퍼지 및 지능시스템학회 논문지 2002, Vol. 12, No. 1, pp. 52-60

컴퓨터 면역시스템 개발을 위한 인공면역계의 모델링과 자기인식 알고리즘

Modelling of Artificial Immune System for Development of Computer Immune system and Self Recognition Algorithm

심귀보* · 서동일**, 김대수*** · 임기욱****

Kwee-Bo Sim, Dae-Su Kim, Dong-Il Seo, and Kee-WooK Rim

* 중앙대학교 전자전기공학부

** 한국전자통신연구원 정보보호기술 연구본부

*** 한신대학교 컴퓨터학과

**** 선문대학교 산업공학과

요 약

최근 컴퓨터의 사용이 보편화되면서 악의적 사용자에 의해 발생하는 컴퓨터 바이러스와 해킹에 의한 피해가 급속히 증가하고 있다. 남의 컴퓨터에 침입하는 해킹이나 데이터를 파괴하는 컴퓨터 바이러스에 의한 피해를 막기 위해 최근에 생명체의 면역시스템의 특징을 이용해 인공면역계를 구성해 시스템 침입탐지와 바이러스 탐지 및 치료에 대한 연구가 활발히 진행 중에 있다. 생체 면역계는 외부에서 침입해 세포나 장기에 피해를 주는 물질인 항원을 스스로 자기세포와 구분해 인식·제거하는 기능이 있다. 이러한 면역계의 특징인 항원을 인식하는 기능은 자기세포의 확실한 인식을 가지고 있는 상태에서 다른 물질을 구분하는 자기·비자기 인식방법으로 볼 수 있다. 본 논문에서는 생체 면역계에서 세포독성 T세포의 생성과정의 하나인 Negative 및 Positive Selection을 모델링하여 침입에 의한 데이터 변경과 바이러스에 의한 데이터 감염 등을 탐지할 때 가장 중요한 요소인 자기 인식 알고리즘을 구현한다. 제안한 알고리즘은 큰 파일에서의 Detection을 구성하기 용이한 점을 가지며 국소(cell)변경과 블록(string)변경에 대한 자기인식률을 통해 알고리즘의 유효성을 검증한다.

Abstract

According as many people use a computer newly, damage of computer virus and hacking is rapidly increasing by the crucial users. A computer virus is one of program in computer and has abilities of self reproduction and destruction like a virus of biology. And hacking is to rob a person's data in a intruded computer and to delete data in a person's computer from the outside. To block hacking that is intrusion of a person's computer and the computer virus that destroys data, a study for intrusion detection of system and virus detection using a biological immune system is in progress. In this paper, we make a model of positive and negative selection for self recognition which have a similar function like T-cytotoxic cell that plays an important role in biological immune system. We embody a self-nonself distinction algorithm in computer, which is an important part when we detect an infected data by computer virus and a modified data by intrusion from the outside. And we showed the validity and effectiveness of the proposed self recognition algorithm by computer simulation about various infected data obtained from the cell change and string change in the self file.

Key Words : 인공면역계(AIS), 생체면역계(BIS), 자기인식 알고리즘(Self-Recognition Algorithm), MHC

1. 서 론

최근 컴퓨터의 사용이 보편화되면서 사회의 정보화 또는 컴퓨터화란 말이 나오고 있을 정도로, 컴퓨터는 기업활동은 물론 시민생활과 여러 분야에서 중요한 역할을 수행하고 있다. 이러한 많은 컴퓨터의 사용과 더불어 악의적 사용자에 의해 발생하는 컴퓨터 바이러스와 해킹의 피해가 급속히 증가하고 있다. 컴퓨터 바이러스는 바이러스가 생체에 침입하여 병을 일으키는 것처럼 컴퓨터 내에 침입하여 자료를 손상시키거나 다른 프로그램들을 파괴해 컴퓨터의 작동을 방해하고 자신을 복제하여 다른 컴퓨터에 전염시키는 프로그램의 한 종류이다. 해킹은 다른 사람의 컴퓨터에 침입하여 데이터를 빼내거나 파괴하는 행위로 컴퓨터 통신망이 발전하면서 전 세계의 컴퓨터들이 통신망에 연결된 인터넷의 활용이 많아지면서

접수일자 : 2001년 11월 1일
완료일자 : 2002년 1월 31일
본 연구는 한국전자통신연구원의 인공면역기반 차세대 인터넷 보안기술 개발의 용역으로 수행되었습니다. 연구비 지원에 감사드립니다.

해커들에 의한 피해가 확산되고 있다. 이렇게 다른 사람의 컴퓨터에 침입하는 해킹이나 데이터를 파괴하는 컴퓨터 바이러스 등에 의한 피해를 막기 위해 최근에 생명체의 면역시스템의 특징을 이용한 시스템 침입탐지[1-6]와 바이러스 탐지[7-8] 및 치료에 대한 연구가 활발히 진행중에 있다.

생체의 면역계는 외부에서 침입해 세포나 장기에 피해를 주는 물질인 항원을 스스로 자기세포와 구분해 인식하고 제거하는 기능을 가지고 있다. 생체 면역계의 특징 중의 하나인 항원을 인식하는 기능은 자기세포의 확실한 인식을 가지고 있는 상태에서 자기세포와 다름으로 구분되는 물질을 분류하는 자기/비자기(self/non-self) 인식방법으로 볼 수 있다. 이러한 기능을 가장 잘 보여주는 면역 T세포 중의 하나인 세포독성 T세포(T-cytotoxic Cell)는 자기세포를 인식하는 부분과 항원으로 인식하는 부분으로 구성되어 항원에 의해 감염된 자기세포를 찾아 제거하는 역할을 한다[9-10]. 이러한 생체 면역계의 특성을 모델링하여 생체 면역계와 같이 외부에서 침입해 자신에게 피해를 주는 것을 방어하며 제거하는 시스템을 컴퓨터 면역 시스템[11-12]이라고 한다.

본 논문에서는 생체 면역계의 자기 세포와 항원을 구별하여 인식하는 방법을 모델링한 자기-인식 알고리즘을 제안한다. 이와 같이 생체 면역계의 자기-인식 특성을 모델링한 사례로는 D. Dasgupta와 S. Forrest의 Anomaly Detection Algorithm[7,13-14]이 있다. 이 알고리즘은 면역세포의 생성 과정 중의 하나인 Negative selection을 이용하여 Anomaly detector를 구성하고 이를 자기-인식 알고리즘에 적용하였다. 제안되었던 알고리즘은 자기 공간의 변경에 대한 인식과 추가되는 자기 공간에 대한 인식에 좋은 반면 자기 공간의 삭제에 대한 인식률[7]과 블록단위의 변경에 대해 인식률이 떨어지는 단점이 있다. 본 논문에서는 생명체의 면역세포의 생성 과정 중의 하나인 Positive Selection을 모델링하여 자기-인식 알고리즘을 제안한다. 제안된 자기-인식 알고리즘은 자기 공간의 변경에 대한 인식과 자기 공간의 삭제에 대한 인식률, 그리고 블록 단위의 변경에 대한 인식률이 좋은 특성을 가진다. 또한 이 두 가지의 자기-인식 알고리즘을 사용할 때의 자기-인식률의 향상을 가져온다. 이를 이용하여 침입에 의한 데이터 변경과 바이러스에 의한 데이터 감염 등을 탐지할 때 가장 중요한 요소인 자기-인식 알고리즘을 구현하였다. 제안한 알고리즘은 자기 공간의 국소 변경과 블록 변경에 대한 자기인식률을 통해 알고리즘의 유효성을 검증한다.

2. 생체 면역시스템

생명체의 방어체계인 면역계는 박테리아, 기생균, 병원균, 독소, 바이러스 등과 같이 항원이라고 통칭하는 매우 다양한 외부 유기체나 단백질에 대하여 생명체의 세포와 장기를 방어할 수 있는 매우 정교하고 복잡한 시스템이며 개체를 건전한 상태로 유지시키기 위해 반드시 필요한 기능이다. 또한 면역계는 virus 감염과 종양발생에 의해 변이한 자기세포를 배제하는 작용도 가지고 있다. 이러한 생명체의 면역계는 중앙처리장치인 뇌의 명령에 따르는 것이 아닌 각 요소의 자율적인 행동이 유기적으로 결합되어 형성된 자율분산시스템으로 항원을 인식하는 기능, 정보처리 기능, 학습 및 기억능력, 자기와 비자기의 구별능력, 분산시스템으로서 전체의 조화를 유지하는 능력 등을 가지고 있다.

2.1 생체 면역계 요소

생체 면역계(biological immune system : BIS)에는 복잡한 면역 반응을 구현하는 많은 면역 세포들이 유기적으로 결합하여 하나의 시스템을 형성하고 있다. 이렇게 복잡하게 형성된 면역계를 구성하는 기본요소는 두 가지 형태의 림프구로 각각 B세포와 T세포이다. B세포는 항원을 죽이는 항체를 생산, 분비하는 체액성 반응(humoral response)을 하며, T세포는 면역에 관련된 세포를 자극 또는 억제하거나 항원에 의해 감염된 자기세포를 죽이는 세포성 반응(cell-mediated response)을 주로 담당한다. 그 밖의 다른 면역 세포들도 직·간접으로 면역반응에 영향을 준다. 다음은 면역반응에 관여하는 요소들이다.

▪ 항원(Antigne)

항원 혹은 면역원(immumogen)은 단백질, 핵산, 탄수화물 등이며 일반적으로 분자량이 5000 이상인 것으로 미생물을 포함하여 각종 생물의 구성성분 혹은 산물이다. 생체에 침입하여 세포나 장기에 피해를 주어 병을 형성한다. 항원에는 항체에 결합하여 반응을 자극하는 특수한 부위가 있는데, 이를 항원결정소(antigenic determinant)라 한다.

▪ 항체(Antibody)

항체 혹은 면역글로불린(immunoglobulin)이라 불리는 물질은 B세포에서 생산되는 특수한 단백질로서 외부에서 들어온 항원이라 불리는 이물질에 대한 반응으로 생성된다. 항체는 항원을 제거하는데 사용되도록 구조가 상당히 특수화되어 있고, 그러한 다양성은 수많은 종류의 항원에 반응할 수 있게 한다.

▪ 대식세포(Macrophage)

대식세포로 불리며 하는 역할은 상처부위로 모여들어 박테리아, 침입 생물, 세포 부스러기, 일부 유리된 항원들을 삼키고 파괴하는 동시에 항원에 대한 정보를 수집한 후 T세포 중에서 그 정보에 적절한 세포를 찾아 자극하는 것이다. 비특성 방어에 담당하면서 특성방어에 중요한 역할을 한다.

▪ T세포(T-cell)

특성방어에서 중요한 역할을 수행하는 면역세포로서 다른 면역세포의 자극 및 억제와 감염된 자기 세포의 제거 등 세포성 반응을 주로 한다. 3가지의 종류로 구분되고 있으며 각각 보조 T세포(T-helper cell), 세포독성 T세포(T-cytotoxic cell) 그리고 억제 T세포(T-suppressor cell)이다. 보조 T세포는 대식세포로부터 전달받은 항원의 특성을 B세포와 세포독성 T세포에 전달하는 역할을 한다. 세포독성 T세포는 항원에 의해 감염된 자기세포를 제거하는 역할을 한다. 억제 T세포는 면역반응에 의해 항원의 수가 감소하면 면역반응을 억제하는 역할을 한다.

▪ B세포(B-cell)

특성방어에서 T세포와 더불어 중요한 역할을 수행하는 면역세포이다. 보조 T세포로부터 전달받은 항원의 특성에 맞는 항체를 생산하는 역할을 한다.

이와 같은 면역세포 각각의 활동이 유기적으로 결합

하여 하나의 면역반응을 형성해 외부에서 침입한 항원에 대해 효과적으로 대체하여 생체를 보호하고 있다.

2.2 면역 반응

면역계는 크게 비특성적 방어(nonspecific defense)와 특성적 방어(specific defense)의 두 종류가 있다. 비특성적 방어는 직접적이며 즉각적인 방어 방법으로 화학물질과 특정한 백혈구들이 사용되는데, 이러한 방어물질들은 항상 대기 상태에 있어 언제라도 작용할 수 있다. 반면에 특성적 방어는 좀더 복잡하고 방어물질들이 준비되는 데 일정한 시간이 필요하다.

비특성적 방어는 식세포들에 의해 박테리아와 망가진 세포 조각들을 삼키는 작용으로 항원의 특성에 상관없이 작용을 한다. 이러한 역할을 하는 면역세포 중에 대식세포가 있다. 대식세포는 주로 침입 세포와 세포 조각을 제거하는 일을 하며, 이 과정에서 항원의 특정부위인 항원결정소(Antigenic Determinant)를 수집하여 보조 T세포에게 전달하는 항원제공세포(Antigen-Presenting Cell, APC)로서의 역할을 한다.

특성 방어는 면역반응(Immune response)을 지칭한다. 이러한 방어는 침입 물질이 침투한 후에야 그 특정 침입자에 대한 특정한 방어 능력이 갖추어지기 때문에 반응이 즉각적이지 못하고 상당한 시간이 소요된다. 이 면역반응에 작용하는 면역세포가 B세포와 T세포이다. 대식세포로부터 항원결정소를 전달받은 보조 T세포는 항원의 특성을 가지는 세포독성 T세포(T-cytotoxic Cell)와 B세포(B-Cell)를 찾아 각각의 세포를 자극한다. 이때 각 면역 세포들은 다양한 종류의 항원에 작용할 수 있도록 다양하게 분화된 상태이다. 그러나 각 항원에 작용할 수 있는 세포의 숫자가 적기 때문에 효과적으로 방어를 할 수 없다. 따라서 자극 받은 T세포와 B세포와 같이 필요한 면역세포들만을 선택하여 빠른 속도로 발달시키는데, 이러한 과정을 클론선택(clonal selection)이라 한다.

클론선택 되어진 면역세포는 항원의 특성을 가지고 있기 때문에 항원을 제거할 수 있는 능력을 가지고 있다. 선택된 세포독성 T세포는 이중 수용체(dual receptor)를 사용해 정상적인 세포들을 공격하기 시작한다. 만일 그 세포에 항원이 없으면 그냥 지나치지만, 세포 표면에 바이러스가 있어서 MHC 단백질과 항원이 모두 들어맞으면 바이러스에 감염된 세포로 판단해 죽인다. 선택된 B세포는 활성화되어 커다란 클론이 형성하게 된다. 이러한 B세포의 클론은 커지면서 형질세포(plasma cell)와 기억 B세포(memory B-cell)로 된다. 형질세포는 항원에 대해 특이성을 갖는 엄청난 양의 항체를 생산하고 분비하게 되며 생산된 항체는 혈액과 체액을 순환하면서 자신에 들어맞는 항원과 결합해 항원의 제거를 도와주게 된다. 기억 B세포는 이 후의 같은 항원의 재차 침입을 위해 항원의 항원결정소를 기억하게 된다.

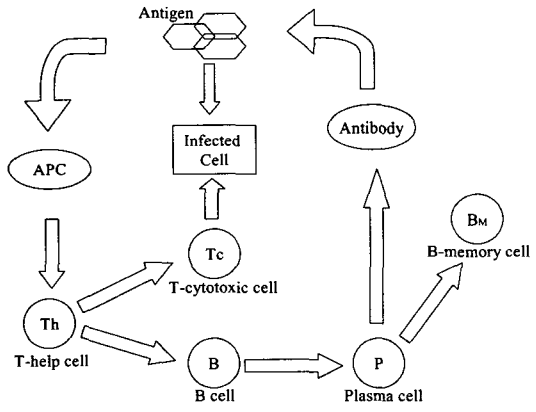

그림 1. 생체 면역시스템의 상호작용도
Fig. 1. Interaction of biological immune system

항원이 감소하기 시작하면 T세포의 일종인 억제 T세포(T-suppressor Cell)에 의해 B세포와 T세포의 활동이 억제되어 면역세포의 수가 감소하게 되며 생체는 정상상태를 유지하게 된다. 이러한 일련의 과정이 면역반응이다. 그림 1은 생체면역시스템의 상호작용을 보여주고 있다.

2.3 면역세포 형성 원리

면역 세포가 외부에서 침입한 항원을 제거하는 면역반응을 정상적으로 수행하기 위해서 각 면역세포들은 2가지의 요소에 의존하게 된다. 하나는 각각의 세포사이의 협력과 공조이다. 또 다른 하나는 항원의 인지 능력과 구별 능력이다. 면역 세포의 항원을 인지하는 능력은 자기 세포와 구별되는 항원을 구별하고 이의 항원결정소의 특성을 가지고 있는 면역 세포를 통해 항원을 제거하는 면역 반응을 일으키는 가장 중요한 능력인 것이다.

면역 세포가 자기 세포를 인지하는 방법으로는 MHC 단백질을 이용한다. 개체에는 각각 개인적인 특징을 이루는 단백질이 존재하며, 단백질을 생성하는 유전들을 주조직 적합성 복합체(major histocompatibility complex, MHC)라 하며, 이렇게 생성된 단백질을 MHC 단백질이라고 한다. 이 MHC 단백질을 인식하는 부분을 면역세포에 존재하며 이를 이용해 자신의 세포인지를 판단하게 된다. B세포나 T세포와 같이 특정 항원에 대해 적용되는 면역 세포는 생성될 때 다양한 항원들의 특성에 부합되는 부분이 존재하며 이를 항원수용체(Antigen Receptor)라 한다. 항원수용체는 면역 세포가 생성될 때 유전자의 돌연변이 및 교차를 이용하여 다양성을 내포하며 생성된다.

자기를 판별해주는 MHC 단백질을 인식하는 부분과 항원의 종류를 판별하는 항원수용체의 특성을 지니는 대표적인 면역 세포는 세포독성 T세포이다. 세포독성 T세포는 항원에 감염된 자기 세포를 제거하는 역할로 먼저 자기 세포인지를 판별하고 자기 세포에 항원이 존재하는가를 검사하므로 이 두 가지의 인식부를 모두 가지고 있다. 이러한 T세포의 인식부를 T세포 수용체(T-cell receptor)라고 한다. T세포 수용체가 면역계에서 정상적으로 동작되지 않으면 자기 세포를 항원으로 인식하게 되어 공격하게 된다. 따라서 면역계는 면역 세포 초기 생성시 MHC 인식부와 항원수용체의 정상적인 동작여부를 확인하면서 면역 세포를 생성하여 면역계를 구성한다. 수용체의 정상적인 동작여부를 가리는 방법으로 사용되는 것이 Positive Selection과 Negative Selection이다.

Positive Selection은 각 면역세포의 MHC 인식기능을 확인하는 선택방법이다. 자기세포에서 분비되는 MHC 단백질을 정확히 인지할 수 있는 면역세포만이 사용가능하기 때문에 갓 생성된 면역세포에 MHC 단백질을 결합시켜 긍정적인 선택이 되는 세포들로만 면역 세포를 구성하게 되며 선택되지 않은 면역 세포들은 자기 세포를 인지하지 못하는 것이므로 제거 또는 재배열 등의 방법

을 사용하여 면역계를 유지한다.

Negative Selection은 항원의 인식에 있어서 자기를 항원으로 인식하는 것을 배제하기 위해 방법이다. 항원수용체가 MHC 단백질을 항원으로 인식하면 모든 자기 세포를 항원으로 인식하게 된다. 때문에 항원으로 MHC 단백질을 인식하지 못하게 하기 위해 면역세포에 MHC 단백질을 결합시켰을 때 항원수용체가 부정적인 선택을 하는 세포만으로 구성된다. 이때 긍정적인 선택을 하는 면역세포는 MHC 단백질을 항원으로 인식하는 세포들이므로 죽이거나 다시 항원수용체를 형성하는 단계를 거치게 된다.

이 두 가지 선택을 거친 면역세포는 MHC 단백질을 자신으로 인식하면서 이를 항원으로 인식하지 못하게 구성되어 생명체에서 정상적인 면역반응을 형성한다. 그림 1은 생체 면역계에서 정상적인 면역 세포의 형성과정을 보여주고 있다.

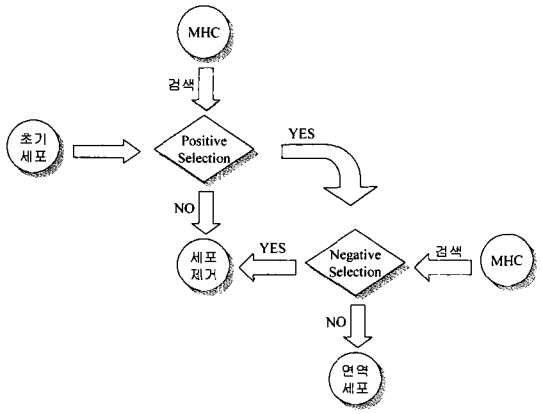

그림 2. 생체 면역계(BIS)의 정상 면역세포 형성 과정
Fig 2. Mechanism of immune cell in BIS

3. 자기 인식 알고리즘

생체 면역계의 중요한 특징 중의 하나가 자기 인식으로서 자기 세포를 인식하면서 자기 세포와 다른 항원을 인식함으로서 자기(Self)와 비자기(Non-self)를 구별하는 것이다. 자기를 확실하게 구분함으로서 자기와 다른 존재를 인식하는 방법이다. 이와 같은 자기와 비자기를 인식하는 방법을 구현한 알고리즘을 자기-인식 알고리즘이라 한다. 특히 생체 면역계의 특징을 이용하여 자기-인식 알고리즘을 구현한 경우가 있다. D. Dasgupta와 S. Forrest는 생체 면역계의 면역 세포의 생성원리 중의 Negative Selection을 이용한 Anomaly Detection Algorithm[7,13-14]으로 자기-인식 알고리즘을 구현하였다. 이 알고리즘은 자기로 보호해야 하는 공간의 변경 및 추가에 대해 강한 자기-인식률을 보인다. 하지만 자기 공간 제거에 따른 자기 인식[7], 블록화된 변경에서의 자기-인식률 저하 등이 나타났다. 이에 본 논문에서는 생체 면역계의 Positive Selection을 이용하여 보안된 알고리즘을 제안한다.

3.1 Anomaly Detection Algorithm

Anomaly Detection algorithm[7,13-14]은 D. Dasgupta 교수와 S. Forrest교수에 의해 제안된 자기-인식 알고리즘의 방법이다. 이는 자기 공간에 대해서 Negative Selection을 매칭해 detector set을 구성하고 이를 자기-인식 알고리즘에 적용하였다. 이 anomaly detector를 이용한 자기-인식 알고리즘은 확률적인 인식률을 나타내며 자기 공간의 변경의 인식률과 추가된 공간의 자기 여부를 판단하는데 유용하다[7].

이 알고리즘은 크게 두 부분으로 구성된다. 하나는 자기 공간을 검사하기 위한 anomaly detector를 구성하는 부분이며 다른 하나는 구성된 anomaly detector를 이용하여 자기 공간을 모니터하며 변화의 발생을 검사하는 부분이다.

Anomaly detector는 자기 공간에 매칭되지 않는 스트링을 이용하여 구성한다. 스트링은 detector의 길이이며 셀의 집합이다. 셀은 하나의 심벌을 가지는 공간으로 정의한다. 먼저 알고리즘은 자기 공간을 인식되고 보호되어야 하는 자기 스트링의 집합 S을 만든다. 두 번째 단계는 랜덤 스트링의 집합 R_0 을 만들어 S의 스트링에 대해 매칭 검사를 한다. 이때 매칭되는 R_0 의 스트링은 제거하며 S의 어떤 스트링과도 매칭되지 않는 스트링들로 anomaly detector 집합 R을 만들며 이 단계를 censoring이라고 부른다[7].

이때 같은 길이의 두 개의 스트링 사이의 완벽한 매칭이라는 것은 스트링의 각 위치에 자리하고 있는 각각의 셀의 심벌이 완전히 동일함을 의미한다. 이러한 매칭은 자기 공간의 스트링이 어느 정도 이상 커지면 매칭되지 않는 스트링들을 찾는 것이 어려워지기 때문에 부분적 매칭 룰을 이용하고 있다. Anomaly detection algorithm에서 사용한 매칭 룰은 인접한 위치 안의 r개 셀들의 매칭됨을 확인하는 방법이다. 그래서, 어떤 두 스트링에 대해, 적어도 r개 인접한 셀들에서 같은 셀들이 존재한다면 매칭되었다고 정의한다.

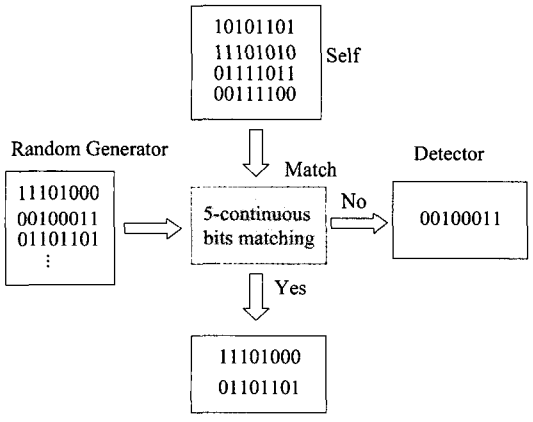

그림 3. Anomaly Detector의 구성 방법[7]
Fig. 3. Construction method of anomaly detector

위와 같은 방법으로 구성된 detector 집합을 이용하여 자기-인식 알고리즘의 구현은 다음과 같다. 구성된 detector 집합은 자기 공간의 속하지 않는 스트링들로 구성되어 있다. 따라서 자기 공간의 상태는 detector 집합 R에 의해 매칭되는 것으로 알 수 있다. 자기 공간 S와

R로부터 하나의 스트링을 선택하고 이 스트링들이 매칭되는 지를 검사함으로서 자기와 비자기를 인식한다. 이때 detector들은 그들이 만들어진 매칭 방법으로 검사된다. 그림 3는 Anomaly detector를 구성하는 방법의 예로서 연속적으로 5개의 셀이 같음을 매칭으로 정의하여 구성한다.

3.2. MHC set을 이용한 자기 인식 알고리즘

본 논문에서는 면역 세포에서 MHC 유전자가 만들어내는 MHC 단백질을 인식해 자기가 아닌 물질에 대해서 자기를 구분하는 방법을 이용하여 자기-인식 알고리즘을 구현하였다. 구현한 자기-인식 알고리즘은 자기 공간에서 자기로 인식할 특징점이나 일정공간을 가지는 MHC를 생성하고 이를 자기 공간의 MHC Set으로 정한다. 이러한 MHC Set은 자기 공간의 특징을 가지고 있으므로 이를 기반해 자기가 아닌 다른 것을 구별해 낼 수 있다. 자기 공간의 특징으로 인식부를 형성하는 방법이 초기 면역세포의 생성시 MHC 인식부를 검사, 형성해 주는 Positive Selection의 방법인 것이다. 자기 공간의 검사는 검사 공간에서 MHC Set을 가지고 있으면 자기 공간으로 인식하며 그러한 MHC Set을 가지고 있지 않으면 비자기 공간으로 인식한다. 따라서 이 방법에 의한 자기-인식 알고리즘은 자기 공간의 변경과 삭제에 대해서 자기를 판별하는 알고리즘으로 작용한다.

MHC는 자기 공간으로 설정된 스트링의 집합 S에서 생성한다. 이러한 스트링은 일정 길이의 셀로 구성되며 셀은 정해진 알파벳 심벌들 중 하나의 값을 가진다. MHC set은 하나의 스트링의 길이를 가지는 MHC들로 정의한다. 생성과 매칭의 기본 단위인 코드는 일정한 개수 r개의 연속적인 셀들을 의미하며 MHC는 교차되는 코드들을 이용하여 생성한다. 그림 4는 스트링의 구성요소를 보여주고 있다.

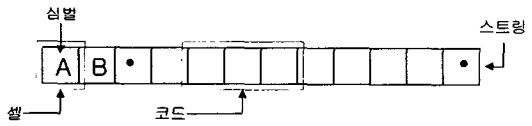

그림 4. MHC set 스트링의 구성 요소
Fig. 4. Unit of MHC set string

MHC set을 구성하는 알고리즘은 다음과 같다.

STEP 1. 스트링들로 구성된 자기 공간 S를 설정한다. 자기 공간은 자기-인식 알고리즘으로 매칭을 실행했을 때 자기로 인식되는 공간을 의미한다. 이렇게 구성된 자기 공간은 스트링의 위치 값을 가지는 코드들을 생성할 수 있다.

STEP 2. 자기 공간 스트링의 첫 번째 위치에 자리 잡은 코드들의 집합에서 중복되지 않는 코드를 랜덤하게 선택한다. 선택된 코드는 하나의 MHC 스트링의 첫 번째 위치에 자리 잡는 코드가 되며 MHC를 구성하는 시드가 된다.

STEP 3. 자기 공간 스트링의 두 번째 위치에 자리 잡은 코드들의 집합에서 첫 번째 위치의 코드와 동일한 부분을 가지는 코드를 선택한다. 선택된 코드는 MHC 스트링의 두 번째 위치에 자리 잡는 코드가 된다.

STEP 4. 스트링의 각 위치에 자리 잡은 코드들의 집합에서 앞 위치의 코드와 동일한 부분을 가지는 코드를 선택하여 각 코드의 위치에 자리 잡아 하나의 스트링을 구성하고 이를 MHC 스트링으로 설정한다.

STEP 5. 위의 2,3,4의 방법을 반복하여 일정 개수의 MHC set을 설정한다.

위와 같은 알고리즘으로 MHC set을 구성한다. 그림 5는 코드를 3으로 설정하여 MHC set을 구성하는 방법을 보여주고 있다.

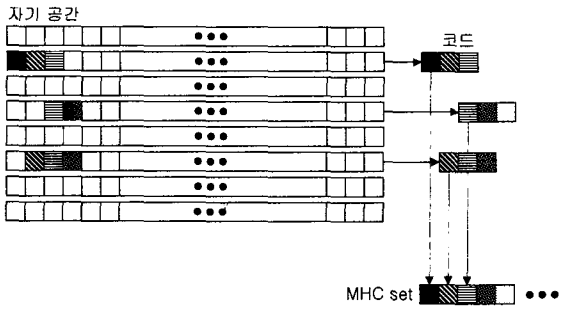

그림 5. MHC set의 구성 방법의 예
Fig. 5. Example of constructing MHC set

자기-인식 검사는 구성한 MHC set을 이용하여 이루어진다. 각의 MHC set의 코드들이 자기 공간 스트링의 위치에 존재하는가를 검사한다. 하나의 MHC 스트링을 이루고 있는 모든 코드가 검사 공간 스트링의 각 위치 값에 존재하면 해당 MHC 스트링에 의해서는 검사 공간을 자기로 인식하고 그렇지 않으면 비자기로 인식한다. 또한 모든 MHC set이 검사 공간을 자기로 인식하면 검사 공간을 자기 공간으로 인식한다. 그림 6은 MHC set을 이용한 검사 공간 검색 과정이다.

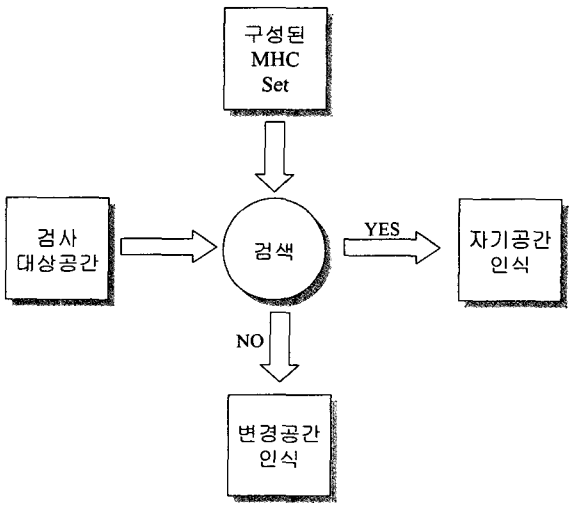

그림 6. MHC set을 이용한 자기/비자기 판단 과정
Fig. 6. Process of self-recognition algorithm using MHC set

3.3. 두 선택 법을 이용한 자기 인식 알고리즘

모델링된 면역세포는 위의 방법의 Positive selection과 Negative selection을 이용하여 정상적인 면역세포를 찾는다. 이러한 방법을 이용하여 Negative selection을 이용한 Anomaly detection algorithm과 Positive slection을 이용한 MHC set algorithm을 복합적으로 사용하여 인식부를 구성하는 자기 인식 알고리즘을 제안한다. 이는 Anomaly detection algorithm의 자기 공간의 삭제에 대한 자기 인식과 MHC set의 자기 공간 추가에 대한 자기 인식의 영역을 서로 보완해주며 제안된 MHC set algorithm의 자기 공간변경 방법에 따른 자기 인식률의 특성을 가지고 있다. 그림 7은 두 가지 선택 법을 모두 이용한 자기 인식 과정을 보여준다.

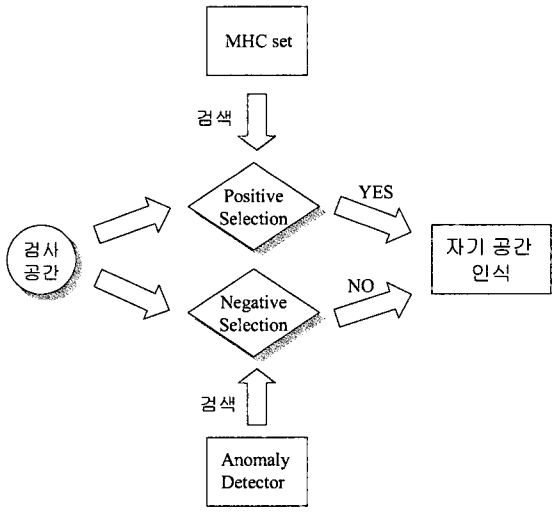

그림 7. 두 가지 선택 법을 이용한 자기 인식 과정
Fig. 7. Process of self-recognition using two-methods

4. 시뮬레이션

생체 면역계의 중요한 특징 중에 하나인 자기 인식을 모델링하여 구현한 자기-인식 알고리즘의 유효성을 검증하기 위해 기존에 S. Forrest가 제안한 Anomaly detection 알고리즘과 본 논문에서 제안한 MHC set을 이용한 자기-인식 알고리즘의 자기-인식에 대해 시뮬레이션하였다. 시뮬레이션은 컴퓨터에서 자기 공간을 형성하고 이를 각 알고리즘에 적용하여 Anomaly detector와 MHC set을 구성한다. 이 후 두 가지 방법으로 자기 공간을 변경시켜 만든 검사 공간에 적용하여 알고리즘의 자기-인식의 정확도를 검사하여 자기-인식 알고리즘의 성능을 비교, 평가하였다. 또한 두 알고리즘을 동시에 적용하였을 경우의 자기-인식의 정확도도 비교, 평가하였다.

4.1 시뮬레이션 조건

자기-인식 알고리즘의 유효성을 검증하기 위한 시뮬레이션의 조건을 다음과 같이 설정하였다. 스트링의 집합인 자기 공간은 자기로 인식하는 공간으로 시뮬레이션에서는 일정 크기의 파일을 만들어 자기 공간으로 설정하였다. 자기 공간의 스트링의 개수는 시뮬레이션에서 각 800, 1600, 3200으로 변화시켜 시뮬레이션하였다. 각 스트링은 32개의 셀들로 구성하였으며 셀이 사용하는 심벌의 개수는 심벌의 개수는 256개로 설정하였다. 이는 2진 8비트의 크기로 문자를 나타내는 컴퓨터의 단위인 Character를 기본으로 스트링을 형성하기 위해서이다. 매칭과 생성에 가장 중요한 요소인 코드는 2로 설정하였다. 따라서 하나의 MHC 스트링은 각 위치에 따른 31개의 코드로 구성된다. 각의 MHC set과 Anomaly detector는 5개, 10개, 20개 그리고 30개로 구성하여 인식부의 개수에 따른 자기-인식 검사의 유효성을 검증하였다.

시뮬레이션은 각 조건에 대해 변경된 자기 공간 10,000개를 형성해 각 알고리즘에 적용하고 이의 자기인식률을 평가하였다. 이 때 자기 공간의 변경방법으로 두 가지를 사용하였다. 하나는 자기 공간의 부분적 국소 변경에 의한 자기-인식 알고리즘의 자기-인식률을 검증하기 위해 자기 공간의 몇 개의 셀을 변경하는 셀 변경(Cell Change) 방법과 실제적으로 해킹이나 바이러스에 의해 발생하는 자기 공간의 일정부분의 변경에 따른 자기-인식률을 확인하기 위해 자기 공간의 일정 개수의 스트링의 모든 셀을 변경하는 스트링 변경(String Change) 방법을 사용하였다.

4.2 셀 변경에 의한 시뮬레이션 결과

본 시뮬레이션은 자기 공간에서 셀을 변경한 경우의 시뮬레이션 결과이다. 자기 공간과의 유사도, MHC set과 Anomaly detector의 개수, 자기 공간의 크기를 변화시켜 시뮬레이션 하였으며 변경된 자기 공간 각 1,000개에 대해서 10번을 수행해 각 자기-인식 알고리즘이 자기로 잘못 인식한 회수의 평균을 결과로서 나타내고 있다. 표 1과 2는 MHC set과 Anomaly detector를 5개와 30개로 각각 구성하고, 자기 공간의 스트링의 개수를 각 800, 1600, 3200으로 설정하여 자기 공간과의 유사도에 따른 자기-인식 알고리즘의 인식을 나타낸다. 결과에서 보여주듯이 자기-인식 알고리즘은 유사도와 구성된 인식부의 개수에 따라 인식률의 정도가 다르게 나타남을 알 수 있다. 제안된 MHC set algorithm은 셀 변경에 의한 자기인식률에서 기존의 Anomlay detection algorithm과 유사한 자기인식률을 보임을 알 수 있다.

표 1. 셀 변경에 대한 시뮬레이션/인식부 개수: 5
Table 1. Simulation result/No. of detector: 5

Cell Change Method							
인식부 개 수		자기 공간의 크기					
5		800		1600		3200	
		MHC set	Anomaly	MHC set	Anomaly	MHC set	Anomaly
자기 공간과의 유사도	0.99	907.7	960	912.2	922.5	905	857.2
	0.98	827.2	932.1	817.5	863.1	817.1	735.4
	0.97	743	890.7	730	798.4	742.2	632.4
	0.96	610.3	836.1	682.3	747.9	652.4	550
	0.95	608	831.7	618.5	697.3	613.2	477.6
	0.94	546.3	883.9	543.7	635.3	554.5	420.1
	0.93	522.6	777.4	495.7	598.1	497.4	363.6
	0.92	453	758.2	444.2	604.1	440.1	320.6
	0.91	410.1	721.3	437	534.8	410.4	274.8
	0.90	388.6	700.8	390	491.2	351.5	251.5

표 2. 셀 변경에 대한 시뮬레이션/인식부 개수: 30
Table 1. Simulation result/No. of detector: 30

Cell Change Method							
인식부 개 수		자기 공간의 크기					
30		800		1600		3200	
		MHC set	Anomaly	MHC set	Anomaly	MHC set	Anomaly
자기 공간과의 유사도	0.99	584.1	804	557.3	634.6	552.5	400.8
	0.98	317.7	634	325	403.7	336	164.8
	0.97	192.7	504.4	190.5	258.4	182.9	71.5
	0.96	62.1	331.6	107.7	163	109.4	27.3
	0.95	54.1	343.3	66.2	111	63.1	13.5
	0.94	33.4	270.3	33.1	70	31.4	5.6
	0.93	19.4	218.1	21.1	53.1	18.7	2.2
	0.92	10.5	180.2	13.8	32.3	10.6	0.7
	0.91	6.5	140.5	5.2	19.2	4.7	0.3
	0.90	4.8	125.2	2.5	14.8	3.2	0.2

4.3 스트링 변경에 의한 시뮬레이션 결과

본 시뮬레이션은 스트링의 변경에 따른 자기-인식 알고리즘들의 자기-인식율을 보여주고 있다. 자기 공간의 일정 스트링의 모든 셀을 변경한 1,000개의 변경 공간에 대해 MHC set과 Anomaly detector를 5개와 30개로 변화시켜 10번을 수행해 자기 공간으로 잘못 인식한 회수의 평균을 결과 값으로 보여주고 있다. 이때 자기 공간의 크기와 자기 공간의 변경 정도에 따른 자기-인식 알고리즘의 인식률을 보여주고 있다. 표 3과 4는 MHC set과 Anomaly detector를 5개와 30개로 구성하고 자기 공간의 스트링의 개수를 각 800, 1600, 3200으로 설정하여 자기 공간과의 유사도에 따른 자기-인식 알고리즘의 인식을 나타내고 있다. MHC set algorithm의 스트링 변경에 따른 인식률이 Anomaly detection algorithm에 비해 높은 자기-인식률을 보임을 알 수 있다. 이는 자기-특성의 공간을 가지고 있는 MHC set algorithm이 블록 변경으로 생기는 자기 공간의 변경의 인식에 유리함을 보여주고 있다.

표 3. 스트링 변경에 대한 시뮬레이션/인식부 개수: 5
Table 3. Simulation result/ No. of detector: 5

String Change Method							
인식부 개 수		자기 공간의 크기					
5		800		1600		3200	
		MHC set	Anomaly	MHC set	Anomaly	MHC set	Anomaly
자기 공간과의 유사도	0.99	908.8	979.8	789.3	960.9	732.3	926.7
	0.98	845.3	958.8	695.6	927	625.2	856.9
	0.97	700.1	944.4	658.9	889.7	403.8	791.7
	0.96	675.7	906.6	473.7	856.2	331.9	733.6
	0.95	609.8	905.9	287.5	804.8	177.3	670.8
	0.94	604.9	887.3	257.5	792.1	93.3	623.1
	0.93	560.2	875.8	191.5	757.7	78.8	585.1
	0.92	395.2	852.9	99.4	739.5	71.5	533.7
	0.91	322.8	840.6	94.6	691.8	58.6	501.6
	0.90	100.3	821.3	1.7	689.4	31.1	455.5

표 4. 스트링 변경에 대한 시뮬레이션/인식부 개수: 30
Table 4. Simulation result/No. of detector: 30

String Change Method							
인식부 개 수		자기 공간의 크기					
30		800		1600		3200	
		MHC set	Anomaly	MHC set	Anomaly	MHC set	Anomaly
자기 공간과의 유사도	0.99	370.8	891.6	93.1	787.2	76.1	620.3
	0.98	342.7	788.8	61.6	625.8	12.2	397.1
	0.97	90.9	710.2	0.1	508.4	0	245.8
	0.96	0.1	554.9	0	396.9	0	161.3
	0.95	0	551.1	0	306.6	0	96.7
	0.94	0	495.4	0	239	0	63.9
	0.93	0	439.9	0	196	0	39
	0.92	0	393.3	0	157.8	0	24.1
	0.91	0	353.5	0	123.9	0	15.
	0.90	0	307.4	0	95.8	0	9.2

4.4 두 Detector에 의한 시뮬레이션 결과

본 시뮬레이션은 인식하는 방법이 다른 MHC set과 Anomaly detector를 동시에 적용시킨 자기-인식 알고리즘의 시뮬레이션을 하였다. 표 5는 인식부 10, 20개에 대해 셀 변경에 의한 시뮬레이션의 결과이며, 표 6은 인식부 10, 20개에 대해 스트링 변경에 의한 시뮬레이션 결과이다. 시뮬레이션은 두 인식부의 특성을 충분히 가지고 있는 자기-인식의 결과를 보여주고 있다.

그림 8과 9는 두 인식부를 이용한 자기 인식률과 MHC set, Anomaly detector를 이용한 자기 인식률을 비교하는 그래프로 그림 8은 자기 스트링이 1600, 인식부의 개수가 10일 때의 인식부들의 셀 변경에 따른 자기 인식률을 보여주고 있으며, 그림 9는 자기 스트링이 1600, 인식부의 개수가 20일 때의 스트링 변경에 따른 자기 인식률을 각각 보여주고 있다. 이와 같이 두 인식부를 이용한 자기 인식률이 같은 수의 각 인식부의 인식률과 유사하면서 동시에 두 인식부의 검색 영역을 공유한다.

표 5. 셀 변경에 대한 시뮬레이션
Table 5. Simulation result under cell change

Cell Change Method							
자기 공간의 크기		800		1600		3200	
인식부 개 수		10	20	10	20	10	20
자기 공간과의 유사도	0.99	871.9	762.8	830.5	701.1	774.8	535.4
	0.98	771.3	584.1	706.5	500	603	368
	0.97	661.3	452.7	581.6	350.1	470.8	224.3
	0.96	514.2	345.1	512.6	251.7	357.1	137.1
	0.95	509	255.1	434.4	184.6	281.1	85.3
	0.94	430.7	199.5	346.6	124.3	235	54
	0.93	406.3	156.6	297.5	84.7	181.1	32.8
	0.92	344.2	131.7	247.1	66.6	137.5	18.8
	0.91	294.7	98.9	237.9	52.3	113.5	13.3
	0.90	272.4	69	192.4	35.3	90.6	7.7

표 6. 스트링 변경에 대한 시뮬레이션
Table 6. Simulation result under string change

String Change Method							
자기 공간의 크기		800		1600		3200	
인식부 개 수		10	20	10	20	10	20
자기 공간과의 유사도	0.99	890.6	777.1	720.1	605.4	679.9	315
	0.98	767.5	386.9	643.8	439.7	537.3	109.4
	0.97	649.1	485.6	584.7	62.9	321.7	20.9
	0.96	633.6	301.4	404	109.3	242.1	35.7
	0.95	570.4	252.4	209.1	101.3	116.9	85.3
	0.94	517.2	87.8	197.3	6.1	50	13
	0.93	500	18	146.2	26.6	48	0
	0.92	332.8	148.3	80.6	0.1	43	0
	0.91	284.2	104.1	72.4	0	30.1	0
	0.90	81.9	49.6	1.1	0.2	15.2	0

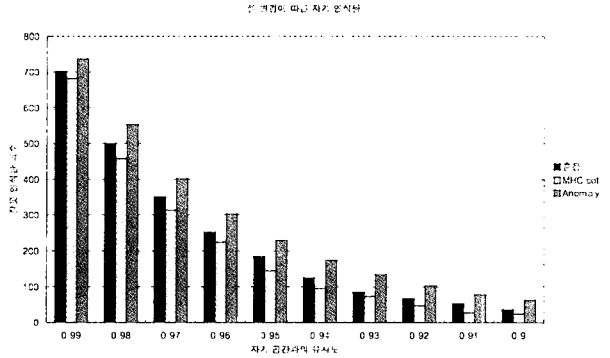

그림 8. 셀 변경에 따른 자기 인식률의 비교.
Fig. 8. Comparison of self-recognition rate under cell change

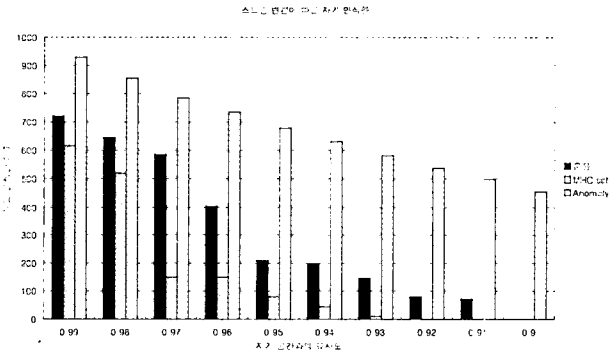

그림 9. 스티링 변경에 따른 자기 인식률의 비교.
Fig. 9. Comparison of self-recognition rate under string change

5. 결 론

최근에 컴퓨터의 보급과 인터넷의 발달로 많은 사람들이 컴퓨터를 사용하고 있다. 이러한 발전 뒤로 다른 사람의 컴퓨터를 침입하여 정보를 빼내거나 컴퓨터의 작동을 막는 컴퓨터 바이러스 등의 확산으로 피해가 증가하고 있다. 이에 본 논문에서는 이러한 해킹이나 컴퓨터 바이러스로부터 컴퓨터의 정보를 보호하기 위해 외부 침입 물질로부터 자기를 방어하는 시스템인 생체 면역계를 모델링한 컴퓨터 면역 시스템의 구현을 위한 자기 인식 알고리즘을 제안하였다. 이 알고리즘은 생체 면역계의 면역세포 특성인 자기를 구분하는 MHC 단백질을 인식하는 MHC 인식기능을 모델링하여 컴퓨터에서 자기 공간을 인식하는 알고리즘을 구현하였다. 구현된 MHC set을 이용한 자기 인식 알고리즘은 기존의 알고리즘과 비교해 자기 공간의 삭제에 대한 자기인식률과 일정 공간의 변경에 따른 인식률에서 그 유효성을 검증하였으며 두 자기 인식 알고리즘의 복합적 인식률은 상호보완적인 알고리즘으로 그 유효성을 입증하였다. 또한 자기를 인식하는 분야로의 적용 가능성을 입증하였다. 제안한 방법은 컴퓨터 면역시스템의 구현을 위한 기본적인 알고리즘으로 사용될 수 있을 것으로 기대된다.

참 고 문 헌

[1] S. Hofmeyr, S. Forrest, and A. Somayaji, "Intrusion Detection Using Sequences of System Calls." Journal of Computer Security vol. 6, pp. 151-180, 1998.

[2] C. Warrender, S. Forrest, B. Pearlmutter, "Detecting intrusions using system calls: Alternative data models," 1999 IEEE Symposium on security and Privacy (1999).

[3] D. Dasgupta, "An Immune Agent Architecture for Intrusion Detection.", Proceedings of The 2000 Genetic and Evolutionary Computation Conference (GECCO 2000) Workshop Program, pp. 42-44, 2000

[4] J. Gu, D. Lee, S. Park, and K. Sim, "An Immunity-based Security Layer Model," Proceedings of The 2000 Genetic and Evolutionary Computation Conference (GECCO 2000) Workshop Program, pp. 47-48, 2000

[5] J. Gu, D. Lee, K. Sim, and S. Park, "An Antibody Layer for Internet Security," Proceedings of Global Telecommunication Conference(GLOBECOM 2000) , pp. 450-454, 2000.

[6] J. Gu, D. Lee, K. Sim, and S. Park, "An Immunity-based Security Layer against Internet Antigens," Transactions on IEICE, vol. E83-B, no.11, pp. 2570-2575, 2000

[7] S. Forrest, A.S. Perelson, L. Allen, R. and Cherukuri, "Self-Nonself Discrimination in a Computer," In Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy, Los Alamitos, CA: IEEE Computer Society Press, 1994.

[8] P. Harmer, and G. Lamont, "An Agent Based Architecture for a Computer Virus Immune System," Proceedings of The 2000 Genetic and Evolutionary Computation Conference(GECCO 2000) Workshop Program, pp. 45-46, 2000.

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

[10] R. A. Wallace, G. P. Sanders, and R. J. Ferl, BIOLOGY : The Science of Life, 3rd eds., HarperCollins Publishers Inc., 1991.
[11] A. Somayaji, S. Hofmeyr, and S. Forrest, "Principles of a Computer Immune System," 1997 New Security Paradigms Workshop pp. 75-82, 1998.
[12] "Computer Immunology." S. Forrest, S. Hofmeyr, and A. Somayaji. Communications of the ACM vol. 40, no. 10, pp. 88-96, 1997.
[13] P. D'haeseleer, S. Forrest, and P. Helman, "An Immunological Approach to Change Detection: Algorithms, Analysis, and Implications,." In Proceedings of the 1996 IEEE Symposium on Computer Security and Privacy, 1996.
[14] D. Dasgupta, and S. Forrest, "An Anomaly Detection Algorithm Inspired by the Immune System." Artificial Immune Systems and Their Applications, Springer, pp. 262-276, 1999.
[15] D. Dasgupta and S. Forrest, "Novelty Detection in Time Series Data using Ideas from Immunology," In Proceedings of The International Conference on Intelligent Systems, 1999.
[16] D. Dasgupta and S. Forrest, "Artificial Immune Systems in Industrial Applications.", In International conference on Intelligent Processing and Manufacturing Material (IPMM), 1999.
[17] L. N. Castro, and F. J. Zuben, "The Clonal Selection Algorithm with Engineering Applications.", Proceedings of The 2000 Genetic and Evolutionary Computation Conference(GECCO · 2000) Workshop Program, pp. 36-37, 2000.
[18] K. Mori, K. Abe, M. Tsukiyama, and T. Fukuda, "Artificial Immune System based on Petri Nets and its Application to Production Management System.", Proceedings of The 2000 Genetic and Evolutionary Computation Conference(GECCO 2000) Workshop Program, pp. 51-52, 2000.
[19] M. Kawagoe and A. Tojo, "Fingerprint Pattern Classification," Pattern Recognition, vol. 17, no. 3, pp.295-303, 1984.
[20] L. O'Gorman, and J. V. Nickerson, "An approach to fingerprint filter design," Pattern Recognition, vol. 22, no. 1, pp. 29-38, 1989.
[21] P. Baldi, and Y. Chauvin, "Neural Networks for Fingerprint Recognition," Neural Computation, vol. 5, pp. 402-418, 1993.
[22] B. M. Mehtre, "Fingerprint Image Analysis for Automatic Identification," Machine Vision and Applications, vol. 6, no. 2-3, 1993.

저 자 소 개

심귀보(Kwee-Bo Sim)
1984년 : 중앙대학교 전자공학과 공학사
1986년 : 동 대학원 전자공학과 공학석사
1990년 : The University of Tokyo 전자공학과 공학박사
1997년~현재 : 한국퍼지 및 지능시스템학회 편집이사 및 논문지 편집위원장
2000년~현재 : 제어자동화시스템공학회 이사 및 직선평위원
2000년~현재 : 대한전기학회 제어및시스템부문회 편집위원 및 학술이사
1991년~현재 : 중앙대학교 전자전기공학부 교수

관심분야 : 인공생명, 진화연산, 지능로봇시스템, 뉴로-퍼지 및 소프트 컴퓨팅, 자율분산시스템, 로봇비전, 진화하드웨어, 인공면역계 등
Phone : +82-2-820-5319
Fax : +82-2-817-0553
E-mail : kbsim@cau.ac.kr

서동일(Dong-Il Seo)
1989년 : 경북대학교 전자공학과 공학사
1994년 : 포항공과대학교 정보통신학과 공학석사
2002년 : 충북대학교 전자계산학과(박사과정 수료)
1989년 1월~1992년 2월 : 삼성전자 종합연구소
1994년 3월~현재 : 한국전자통신연구원 사이버테러기술분석팀장

관심분야 : Network Security, 인터넷정보보호, Computer Network
Phone : +82-42-860-3814
Fax : +82-42-860-5611
E-mail : bluesea@etri.re.kr

김대수(Dae-Su Kim)
한국퍼지 및 지능시스템학회 논문지 제 11권 제9호 참조

임기욱(Kee-Wook Rim)
한국퍼지 및 지능시스템학회 논문지 제 11권 제9호 참조