한국지능정보시스템학회
02 춘계학술대회 논문집, pp. 417~422

초돌연변이(Hypermutation)를 이용한 유전자 라이브러리 진화와 동적 선택 알고리즘

김정원* · 최종욱** · 김상진**

Dynamic Clonal Selection Algorithm with Gene Library Evolution using a Hypermutation

Jungwon Kim* · Jong Uk Choiu** · Sang Jin Kim**

요 약

인공면역시스템을 이용한 침입탐지시스템 개발을 위해 적용한 동적 클론 선택(Dynamic Clonal Selection) 알고리즘과 그의 문제점을 소개하고 보다 개선된 동적 클론 선택 알고리즘을 제안한다. 이전 연구에서 침입탐지시스템이 흔히 접하게 되는 상황, 즉 과거 안정적으로 관찰되었던 정상행위가 합법적인 요인들로 인하여 갑작스러운 변화를 보일 경우 과거 생성되었던 기억탐지자가 정상행위를 비정상행위로 오류 판단하는 것을 막기 위하여 인간면역시스템의 체세포 돌연변이 (somatic hypermutation)를 이용하여 유전자 라이브러리를 진화시키는 방법을 첨가한 동적 클론 선택 알고리즘을 소개한다.

Key words: 인공면역시스템, 동적 클론 선택 알고리즘, 체세포 돌연변이, 침입탐지시스템

1. 서 론

네트워크 정보 보안 시스템 기술 개발 노력중의 하나로 개발되고 있는 침입탐지시스템은 시스템의 불법적인 오용이나 남용을 탐지하는 시스템을 칭한다. 현재 사용되고 있는 침입탐지시스템은 대부분 이미 알려진 침입정보를 이용하는 것으로, 새로운 시스템의 허점을 이용한 알려지지 않은 침입에는 많은 허점을 드러내고 있다. 이러한 문제점을 극복하기 위한 국내외 연구 노력 중의 하나로, 외부에서 침입한 병원균을 효과적으로 탐지/파괴하는 인간의 면역 시스템을 응용하여 외부침입 탐지 시스템을 개발하는 연구들이 보고되고 있다 [1].

본 논문에서는 이러한 노력의 일환으로 소개된 동적 클론 선택 (Dynamic Clonal Selection) 알고리즘의 문제점으로 지적된 false-positive (FP) 오류를 감소시키기 위해 개선된 동적 클론 선택 알고리즘을 소개한다. Kim과 Bentley[3]는 Hofmeyr[2]의 인공면역시스템을 모델로 한 동적 클론 선택 알고리즘의 연구보고에서, 이 알고리즘이 지금껏 관찰되었던 정상행위가 합법적인 요인들로 인하여 갑작스러운 변화를 보일 경우, 이전에 생성되었던 기억 탐지자의 탐지결과는, 정상행위를 비정상행위로 판단하는 FP오류가 높게 나타나는 것을 지적하였다. 이러한 문제를 개선하기 위하여, 정상행위를 비정상행위로 오류 판단하는 기억 탐지자들을 제거하는 방법이 제안되었고 [4], 그 결과 동적 클론 선택 알고리즘은 약0.1% 미만의 낮은FP 오류율을 보였다.

그러나, 이러한 만족스러운 결과는 사용자의 참여에 의해서만 가능하였다. 정상/비정상행위 판별을 근간으로 하는 침입탐지시스템의 가장 심각한 문제는 높은 FP 오류율로 인한 사용자의 시스템에 대한 신뢰 상실인 것으로 분석된 바 [1], 제안된 인공면역시스템은 비정상행위를 탐지할 경우만이 사용자에게 최종탐지 결과를 알리는 것으로, 탐지자가 최종탐지결과를 사용자에게 보내는 횟수를 최소화시킬 때만이 개선된 알고리즘을 효과적으로 사용할 수 있다 하겠다.

이러한 문제점을 보완하기 위하여, 본 논문에서는 높은 TP탐지 율과 낮은 FP 오류율을 유지하면서, 동시에 탐지자가 최종탐지결과를 사용자에게 보내는 횟수를 줄일 수 있는 방안을 제안한다. 그러한 방안으로 인간면역시스템의 체세포 돌연변이(somatic hypermutation)를 이용하여 유전자 라이브러리를 진화시키는 방법을 첨가한 동적 클론 선

※ 본 논문은 한국과학재단 특정기초연구사업 (1999-2-511-001-3)의 지원으로 수행되었음
* 런던킹스컬리지
** 상명대

택 알고리즘을 소개한다. 소개된 방법은 기계학습 (Machine Learning) 연구에서 쓰이는 벤치마킹 데이터를 모의 적으로 선택하여 동적 클론 선택 알고리즘에 제공하는 방식으로 그 성능이 평가되었다.

2. 동적 클론 선택 (Dynamic Clonal Selection) 알고리즘

동적 클론 선택 (Dynamic Clonal Selection) 알고리즘은Hofmeyr[2] 인공면역시스템을 모델로 한 새로운 인공면역알고리즘으로, 세개의 다른 탐지자 개체군의 상호작용으로 인공면역반응을 생성한다. 이 알고리즘은 [3]에 자세히 기술되어 있으며, 본 논문에서는 그의 주요 운영 메커니즘만을 소개한다.

동적 클론 선택 알고리즘은 관찰대상의 비정상적인 개별행위를 탐지할 수 있는 탐지자 개체군을 지속적으로 생성, 갱신하는 것을 주 골자로 운영된다. 처음 무작위로 생성되는 미성숙 탐지자들은 음성선택(Negative Selection)을 통과하는 것으로 성숙탐지자가 된다. 음성선택이란 현재 알고리즘이 관찰하고자하는 어떤 행위에 대한 데이터를 항원으로 간주하여, 일정 자기내성 기간 (Tolerisation Period) 동안 관찰된 모든 정상항원 데이터에 대해 탐지 신호를 발산하지 않는 탐지자들만이 성숙탐지자로 변환되는 것을 허락하는 선택과정을 칭한다. 첫 번째 탐지자 생성 경로인 음성선택(Negative Selection)으로 인하여 탐지자들은 후에 정상행위에 대해 탐지신호를 보내지 않는 자기내성(Self Tolerance)을 갖게된다.

성숙탐지자들은 곧바로 관찰되는 항원 데이터에 대해 탐지과정을 시작한다. 이때 성숙탐지자들이 새로 관찰되는 항원데이터를 비정상행위로 간주하여 탐지 신호를 발산할 경우, 성숙탐지자의 변수인 탐지총합(Match Count)을 하나씩 증가시킨다. 따라서, 새로운 항원 데이터들이 각 성숙탐지자들에 의해 비정상행위로 간주될 때마다, 각 해당 성숙탐지자들의 탐지총합은 증가하게된다. 증가된 탐지총합이 사용자가 미리 정의한 면역 반응 임계값(Activation Threshold)이 되었을 경우, 성숙탐지자들은 최종 비정상행위 탐지 신호를 사용자에게 보내게 된다. 이는 성숙탐지자들이 음성선택을 통해 자기내성을 갖게되었지만, 자기내성기간동안 관찰된 정상행위가 관찰시스템이 보일 수 있는 모든 정상행위를 포함할 수 없기 때문에 초래되는 FP 오류를 줄이기 위한 방안이다.

사용자는 성숙탐지자가 보내온 탐지결과를 분석하여 그 결과가 정확하게 비정상행위를 탐지하였을 경우, 성숙탐지자를 기억탐지자로 변환시키어 새로운 항원데이터 관찰을 위하여 다시 탐지시스템에 보내게 된다. 기억탐지자들은 새로 관찰되는 항원데이타를 비정상행위로 간주할 때, 탐지총합의 증가 없이 곧바로 비정상행위 탐지신호를 사용자에게 보낸다. 이는 기억탐지자가 성숙탐지자와는 달리 이미 비정상행위를 탐지하여 그 유용성을 검증 받았음으로 성숙탐지자들에 비해 FP 오류가 낮을 것으로 기대되기 때문이다.

또 하나 주목하여야 할 점은 성숙탐지자들은 사용자가 미리 정의한 수명(Life Span)이 주어져 있어서, 만일 주어진 수명 기간이내에 그들의 탐지총합이 임계값을 만족시키지 못할 경우 바로 시스템에서 제거된다. 기억탐지자들은 이와는 달리 무한 수명을 가지고, 한번 생성된 경우 지속적으로 관찰되는 항원들에 대해 탐지활동을 벌인다.

따라서, 무작위로 생성된 하나의 탐지자는 일정 자기내성 기간동안 미성숙탐지자로 음성선택과정을 거친 후, 탐지총합이 면역 반응 임계값에 미치기까지 성숙탐지자로의 성숙기간을 마친다. 성숙기간을 마친 성숙탐지자는 사용자의 확인을 받고서는 기억탐지자로 변환되어 탐지과정을 시작하게 된다. 이러한 세단계 과정은 항원데이터가 제공되는 한 지속적으로 진행된다. 항원데이터가 제공되는 순간 동적 클론 선택 알고리즘은 우선 기억탐지자에 의해 비정상행위의 탐지를 시작하고, 아직 생성된 기억탐지자가 없을 경우엔 항원데이터는 성숙탐지자에게 제공된다. 이때의 항원데이터에 포함되어 있을 비정상행위에 의해 성숙탐지자의 성숙과정을 진행시킨다. 그러나, 생성된 성숙탐지자도 역시 없을 경우 항원데이타는 미성숙탐지자에게로 제공되어 음성선택에 쓰이게 된다. 따라서, 최초의 자기내성 기간동안은 시스템의 자기내성을 가질 수 있는 최초의 성숙탐지자 생성을 위한 훈련과정으로서, 비정상행위가 포함되어 있지 않은 항원데이타만을 제공하는 것을 가정한다.

따라서, 동적 클론 선택 알고리즘 진화와 학습의 한 세대는, 만족할만한 수의 최초의 성숙탐지자가 생성된 이후로는 이상에 서술된 순서, 즉 기억탐지자의 탐지, 성숙탐지자의 성숙, 미성숙탐지자의 내성으로 이루어진다.

2.1 개선된 동적 클론 선택 알고리즘

Kim과 Bentley [3]의 동적 클론 선택 알고리즘의 연구보고에서는 비성숙탐지자들이 음성선택 평가를 받게되는 성숙과정중 제공되는 항원 데이터가 각 세대마다 다른 항원 데이터를 포함하여, 전체 항원 데이터의 오직 일부분만이 제공될 경우, 적절한 자기내성 기간, 성숙탐지자의 반응 임계값과 수명을 부여하는 것으로 높은 탐지 율과 낮은 FP오류율을 보였다.

그러나, 지금까지 관찰되었던 정상행위가 합법적인 요인들로 인하여 갑작스러운 변화를 보일 경우, 기억탐지자들의 탐지결과는 심각한FP오류를 보이는 문제점 역시 발견되었다. 이는 동적 클론 선택 알고리즘이 음성선택을 이용하여 낮은FP오류의 보장을 의미하는 자기내성을 갖게되었던 까닭으로 설명된다. 비성숙탐지자들의 성숙 과정 중에 보였던 정상행위가 이후 탐지 과정 중에 어떤 합법적인 요인들로 인해 갑작스러운 변화를 보일 경우, 기존에 생성되었던 탐지자들은 새로운 정상행위에 대해 내성을 갖게되지 못하며, 이로 인하여 높은 FP오류를 보이는 것이다.

이러한 문제점을 개선하기 위하여 기억탐지자들을 선택적으로 제거하는 방법이 제안되었다 [4]. 기존의 동적 클론 선택 알고리즘에서는 기억탐지자들에게 무한 수명을 부여했었다. 그러나, 실제 인간 면역 시스템의 기억면역세포들은 고정된 수의 기억

면역 세포군을 유지하기는 하지만, 그를 구성하는 기억면역세포들은 지속적으로 생성되고 제거되는 것으로 알려져 있다 [7]. 따라서, 개선된 동적 클론 선택 알고리즘에서는 기억탐지자들의 탐지결과 또한 사용자가 분석하여 오직 그 결과가 비정상행위를 탐지할 경우에만 기억 탐지자들을 지속적으로 기억탐지자들로 남겨둔다. 이와는 반대로, 기억탐지자들이 정상행위를 탐지하는 오류를 범했을 경우에는 기억탐지자들은 바로 시스템에서 제거된다.

2.2 동적 클론 선택 알고리즘의 성능

동적 클론 선택 알고리즘을 침입탐지시스템에 사용하기 위하여 수행된 연구에 의하면 [3], 알고리즘의 성능은 세 가지 변수인 자기내성기간(T), 성숙탐지자의 면역반응임계값(A)과 수명(L)의 값에 의해 크게 좌우됨이 보고되었다. 또한, 개선된 알고리즘은 약0.1% 미만의 낮은FP 오류율을 보였으나, 이 경우 기억탐지자의 탐지결과를 분석하는 사용자의 참여가 필요하였다[4]. 따라서, 기억탐지자가 최종탐지결과를 사용자에게 보내는 횟수가 작을수록 제안된 알고리즘이 효과적일 것으로 기대되었다.

같은 연구에 의하면 [4], 성숙탐지자의 면역반응임계값을 증가시키는 것으로 기억탐지자가 최종탐지결과를 사용자에게 보내는 횟수를 작게 할 수 있었고, 이 경우 낮은FP 오류율은 보였으나, 만족스럽지 못한 낮은 True Positive(TP) 탐지 율을 보였다. 이와는 반대로 성숙탐지자의 면역반응임계값을 감소시키는 것으로 기억탐지자가 최종탐지결과를 사용자에게 보내는 횟수를 크게 할 경우엔 만족스러운True Positive(TP) 탐지 율을 보였다.

이러한 결과를 얻게된 이유는 처음 미성숙탐지자들은 난수발생을 통해 생성되었으므로, 음성선택을 통과한 기억탐지자들 각자가 탐지하는 비정상행위 패턴의 수는 거의 균등할 확률이 높다. 따라서, 더 많은 수의 탐지가가 기억탐지자로 살아남을 경우 시스템의 탐지 율이 높아지게 되지만. 더 많은 수의 탐지결과가 사용자에게 최종탐지확인을 위해 보내지게 된다.

3. 체세포 돌연변이 (Somatic Hypermutation)를 이용한 탐지자군의 진화

개선된 동적 클론 선택 알고리즘이 인간 면역시스템의 다양한 구성요소들을 구현하였으나, 유용한 모든 구성요소들을 구현한 것은 아니다. 그 중 하나로, 인간면역시스템은 기억 면역세포에 체세포 돌연변이를 적용하는 것으로 새로운 면역세포를 생성하고, 이렇게 생성된 면역세포는 보다 많은 수의 유해한 항원들을 탐지하는 것으로 알려져 있다 [7]. 즉, 기억 면역 세포의 체세포 돌연변이는 면역세포들이 현재 몸 안에 퍼져있는 유해한 항원들을 빠르게 탐지하도록 진화를 유도하는 것으로 알려져 있다 [7].

제2절에서 소개된 동적 클론 선택 알고리즘은 이러한 기억 탐지자들의 진화를 유도하는 요소를 갖고있지 않으며, 따라서 보다 많은 수의 기억탐지자가 생성되어 그 탐지횟수가 커져야 만이 전체 시스템의 TP 탐지 율이 높아지는 결과를 보였다. 다시 말해, 지난 세대에 기억탐지자의 항원탐지결과가 반영되어 새로운 탐지자를 생성할 수 있다면, 최근에 침입하는 항원들의 분포에 따라 탐지자군들은 진화할 수 있을 것이고, 따라서 적은 수의 기억탐지자로도 많은 수의 항원을 탐지할 수 있을 것으로 기대할 수 있다.

따라서, 이전에 동적 클론 선택 알고리즘이 미성숙탐지자들을 난수발생을 통해서만 생성했던 것에 반해, 본 논문에서 제안하는 동적 클론 선택 알고리즘은 이미 생성된 기억탐지자들에 돌연변이 오퍼레이터를 적용하여 미성숙탐지자들을 생성하는 것으로 기억 탐지자군의 진화를 유도할 수 있도록 한다. 특히, 기억 탐지자군이 현재 침입한 비정상 항원의 탐지를 하는 동시에 현재 존재하는 정상항원의 탐지를 피하는 방향으로 진화하기 위해서, 이전 세대에 정상항원의 탐지로 제거된 기억탐지자에 돌연변이 오퍼레이터를 적용하여 새로운 미성숙탐지자를 생성한다. 이전 세대에 정상항원의 탐지로 제거된 기억탐지자들은 만족할 만한 수의 비정상 항원들을 탐지하여 기억탐지자가 되었으나, 이후 정상항원을 탐지하는 실수를 범한 탐지자들이다. 따라서, 이들 탐지자들을 완전히 제거하기보다는 돌연변이 오퍼레이터를 통한 변이를 첨가하는 것으로, 여전히 만족할 만한 수의 비정상항원은 탐지할 수 있으나, 정상항원은 탐지하는 실수를 범하지 않은 탐지 자로 거듭날 수 있을 것이다.

이렇게 생성된 미성숙탐지자는 미성숙탐지자군에 첨가되고, 이들은 음성선택에 의해서 일부만이 성숙탐지자로 남게된다. 만일, 이미 생성된 기억탐지자가 없을 경우엔 이전의 동적 클론 선택 알고리즘처럼 미성숙탐지자는 난수발생을 통해서 생성된다.

동적 클론 선택 알고리즘의 돌연변이 오퍼레이터 적용을 통한 새로운 미성숙탐지자 생성에서 주목해야 할 점은 돌연변이 오퍼레이터의 적용 율이다. 동적 클론 선택 알고리즘과 유사한 진화개념을 사용하는 기존의 유전자 알고리즘이 돌연변이 오퍼레이터 적용을 통하여 새로운 개체를 생성할 경우, 그 적용 율이 0.01- 0.05%로 아주 낮은 반면 [6], 본 논문에서 동적 클론 선택 알고리즘이 적용하는 돌연변이 오퍼레이터의 적용률은 0.1 과 0.2%로 높은 편이다. 이 역시 인간면역시스템의 체세포 돌연변이를 그대로 따른 것으로, 인간면역시스템은 빠른 속도로 퍼져나가는 많은 수의 유해한 항원들을 보다 빠른 속도로 탐지해내기 위해, 높은 돌연변이 적용률을 사용하는 것으로 알려져있다. [7]

4. 데이터와 변수값 설정

실험은 UCI 기계학습 벤치마킹 데이터 모음 사이트에서 제공하는 위스콘신 유방암 데이터를 사용하였다 (ftp://ftp.ics.uci.edu/pub/machine-learning-databases). 이 데이터는 악성종양 (Malignant)과

양성종양 (Benign) 두 그룹으로 나뉘어지는데, 악성종양에 해당되는 데이터를 비정상행위로, 양성종양에 해당되는 데이터를 정상행위로 다루어 동적 클론 선택 알고리즘에 제공하였다. 악성종양의 경우 240개의 표본을 갖고 있으며, 양성종양의 경우 460개의 표본을 갖는다.

동적으로 변화하는 분포를 가지는 항원데이터를 자기내성기간과 성숙기간, 탐지기간 중에 제공하기 위해, 정상행위와 비정상행위에 해당하는 데이터를 클러스터링으로 잘 알려진 Expectation Maximization (EM) 알고리즘[4]을 이용하여 각각 3개의 부군(sub-groups)으로 나누었다. 동적 클론 선택이 진행되는 매N 세대동안 오직 세개의 부군중 하나의 부군에 해당하는 항원데이타들의 80%에 해당하는 비정상, 정상의 항원데이터만이 무작위로 선정되어 알고리즘에 제공되었다. N이 커질수록, 생성되는 탐지자들은 가장 최근에 선택된 항원부군에 해당하는 정상행위와 비정상행위만의 분포를 인식하게 된다. 따라서, 비교적 큰 값을 갖는N 세대이후 항원부군을 갑자기 대체하는 것으로 항원데이터의 분포를 바꾸고, 동적 클론 선택 알고리즘이 새롭게 생성하는 탐지자들이 얼마나 빠르게 새 항원부군의 정상행위와 비정상행위 데이터를 판별할 수 있는지를 관찰, 분석하는 것을 실험의 목표로 한다.

동적 클론 선택 알고리즘은 다양한 변수를 갖고 있는데, 그 변수 값에 따른 알고리즘의 성능은 [3]에 보고된바 있다. 본 논문의 실험에 쓰인 변수들의 값은 [3]에 보고된 실험결과에 따라 가장 적절한 값으로 선택되었으며, 그 값들은 표 1에 요약되어 있다.

<표 1> 동적 클론 선택 알고리즘에 쓰인 변수들의 값

변수	값
자기내성기간(T)	30
성숙탐지자의 수명(L)	10
성숙탐지자의 면역 반응 임계값	{10, 20, 40}
항원데이터가 동일 항원부군에서 지속적으로 선택되는 세대수 (N)	30

5. 실험결과

돌연변이 오퍼레이터를 적용하여 탐지자군 진화를 유도한 동적 클론 선택 알고리즘과 이러한 탐지자군의 진화를 유도하지 않은 알고리즘을 각각 같은 변수 값을 부여하여 수행하고 난후 그 결과들을 분석하였다. 각 변수선택 조합에 따른 하나의 실험은 총 2000세대동안 수행되었고, 각 실험을 5회 반복 수행한 평균 결과 값이 그림 1, 그림 2, 그림 3에 나타나있다.

그림 1은 2.1 개선된 동적 클론 선택 알고리즘에 소개된 알고리즘을 수행한 결과이고, 그림 2와 그림 3은 돌연변이 오퍼레이터를 적용하여 탐지자군의 진화를 유도한 이후의 동적 클론 선택 알고리즘의 결과이다. 이 경우, 돌연변이 오퍼레이터 적용율을 0.1 %와 0.2%로 달리하여 두 종류의 실험을 실시하였다. 각 그림에 있는 그래프들의 X축은 면역과정 진행 세대수를 나타내고, Y축은 탐지 율을 나타낸다. 특히 X축의 보조 선은 매 100세대마다 그려져있다.

돌연변이 오퍼레이터 적용에 따라 달라진 결과는 각 경우 관찰된 TP 탐지 율과 FP오류율에 따라 분석된다. 우선 돌연변이 오퍼레이터 적용된 결과를 보이는 그림2 와 3에서, 돌연변이 적용 율이 0.2%이고 A=5 인 경우를 제외하고는 모든 실험결과의 FP 오류율이 아주 낮은 것을 관찰할 수 있다.

특히 이 결과들에서 주목할 점은 돌연변이 오퍼레이터를 적용함에 따라, TP탐지 율이 상승하고 있으며, 또한 돌연변이 오퍼레이터 적용 율이 증가함에 따라 TP탐지 율이 크게 상승하고 있는 점이다. 특히, A값이 상대적으로 큰 경우 이러한 결과는 눈에 띄게 나타나는데, 예를 들어 A= 40 인 경우 돌연변이 오퍼레이터를 적용하지 않은 그림 1의 결과를 보면 TP 탐지 율이 50-90% 사이를 나타내고 있는 반면, 돌연변이 오퍼레이터를 적용률 0.2%을 적용하여 탐지자군의 진화를 유도한 경우의 A = 40 인 경우, TP 탐지 율이 85-95% 사이로 높게 나타나고 있다(그림 3). 또하나 중요한 결과로는 이러한 TP탐지 율의 증가가 FP오류율의 감소 없이 얻어졌다는 점이다. 이렇듯, 돌연변이 오퍼레이터를 통한 기억탐지자군의 진화가 인공면역시스템의 성능을 긍정적으로 향상시키는 역할을 했다는 것을 이상의 실험결과에서 확인할 수 있었다.

또한, 이전 돌연변이 오퍼레이터의 적용에 의한 탐지자군의 진화를 유도하지 않았을 경우, 낮은 FP 오류율과 높은 TP 탐지 율을 탐지자가 사용자에게 보내는 최종탐지결과 횟수를 증가시키는 것으로 얻을 수 있었고, 이를 개선하기 위해 돌연변이 오퍼레이터를 적용한 탐지자군의 진화를 유도하였다. 따라서, 그림 2과 3에서 나타난 결과가 탐지자가 사용자에게 보고하는 최종탐지결과 횟수를 증가시키지 않은 상태에서 얻을 수 있는 것인지를 알아보기 위해 동적 클론 선택 알고리즘의 수행 중 각 세대마다 생성된 기억탐지자의 수, 제거된 기억탐지자의 수와 제거되지 않고 살아남은 기억탐지자의 수를 분석해본다. 표 1, 2, 3에서는 돌연변이 오퍼레이터 적용을 하지 않은 경우와, 적용 율을 달리하여 돌연변이 오퍼레이터를 적용했을 경우 관찰된 결과가 각각 아래의 세표에 나타나 있다. 이 결과 역시5회 반복수행의 평균 결과 값이며 괄호 안의 값은 분산 값을 나타낸다.

<표 1> 각 세대마다 생성, 제거 그리고 잔존된 기억탐지자의 수와 사용자에게 보고된 최종탐지 결과 횟수. 탐지자군 진화가 유도되지 않은 경우.

돌연변이 오퍼레이터 적용이 없는 경우				
	생성	제거	잔존	최종탐지결과
A=10	124.25(50.7)	91.5(33.77)	32.75(18.25)	29.36(6.28)
A=20	78.75(5.62)	54.5(3.83)	24.25(14.25)	20.39(8.35)
A=40	55.25(4.09)	40.75(5.02)	14.5(1.67)	16.43(11.76)

<표 2> 각 세대마다 생성, 제거 그리고 잔존된 기억탐지자의 수와 사용자에게 보고된 최종 탐지 결과 횟수. 돌연변이 오퍼레이터 적용률 = 0.1%

돌연변이 오퍼레이터 적용률 = 0,1 %				
	생성	제거	잔존	최종탐지결과
A=10	376(1444.67)	339(1456.67)	37(4)	31.39(1.43)
A=20	259.5(176.3)	227(172)	32.5(7)	28.08(2.99)
A=40	203.5(149.5)	176(13778)	27.5(24.5)	22.56(6.66)

<표 3> 각 세대마다 생성, 제거 그리고 잔존된 기억탐지자의 수와 사용자에게 보고된 최종 탐지 결과 횟수. 돌연변이 오퍼레이터 적용률 = 0.2%

돌연변이 오퍼레이터 적용률 = 0.2 %				
	생성	제거	잔존	최종탐지결과
A=10	193.5(539)	160.8(393.6)	32.8(24.9)	27.52(14.62)
A=20	126.5(53.7)	97.5(67)	29(8.67)	24.48(6.94)
A=40	98(107)	78.5(101)	19.5(0.33)	16.75(1.12)

이상의 세 가지 다른 결과들에서 공통적으로 발견할 수 있는 점은 A가 가장 큰 경우인 40일 때, 최종탐지결과횟수가 가장 적은 것으로 나타나고 있다. 특히 이들 결과 중 돌연변이 오퍼레이터 적용율이 0.2%로 큰 경우에, 같은 A = 40이 주어진 다른 경우에 서 얻어진 최종탐지결과횟수들 보다는 더 작은 횟수를 보이는 것으로 관찰되고 있다. 앞서 보고된 실험결과에서, 이 경우 (적용 율이 0.2% 이고 A가 40인 경우)에 높은 TP 탐지 율과 낮은 FP 오류율을 보이고 있으며 (그림 3) 동시에 최종탐지결과횟수 또한 낮은 수준을 유지하고 있어 (표 3), 돌연변이 오퍼레이터를 적용한 탐지자군 진화의 유도가 동적 클론 선택 알고리즘을 침입탐지시스템에 사용될 수 있도록 개선시키고 있음을 보이고 있다.

6. 관련연구

본 논문에서 소개된 동적 클론 선택 알고리즘은 침입탐지시스템에게 특별히 요구되는 중요한 요소들을 갖추기 위해, 다양한 인간면역시스템의 메커니즘들을 구현시키고 있다, 동적 클론 선택 알고리즘이 구현하고 있는 인간면역시스템의 메커니즘들로는 미성숙, 성숙, 기억 세 종류 독립적으로 생성, 유지되는 탐지군들, 자기 내성 기간, 성숙탐지자 면역 반응 임계값, 성숙탐지자들의 수명, 체세포 돌연변이의 적용을 이용한 탐지자군의 진화가 바로 그것들이다.

이들 중 체세포 돌연변이의 적용을 이용한 탐지자군의 진화를 제외한 다른 메커니즘들은 이미 이전 Hofmeyr [2] 가 연구 발표한 인공면역시스템인 LYSIS에서 소개된바 있다. LYSIS 의 경우3900개의 정상 행위 스트링으로 구성된50일간 동안 모아진 네트워크 패킷의 헤더에서 비정상행위를 탐지하는 테스트를 하였다. 이 테스트를 위해 LYSIS는 50개의 다른 호스트들에서 각 100개의 미성숙탐지자를 생성하여, 총 5000개의 미성숙탐지자를 생성하여 그 성능을 평가하였다. 그러나, 이 평가에서 LYSIS는 침입탐지시스템이 다루어야 할 중요한 경우, 즉 지금껏 관찰되었던 정상행위가 합법적인 요인들로 인하여 갑작스러운 변화를 보일 경우에 대한 평가는 이루어지지 않았다. 그 대신 Hofmeyr [2] 는 관찰시간동안 정상행위의 일부분만을 나타내는 항원데이터만을 제공받게 되었을 때, LYSIS가 점진적으로 모든 정상행위를 학습할 수 있는가를 평가하였다. 이러한 점에서 본 논문에서 소개된 동적 클론 선택 알고리즘은 지금껏 관찰되었던 정상행위가 합법적인 요인들로 인하여 갑작스러운 변화를 보일 경우 대한 평가가 이루어졌고, 이러한 경우 나타나는 문제점들을 돌연변이 오퍼레이터를 이용한 탐지자군의 진화를 유도하여 해결하였다.

7. 결론 및 향후연구

이전 연구에서 [4], 침입탐지시스템이 흔히 접하게 되는 상황, 즉 과거 안정적으로 관찰되었던 정상행위가 합법적인 요인들로 인하여 갑작스러운 변화를 보일 경우 과거 생성되었던 기억탐지자가 정상행위를 비정상행위로 오류판단하는것을 막기 위하여, 동적 클론 선택은 기억탐지자의 탐지결과에 따라 기억 탐지자들을 제거하는 방법을 사용하였고, 그 결과 약0.1% 미만의 낮은FP 오류율을 보였다. 그러나, 이러한 만족스러운 결과는 사용자의 참여에 의해서만 가능하였고, 비정상행위를 탐지할 경우만이 사용자에게 최종탐지 결과를 알리는 것으로, 탐지자가 최종탐지결과를 사용자에게 보내는 횟수를 최소화시키는 경우엔 TP 탐지 율이 만족스럽지 못하게 나타났다.

이러한 문제점을 보완하기 위하여, 본 논문에서는 높은 TP탐지 율과 낮은 FP 오류율을 유지하면서, 동시에 탐지자가 최종탐지결과를 사용자에게 보내는 횟수를 줄일 수 있는 방안으로 인간면역시스템의 체세포 돌연변이를 이용하여 탐지자군을 진화시키는 방법을 이용한 동적 클론 선택 알고리즘을 소개하고 평가하였다. 알고리즘 평가를 위해, 과거 안정적으로 관찰되었던 정상행위가 합법적인 요인들로 인하여 갑작스러운 변화를 보일 경우 모의실험하였고, 그 탐지 율과 오류율을 분석하였다. 실험은 기계 학습 벤치마킹에 쓰이는 유방암의 데이터의 악성종양의 경우를 비정상행위로 양성종양의 경우를 정상행위로 간주하여 실시되었다.

본 논문에서 소개된 동적 클론 선택 알고리즘의 새로운 점은, 정상행위를 탐지하는 오류를 범하여 제거되는 기억탐지자에 대해 돌연변이 오퍼레이터를 적용하여 새로운 미성숙탐지자를 생성하는 것이다. 또한, 기존 유전자 알고리즘에서 흔히 돌연변이 오퍼레이터가 0.01-0.05% 내외의 아주 낮은 적용율에 따라 적용되었던 것에 반해 0.1%와 0.2%라는 높은 적용 율에 따라 돌연변이 오퍼레이터가 적용되었다.

실험결과, 성숙탐지자 면역 반응 임계값을 증가시켜 탐지자의 최종탐지횟수를 최소로 하는 경우에도, 0.2% 적용률로 돌연변이 오퍼레이터를 적용 미성숙탐지자를 생성하면, 5% 미만의 낮은 FP 오류율과 85-95% 의 높은TP 탐지 율을 보였다. 또한, 돌연변이 오퍼레이터 적용 율이 더 큰0.2%가 사용되었을 때가, 0.1%의 돌연변이 오퍼레이터 적용 율이 사용된 경우보다 더 작은 횟수의 최종탐지결과를 보였다.

이러한 결과에서 보듯. 기존의 돌연변이 오퍼레이터 적용 율보다 훨씬 큰 적용 율을 사용하여 탐지자군을 진화시키는 경우에 동적 클론 선택 알고리즘이 침입탐지 시스템에 사용되기 더 적합한 것으로 평가될 수 있다. 향후 연구로는 본 논문에서 연구된 동적 클론 선택 알고리즘을 실제 침입탐지 관련 데이터에 평가해 보는 것이고, 그와 관련된 연구가 현재 진행중이다.

8. 참고문헌

[1] Allen, J. et al, (2000), "State of the Practice of Intrusion Detection Technologies", Technical Report CMU/SEI-99-TR-028, Software Engineering Institute, Carnegie Mellon University.

[2] Hofmeyr, S., (1999) An Immunological Model of Distributed Detection and Its Application to Computer Security, PhD Thesis, Dept of Computer Science, University of New Mexico.

[3] Kim, J. and Bentley, P. (2002), "Towards an Artificial Immune System for Network Intrusion Detection: An Investigation of Dynamic Clonal Selection", the Congress on Evolutionary Computation (CEC-2002), Hawaii, May 12-17. to appear.

[4] 김정원, 최종욱, 김상진. (2002), "Toward "기억 탐지자의 제거를 통한 동적클론선택 알고리즘의 개선", 정보처리학회, 제9권 제1호 pp923-926, 2002.

[5] Mitchell, T. (1997), Machine Learning, McGraw-Hill. Paul, W. E., (1993), "The Immune System: An Introduction", in Fundamental Immunology 3rd Ed., W. E. Paul (Ed), Raven Press Ltd.

[6] Mitchell, M. (1996), An Introduction to Genetic Algorithm, MIT Press.

[7] Tizard, I. R., (1995), Immunology: Introduction, 4th Ed, Saunders College Publish