

인공면역계의 자기-인식 알고리즘

Self-Recognition Algorithm of Artificial Immune System

선상준* · 이동욱* · 심귀보* · 성원기**

* 중앙대학교 전자전기공학부

** 삼척대학교 전자공학과

Sang-Joon Sun*, Dong-Wook Lee*, Kwee-Bo Sim*,** Won-Ki Sung**

* School of Electrical and Electronic Engineering, Chung-Ang University

** Department of Electronics Engineering, Samchok National University

(E-mail : kbsim@cau.ac.kr)

ABSTRACT

According as many people use a computer newly, damage of computer virus and hacking is rapidly increasing by the crucial users. 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 selection and negative selection of self-recognition process that is ability of T-cytotoxic cell that plays an important part in biological immune system. So we embody a self-nonself distinction algorithm in computer. To prove the efficacy of self-recognition algorithm, we use simulations by a cell change and a string change of self file.

Keywords : 인공면역계, 생체면역계, 자기인식 알고리즘, MHC

I. 서론

최근 컴퓨터의 사용이 보편화되면서 악의적 사용자에게 의해 발생하는 컴퓨터 바이러스와 해킹에 의한 피해가 급속히 증가하고 있다. 컴퓨터 바이러스는 자기 복제와 파괴 능력을 갖고 컴퓨터에서 실행되는 프로그램으로 감염 대상의 프로그램이나 데이터 파일을 파괴한다. 해킹은 다른 사람의 컴퓨터에 침입해 정보를 가져 오거나, 그 컴퓨터가 가지고 있는 정보를 없애는 작용으로, 인터넷의 발전에 따라 많은 컴퓨터로 피해가 확산되고 있다. 이렇게 남의 컴퓨터에 침입하는 해킹이나 데이터를 파괴하는 컴퓨터 바이러스의 피해를 막기 위해 최근에 생체 면역계의 특징을 컴퓨터에 적용한 인공면역계를 구성해 시스템 침입탐지와 바이러스 탐지, 치료에 대한 연구가 진행 중에 있다[1,4].

생체 면역계는 외부에서 침입해 세포나 장기에 피해를 주는 물질인 항원을 스스로 자기세

포와 구분해 인식, 제거하는 기능이 있다. 항원 인식 기능은 자기세포의 확실한 인식을 가지고 있는 상태에서 다른 물질을 구분하는 자기/비자기 인식방법으로 볼 수 있다[5].

면역계를 자기-인식 알고리즘에 적용한 사례는 S. Forrest의 Negative Selection을 이용한 비자기 인식부(Nonsel Self Detector)의 자기 인식 기능을 제안하였다[2,3]. 그러나 이 알고리즘은 전체공간이 커지면 Detector를 형성하기 힘들며 블록단위의 변경에 대한 자기인식률이 떨어진 다. 본 논문에서는 생체 면역계에서 세포독성 T세포 생성과정의 하나인 Positive Selection을 모델링하여 자기-인식 알고리즘을 구현하였다. 제안한 알고리즘은 큰 파일에서의 Detection을 구성하기 용이한 점을 가지며 국소변경과 블록 변경에 대한 자기인식률을 통해 알고리즘의 유효성을 검증한다.

II. 생체 면역시스템

2.1 생체 면역시스템 상호작용

생명체의 방어체제인 면역계는 박테리아, 기생균, 바이러스 등과 같이 항원이라고 통칭하는 매우 다양한 외부 유기체나 단백질에 대하여 생명체의 세포와 장기를 방어할 수 있는 매우 정교하고 복잡한 시스템이다. 면역계를 구성하는 기본요소는 두 가지 형태의 림프구로 B세포와 T세포이다. B세포는 항원을 죽이는 항체를 생산, 분비하는 체액성 반응을 하며, T세포는 면역에 관련된 세포를 자극 또는 억제하거나 항원에 의해 감염된 자기세포를 죽이는 세포성 반응을 주로 담당한다.

외부의 침입 물질인 항원이 발생하면 대식세포 등과 같은 식세포에 의한 면역반응이 형성된다. 이 과정에서 대식세포는 항원의 특정부위인 항원결정소의 정보를 수집, 이를 항원의 모습으로 설정하여 제공한다. 이렇게 모여진 정보는 T세포의 하나인 보조 T세포(T-helper Cell)를 통해 세포독성 T세포(T-cytotoxic Cell)와 B세포(B-Cell)를 자극한다. 자극 받은 B세포는 항체를 생산하며 이 후에 같은 항원의 재침입의 면역반응을 위한 기억세포가 된다. 세포독성 T세포는 항원에 의해 감염된 자기 세포를 제거하는 역할을 한다. 항원이 감소하기 시작하면 T세포의 일종인 억제 T세포(T-suppressor Cell)에 의해 B세포와 T세포의 활동이 억제되어 면역반응이 감소하게 된다. 그림 1은 생체면역시스템의 상호작용을 보여주고 있다.

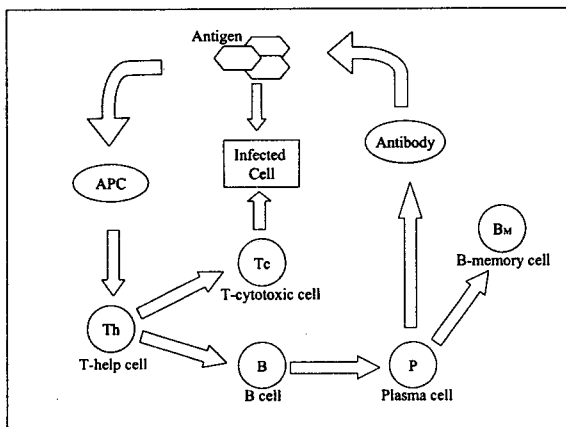


그림 1. 생체 면역시스템의 상호작용도
Fig. 1 Interaction of biological immune system

2.1 Positive & Negative Selection

각 면역세포는 항원과 자기세포들을 인식하

는 각각의 수용체를 가지고 있다. B세포에는 항원수용체가 존재하여 항원을 구분하며, 세포독성 T세포는 항원에 감염된 자기세포를 구분하기 위해 항원수용체와 자신의 세포를 구별하기 위해 자기세포 판별용 단백질인 MHC(Major Histocompatibility Complex, 주조직 적합성 복합체)를 인식하는 MHC 인식부(MHC Receptor)를 가지고 있다.

Positive Selection은 각 면역세포의 MHC 인식기능을 확인하는 선택방법이다. 자기세포에서 분비되는 MHC 단백질을 인지할 수 있는 면역세포만이 사용가능하기에 갖 생성된 면역세포에 MHC 단백질을 결합시켜 긍정적인 선택이 되는 세포들로만 면역 세포를 구성하게 된다.

Negative Selection은 항원의 인식에 있어서 자기를 항원으로 인식하는 것을 배제하기 위해 방법이다. 항원수용체가 MHC 단백질을 항원으로 인식하면 모든 세포를 항원으로 인식하기 때문에 항원으로 MHC 단백질을 인식하지 못하게 하기 위해 면역세포에 MHC 단백질을 결합시켰을 때 항원수용체가 부정적인 선택을 하는 세포만으로 구성된다.

III. 알고리즘 구현

본 논문에서는 면역세포의 생성시 MHC 인식부를 형성해 주는 Positive Selection을 자기-인식 알고리즘으로 구현하여 컴퓨터에서 자기로 인식해야하는 파일이나 기능에 대해 MHC 인식부를 형성해 자기와 비자기를 인식하게 하였다.

MHC 인식부는 자기 파일내에서 가져온 정보로 구성한다. 컴퓨터 상의 자기파일은 심벌로 구성된 스트링(String)들의 집합이다. 이때 일정한 스트링의 길이를 셀(Cell)이라 정의한다. 제한된 MHC 인식부는 자기파일을 셀의 길이로 나누어 생기는, 셀 위치 값을 가진 자기파일의 연속적인 스트링들을 가져와서 하나의 셀을 구성한다. MHC 인식부의 일부를 구성하는 연속적인 스트링을 코드라고 한다. 하나의 MHC 인식부는 자기 파일에서 코드들을 이용하여 구성한다. 이렇게 구성한 하나의 MHC 인식부는 자기파일 셀들의 같은 위치에 존재하는 같은 심벌들을 코드의 크기 값만큼 MHC로 구성해 자기파일의 정보를 저장하게 된다. MHC 인식부를 이용하여 검사할 때는 구성된 MHC 인식부들의 데이터를 토대로 위치에 따른 코드의 존

재여부를 이용해 자기파일을 구분한다.

다음은 자기-인식 알고리즘에서 자기와 비자기를 구분하고 자기파일의 특정 정보를 가지고 있는 MHC 인식부를 구성하는 알고리즘이다.

1. 자기파일(컴퓨터 상에서 자신으로 인식해야 하는 파일)을 일정한 크기인 셀로 나눈다. 이 셀은 하나의 MHC 인식부의 길이가 되며 자기파일 변경검사의 단위가 된다.
2. 자기파일 셀들의 위치 값에 존재하는 자기파일 셀의 연속적인 스트링인 코드를 MHC 인식부 위치의 코드로 구성한다.
3. 셀의 앞의 위치 값에 첫 번째 코드부분을 포함하는 자기파일의 셀을 찾아 MHC 인식부 다음 위치의 코드로 구성한다. 구성된 두 번째 코드는 첫 번째 코드의 부분과 새로운 부분으로 구성되게 된다.
4. 위의 방법을 반복해서 셀의 크기를 구성하는 코드들을 찾아 구성된 셀이 하나의 MHC 인식부가 된다. 따라서 MHC 인식부는 각 위치 값에서 자기파일의 정보를 가지게 된다.
5. 일정 개수의 MHC 인식부 셀들을 만들어 자기-인식부를 구성한다.

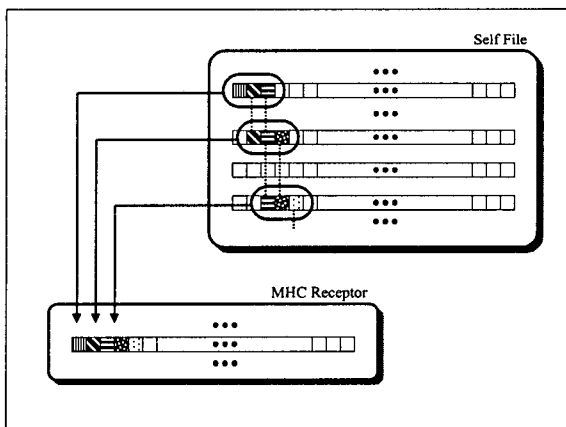


그림 2. 자기파일에서의 MHC 인식부 구성 방법
Fig. 2 Method of constructing MHC receptor in self files

그림 2는 자기 파일에서 MHC 인식부를 구성하는 방법으로 셀의 위치에 존재하는 자기파일의 코드를 모아서 하나의 MHC 인식부를 구성하는 것을 보여준다. 코드의 길이를 3으로해 동일한 패턴의 스트링들을 MHC 인식부의 자기파일 셀과 같은 위치에 구성하여 하나의 인식부가 자기파일의 여러 부분을 내재되어 구성한다.

자기-인식 알고리즘은 구성된 MHC 인식부들을 이용해 이루어진다. 각각의 인식부의 코드들이 자기파일의 해당 위치의 존재를 검사한다. MHC 인식부를 이루고 있는 코드가 모두 존재하면 해당 인식부에 의해서는 자기파일로 인식하고 모든 MHC 인식부가 자기파일로 인식하면 검사한 파일은 자기파일로 인식하게 된다.

IV. 시뮬레이션

생체 면역 시스템에 기반한 자기-인식 알고리즘의 유효성을 검증하기 위해 기존에 제안된 S. Forrest의 자기-인식 알고리즘과 비교하였다. 시뮬레이션은 컴퓨터에서 자기파일을 만들고 변경된 자기파일의 비자기 인식률로 알고리즘의 정확도를 검사하여 알고리즘의 성능을 비교, 평가하였다.

4.1 시뮬레이션 조건

자기파일은 컴퓨터에서 자기로 인식하는 파일로, 시뮬레이션에서는 일정 크기의 파일을 구성해 자기파일로 설정하였다. 자기파일의 각 스트링을 이루는 심벌의 개수는 256개 2진 8비트 크기로 구성하였다. 셀은 32개의 스트링, 그리고 총 1600개의 셀로 자기파일을 구성하였다. 또한 MHC 인식부의 코드의 길이는 2로 설정하며 인식부는 5개, 10개, 20개, 30개를 구성하여 각각의 경우에 따른 자기파일 변경 검색의 유효성을 검증하였다. 비자기 인식부의 시뮬레이션 조건도 동일하게 설정하였다.

시뮬레이션은 10,000개의 변경파일을 형성해 자기인식률을 평가하였다. 이때 두 가지의 자기파일의 변경방법을 사용하였다. 하나는 국소 변경에 의한 자기파일 인식률을 검증하기 위해 몇 개의 스트링을 변경하는 스트링 변경(String Change)방법과 자기파일의 일정부분 변경에 따른 인식률을 확인하기 위한 셀 변경(Cell Change) 방법의 두 가지를 설정하였다.

4.2 시뮬레이션 결과

표 1은 스트링 변경의 정도와 MHC 인식부 개수, 비자기 인식부의 개수를 변화시켜 얻은 자기-인식 알고리즘 시뮬레이션 결과이다. 시뮬레이션 결과 값은 각 10,000개의 파일에 대해 각각의 MHC 인식부와 비자기 인식부가 변경된 파일을 자기파일로 인식한 경우를 나타낸다. 표

에서 보듯이 MHC 인식부와 비자기 인식부를 이용한 자기-인식률은 비슷한 결과를 보여주고 있으며 두 가지의 인식부 모두 변경된 파일간의 유사도와 인식부의 개수가 인식율에 영향을 주어 낮은 유사도와 많은 개수일 때 인식율이 높아짐을 알 수 있다.

표 1. 스트링변경에 의해 변경된 자기파일 10,000개에 대해서 자기-인식알고리즘이 자기로 잘못 인식한 횟수

Table 1 In string-change simulation, this is number that recognizes as self for 10000 changed self file.

인식부	인식부 개수	자기파일과의 유사도					
		0.999	0.998	0.996	0.984	0.976	0.969
MHC	5	9935	9850	9754	8568	7862	7311
	10	9891	9617	9450	7324	6532	6113
인식부	20	9775	9297	8912	5641	3797	3233
	30	9626	8952	8308	4089	2511	1759
비자기	5	9943	9846	9757	8968	8413	7995
	10	9905	9718	9509	7857	7060	6395
인식부	20	9812	9446	9027	6288	4899	3807
	30	9707	9160	8713	4966	3392	2439

표 2. 셀 변경에 의해 변경된 자기파일 10,000개에 대해서 자기-인식 알고리즘이 자기로 잘못 인식한 횟수

Table 2 In cell-change simulation, this is number that recognizes as self for 10000 changed self file

인식부	인식부 개수	자기파일과의 유사도					
		0.990	0.980	0.970	0.960	0.940	0.900
MHC	5	9796	9989	6861	2400	2562	0
	10	6815	2803	612	7	2	0
인식부	20	2659	1762	0	3	0	0
	30	187	366	2	0	0	0
비자기	5	9649	9240	8924	8593	7911	6784
	10	9280	8548	7950	7320	6248	4600
인식부	20	8544	7289	6223	5422	3907	2161
	30	7889	6187	4979	3925	2451	968

표 2는 셀의 변경에 따른 자기-인식 알고리즘들의 자기-인식율을 보여주고 있다. 10,000개의 변경파일에 대해 MHC 인식부와 비자기 인식부의 개수를 변화시켜 잘못 인식한 경우를 나타냈다. 표에서 보듯이 MHC 인식부의 인식률이 비자기 인식부에 의한 인식률에 비해 높은 성능을 보여주고 있다. 특히 적은 MHC 인식부의 개수에 대해서도 높은 자기-인식 효율

성을 보여주고 있다. 이는 블록화된 자기파일의 변경부분에 대해 MHC 인식부에 의한 자기변경검사가 비자기 인식부에 의한 경우보다 정밀하게 인식함을 보여준다.

V. 결론 및 향후과제

본 논문에서는 생체 면역계의 세포독성 T세포에서 MHC 단백질을 인식하는 MHC 인식기능을 이용하여 컴퓨터에서 파일의 자기-인식을 검사하는 알고리즘을 구현하였다. 이렇게 구현한 MHC 인식부는 변경된 자기파일에 대한 자기-인식률을 기존의 알고리즘과 비교해 그 유효성을 검증하였다. 이에 컴퓨터 면역시스템 구축에 필요한 자기-인식 알고리즘의 적용 가능성을 보였다. 구현한 알고리즘은 큰 파일에서도 기존의 알고리즘에 비해 쉽게 구성이 가능하며 일정부분의 변경에 대해서는 높은 자기 인식률을 보여준다. 아직 자기파일과의 유사도가 작은 변경 검사에서는 그 유효성이 미미함의 보안을 위해 MHC 인식부와 비자기 인식부를 통한 종합적인 자기-인식 알고리즘의 수정이 요구된다.

감사의 글

본 연구는 한국과학재단 특정기초연구의 연구비 지원으로 수행되었으며, 또한 2000년도 한국산업자원부 제2차 산업기반기술개발사업에서 연구비를 일부 지원 받았습니다. 연구비지원에 감사 드립니다.

참고문헌

- [1] D. Dasgupta, "An Immune Agent Architecture for Intrusion Detection", Genetic and Evolutionary Computation Conference Workshop Program pp.42 - 44, 2000.
- [2] 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
- [3] S. Forrest, L. Allen, A. S. Perelson, R. Cherkuri "Self-Nonself Discrimination in a Computer" IEEE Symposium on Research in Security and Privacy, 1994.
- [4] P. K. Harmer, G. B. Lamont, "An Agent Based Architecture for a Computer Virus Immune System", Genetic and Evolutionary Computation Conference Workshop Program pp. 45 - 46, 2000.
- [5] I. Roitt, J. Brostoff, D. Male, Immunology, 4th edition, Mosby, 1996.