$

$$\text{if } N_{match} = 0,\ weight = -\frac{w_{slide}}{w_{\max}} \frac{1}{(w_{\max} - w_{slide} + 1)}$$

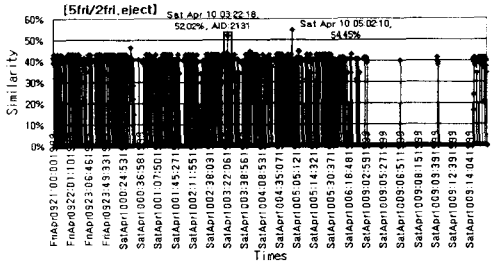
$$\text{if } N_{match} > 0,\ weight = \frac{w_{slide}}{w_{\max}} \frac{1}{(w_{\max} - w_{slide} + 1)} \frac{N_{match}}{N_{sum}}$$

여기서, N_{match}는 항체 집합과 매칭된 개수이고, N_{msets}는 항원 집합의 총 개수이다. 가중치는 윈도우 사이즈와 빈도수를 고려하여 발생빈도 확률값으로 계산한다.

4. 실험 및 분석

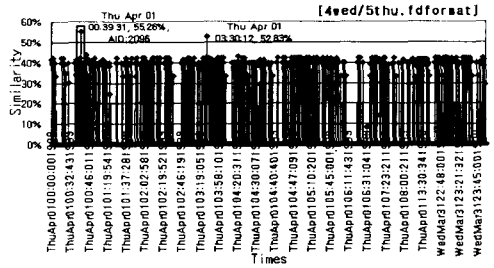
본 논문에서는 사용된 실험 데이터는 DARPA의 1999년 BSM 로그 데이터를 사용하였으며, 특히 u2r(user to root) 공격 형태의 시스템 호출 시퀀스를 대상으로 실험하였다. 공격패턴 추출 및 테스트용으로 사용한 공격으로 eject, fdformat, ffbconfig는 buffer overflow 공격이고 ps는 race condition 공격이다. DARPA BSM 로그 중에서 하나의 공격으로 공격 특징을 추출한 후에 다른 공격이 포함된 로그 데이터에 적용하여 탐지를 수행하는지를 실험하였다. 이러한 실험 방식은 기존의 알려진 공격에 대하여 특징을 추출한 후 알려지지 않는 공격에 대한 탐지가 가능한지를 검사하는 방법으로 해석할 수 있다.

<그림 3>은 2주 금요일 데이터에서 eject에 대한 공격 특징을 추출하여 5주 금요일 데이터를 대상으로 공격 부분을 탐지한 결과이다. 52.02%의 유사도에서 공격된 시간과 공격자의 ID를 표시해 주고 있다.



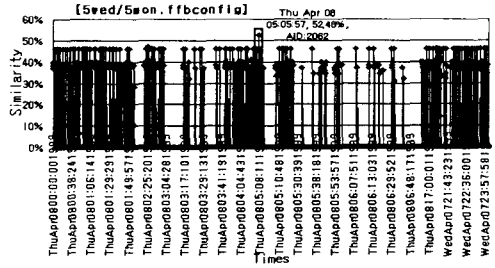
<그림 3> eject 공격에 대한 실험결과

<그림 4>는 5주 목요일 데이터에서 fdformat에 대한 공격 특징을 추출하여 4주 수요일 데이터를 대상으로 공격 부분을 탐지한 결과이다. 유사도가 55.26%인 부분에서 공격시간과 공격자의 ID를 보여주고 있다.



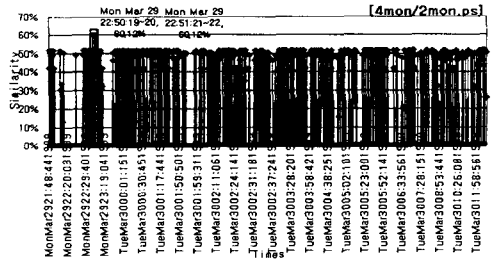
<그림 4> fdformat 공격에 대한 실험결과

<그림 5>는 5주 월요일 데이터에서 ffbconfig에 대한 공격 특징을 추출하여 5주 수요일 데이터를 대상으로 공격 부분을 탐지한 결과이다. 유사도가 52.48%인 부분에서 공격시간과 공격자의 ID를 보여주고 있다.



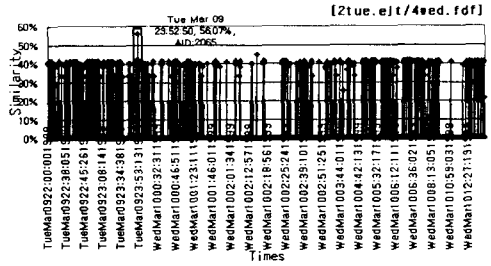
<그림 5> ffbconfig 공격에 대한 실험결과

<그림 6>은 2주 월요일 데이터에서 ps에 대한 공격 특징을 추출하여 4주 월요일 데이터를 대상으로 공격 부분을 탐지한 결과로, 60.12%의 유사도로 네 부분 모두에서 탐지되었다.



<그림 6> ps 공격에 대한 실험결과

다음으로 유사한 공격유형을 대상으로 서로 다른 공격 패턴을 사용하여 상호 탐지 가능성을 알아보는 실험을 하였다. <그림 7>은 4주 수요일 데이터에서 fdformat에 대한 공격 특징을 추출하여 2주 화요일 데이터를 대상으로 공격 부분을 탐지한 결과이다. 56.07%의 유사도에서 eject에 대한 공격 시간과 공격자의 UID를 탐지하였다.



<그림 7> fdformat 패턴의 eject 공격 탐지결과

5. 결론

기존의 오용 기반 침입탐지 시스템에서는 변형되는 해킹 방법에 대한 지속적인 탐지패턴을 지원해 주어야 하는 단점이 있다. 이러한 변형되거나 알려지지 않는 공격에 대한 탐지는 비정상행위 탐지방법으로 본 논문에서는 컴퓨터 면역시스템의 특성인 부정 및 긍정 선택을 결합하여 새로운 공격 특성을 추출하는 시스템을 제시하였다. 알려진 공격으로부터 특성을 추출하여 알려지지 않는 공격에 대응할 수 있도록 특성을 변경하도록 하였다.

실험 데이터 집합으로 DARPA BSM 로그를 사용한 결과 u2r 공격의 buffer overflow 공격에 대하여 50% 이상의 유사도로 공격이 이루어졌던 시간과 공격자를 탐지하였고, race condition 공격에서는 60%의 유사도로 공격을 탐지하였다. 또한, 유사한 공격형태의 서로 다른 공격에 대해서도 탐지가 이루어짐을 알 수 있었다. 본 논문에서 제안한 방법은 시스템에서 다른 유형의 공격이나 네트워크 공격에 대하여 적용할 수 있을 것으로 본다.

6. 참고 문헌

[1] D. E. Denning, "An Intrusion Detection Model," IEEE Trans. on Software Engineering, No.2, Feb., 1987.

[2] I. Roitt, J. Brostoff, D. Male, Immunology, 4th edition, Mosby, 1996.

[3] D. Dasgupta, Artificial Immune Systems and Their Applications, Springer, 1999.

[4] S. Forrest, L. Allen, A. S. Perelson, R. Cherukuri, "Self-Nonself Discrimination in a Computer," IEEE Symp. on Research in Security and Privacy, 1994.

[5] P. D'haeseleer, S. Forrest, P. Helman, "An Immunological Approach to Change Detection: Algorithms, Analysis and Implications," Proc. of IEEE Symp. on Security and Privacy, 1996.

[6] D. Dasgupta, "An Immune Agent Architecture for Intrusion Detection.", Proceedings of The GECCO 2000 Workshop Prog. pp. 42-44, 2000.

[7] D. Dasgupta, S. Forrest, "Novelty Detection in Time Series Data using Ideas from Immunology," IN Proceedings of The International Conference on Intelligent Systems, 1999.

[8] 김민수, 은유진, 노봉남, "UNIX 환경에서 퍼지 Petri net을 이용한 호스트 기반 침입탐지 시스템 설계", 한국정보처리학회 논문지, 제6권, 제7호, 1999년.

[9] 심귀보, 서동일, 김대수, 임기욱, 컴퓨터 면역시스템 개발을 위한 인공면역계의 모델링과 자기인식 알고리즘, 퍼지 및 지능시스템학회 논문지, 제11권, 제10호, pp. 910-918, 2002년 1월.