

컴퓨터 면역시스템 개발을 위한 생체 면역시스템 모델링

Biological Immune System Modeling for Development of Computer Immune System

신상준 · 심재윤 · 심귀보

Sang-Joon Sun, Jae-Yoon Sim, and Kwee-Bo Sim

중앙대학교 전자전기공학부

School of Electrical and Electronic Engineering, Chung-Ang University

요 약

생체 면역시스템은 바이러스나 병원균 등의 외부 침입자로부터 자신을 보호한다. 본 논문에서는 이러한 기능을 컴퓨터 면역시스템에 적용하기 위하여 생체 면역시스템의 기본 요소인 T세포를 모델링하였다. 모델링한 T세포는 MHC 인식부와 항원 인식부를 가지고 있으며, 매칭을 통하여 컴퓨터내의 랜덤 파일로부터 자기 파일과 비자기 파일을 구분할 수 있다. 모델링한 T세포의 유효성을 보이기 위해서 컴퓨터 바이러스의 문제에 적용하여 자기-비자기 구분이 가능함을 보인다.

ABSTRACT

Biological immune system protects self from intrusion such as virus, pathogenic bacteria. In this paper, we make a model of the T-cell of biological immune system to adopt this function into computer immune system. The modeled T-cell possesses the ability to be able to distinguishes self files from random files in computer. The T-cell consists of MHC perception and antigen perception, and distinguishes self from non-self through matching file which we must test. The preliminary experiments illustrate how the modeled T-cell might be applied to the problem of computer virus.

Key Words : 인공 면역시스템, 자기-비자기 식별알고리즘, 컴퓨터 면역시스템, 컴퓨터 바이러스, 매칭 알고리즘.

1. 서 론

생체 면역시스템은 외부의 물질로부터 생명체의 몸을 지키는 방어시스템으로 외부의 물질을 인식하고 자기와 비자기를 구분하며 학습 및 기억 능력과 분산시스템으로서 전체의 조화를 유지하는 능력을 가지고 있는 복잡하고 정교한 시스템이다[1, 2]. 그 중에서 자기-비자기(Self-Nonself) 인식은 외부에서 침입한 존재나 내부에서 변형된 존재를 자기 외의 존재로 인식함으로써 생체 면역시스템의 감시작용을 한다. 이러한 생체 면역시스템의 자기-비자기 인식은 새로운 컴퓨터 바이러스 등에 능동적으로 대처할 수 없는 현재의 바이러스 검출 및 치료 방법에 응용이 가능해 많은 연구가 활발히 진행 중에 있다[3-7].

생체 면역시스템을 자기와 비자기의 구분에 적용한 사례는 Stephanie Forrest와 Lawrence Allen의 항원 인식에 의한 자기-비자기 식별 모델이 있다[3]. 본 논문에서는 항원 인식에 의한 자기-비자기 구분 모델에서 고려하지 않는

MHC(Major Histocompatibility Complex, 주조직 적합성 복합체) 인식부를 이용하여 생체 면역시스템의 자기-비자기 식별 모델을 제안한다. 제안한 모델을 이용하여 자기-비자기 식별 시뮬레이션한 결과 항원 인식부만을 사용한 모델에 비해 MHC 인식부를 같이 사용한 경우의 자기-비자기 인식률이 높아짐을 입증하였다.

2. 자기-비자기 식별 모델

2.1 생체 면역시스템

외부에서 침입한 존재를 감지하고 이를 제거하는 생체 면역시스템을 구성하는 기본 요소는 B세포(B-cell)와 T세포(T-cell)이다.

B세포는 생체에 침입하거나 들어온 외부 물질인 항원에 대항해서 항체를 분비함으로써 항원을 제거하는 역할을 한다. T세포는 기능에 따라 몇 가지로 구분된다. B세포를 활성화시켜 항체의 분비를 촉진하는 역할을 하는 보조 T세포(Helper T-cell)와 항원에 의해 감염된 자기 세포를 식별하여 죽이는 역할을 하는 세포독성 T세포(Cytotoxic T-cell), 그리고 항체에 의한 면역 시스템이 활성화된 이후 시스템을 억제하는 역할을 하는 억제 T세포(Suppressor T-cell)로 구분된다. 특히 세포독성 T세포는 각각의 세포가 자기 세포임을 알리는 MHC 단백질질을 인식하는 부분과 외부 물질인 항

접수 일자 : 2000년 1월 5일

완료 일자 : 2001년 3월 22일

본 연구는 한국산업자원부 2000년 제2차 산업기반기술 개발사업(공통핵심/Spin-Off)의 연구비지원으로 수행되었으며 연구비 지원에 감사드립니다.

원을 인식하는 항원 인식부분으로 구성되어서 세포가 자기 세포인지 아닌지, 그리고 항원에 의해 감염된 세포인지 아닌지를 구분할 수 있다.

2.2 MHC 인식부와 항원 인식부

세포독성 T세포는 항원의 침입에 손상을 받아 감염된 자기 세포를 찾아서 제거하는 역할을 수행한다. 감염된 세포를 찾기 위한 방법으로 T세포는 세포가 자기 세포인가를 확인하고 그 세포 내에 항원의 존재를 검사하여 찾는다. 감염된 세포를 구분하는 식별인자인 MHC와 항원을 이용하여 세포를 구분한다. MHC는 자기세포에서 분비되는 단백질로서 T세포의 MHC 인식부에서 MHC임을 인정해야 자기임을 구분하게 된다. 항원은 외부에서 침입해 온 세포나 박테리아 등으로 이를 인지하는 항원 인식부에서 항원인지를 검사함으로써 외부 물질임을 알아낸다. 이렇게 T세포는 2가지의 검사방법에 의해 감염된 세포를 찾을 수 있다. 각각의 인식부들은 초기 T세포가 형성될 때 2가지 방법에 의해 결정된다. MHC 인식부는 Positive Selection을 통해 MHC를 구분하는 세포만으로 초기 T세포를 구성한다. 항원인식부에서는 MHC를 인식하는 인식부를 형성하면 MHC를 항원으로 인식하게 되므로 MHC를 인식하지 못하게 Negative Selection을 이용하여 MHC를 항원으로 인식하는 T세포를 제거하는 방법으로 구성한다. 이와 같은 2가지의 방법으로 형성된 T세포의 MHC인식부와 항원인식부는 자기세포와 감염된 세포를 구분하는 정상적인 활동을 하게된다. 그림 1은 형성된 T세포가 MHC 인식부와 항원 인식부를 이용하여 세포의 감염 여부를 확인하는 모식도를 나타낸다

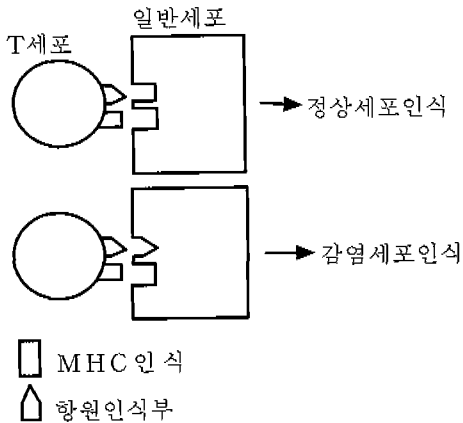


그림 1. 세포의 감염여부를 확인하는 모식도
Fig. 1. Picture that certifies inspected cell

2.3 생체 면역시스템 모델링

본 논문에서는 자기와 비자기를 구분하여 인식하는 생체 면역 시스템의 세포독성 T세포의 특성을 컴퓨터의 면역시스템에 적용 가능하게 하는 모델을 제안하였다. 컴퓨터에서 자기를 인식하고 있으면 외부의 침입으로 인한 자기의 변화를 감지할 수 있다.

모델링된 T세포는 두 부분의 인식부 즉, 자기를 식별하는 MHC 인식부와 비자기를 식별하는 항원 인식부로 구성된다. 이러한 T세포의 MHC 인식부는 자기로 인식되는 파일의 일부를 이용해서 구성하고, 항원 인식부는 랜덤 스트링을 매칭해서 구성한다. 그림 2는 T세포의 인식부를 구성한 예를 나

타낸다.

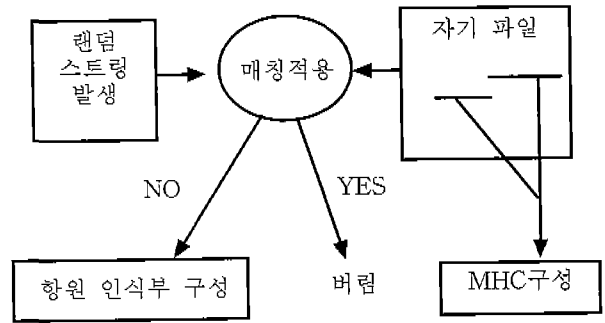


그림 2. T세포의 인식부 구성
Fig. 2. Construction of T-cell receptor

이렇게 구성된 각각의 인식부를 이용해서 자기-비자기 검색을 필요로 하는 파일과의 매칭을 통해 자기-비자기를 식별하게 된다. 그림 3은 T세포를 이용한 자기-비자기 검색 과정을 나타낸 그림이다.

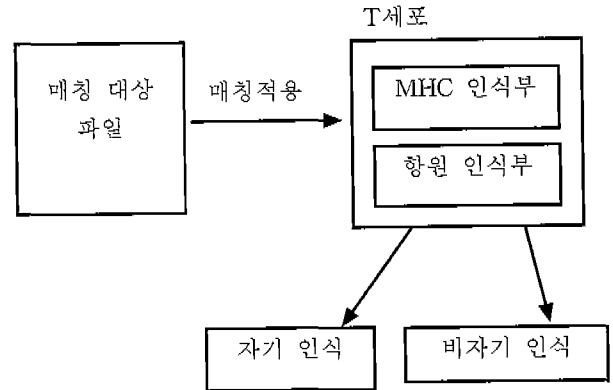


그림 3. T세포를 이용한 자기-비자기 검색
Fig. 3. A search of self-nonself using T-cell

2.3.1 MHC 인식부

MHC는 생체 면역시스템에서 세포들 자신이 자기임을 알리는 단백질로 이를 이용하여 T세포는 그 세포가 자기 인지를 구분한다. 모델링된 MHC 인식부는 자기로 인식되는 파일의 일부를 스트링 단위로 M 개를 선택하여 구성한다. 구성된 MHC 인식부를 통하여 자기-비자기 검색을 했을 때 자기로 인식할 확률(P_M)은 (1)식과 같이 나타낼 수 있다.

$$P_M \approx \frac{s_p C_M}{s C_M} = \frac{S_p!(S-M)!}{S!(S_p-M)!} \quad (1)$$

식에서,

S : 자기 파일의 전체 스트링의 개수

M : 구성된 MHC 인식부의 개수

S_p : 검색 파일 내의 자기 파일과 같은 스트링의 개수

를 각각 나타낸다. 한편 표 1은 S, S_p, M 을 임의로 선정했을 때 MHC 인식부가 자기로 인식할 확률을 나타낸 것이다.

표 1. S, S_p, M 을 임의로 선정했을 때 MHC 인식부가 자기로 인식할 확률 P_M

Table 1. Probability P_M that is recognized as self using MHC receptor, when S, S_p, M is randomly selected

S	S_p	M	P_M
1600	1440	5	0.5901
1600	1440	10	0.3476
1600	1440	20	0.1200
1600	1520	10	0.5978
1600	1520	20	0.3562
1600	1520	30	0.2116

2.3.2 항원 인식부

생체 면역시스템에서 항원 인식부는 MHC를 인식하지 못하도록 구성되어 있기 때문에 자기를 인식하지 못한다. 따라서 항원 인식부는 비자기를 인식하기 위한 부분인 것이다. 모델링된 항원 인식부는 자기 파일과 매칭이 되지 않은 A 개의 스트링들로 구성되어 자기와의 매칭이 되지 않도록 하여 비자기를 인식한다.

자신과 비교해 항원 인식부를 구성할 때 자신과 완전히 다른 스트링으로 구성하는 것은 일반적으로 매우 힘들다. 이렇게 전혀 매칭이 되지 않은 스트링으로 항원 인식부를 구성하기 위해서는 파일의 크기가 매우 작아야 한다. 자기 파일의 크기와 빠른 항원 인식부의 구성을 위해서는 매칭을 검사할 때 자기 파일과 구성할 스트링이 r 개의 심벌이 연속적으로 같지 않도록 항원 인식부를 구성한다. 이러한 항원 인식부를 구성하면 매칭이 되지 않은 스트링을 찾을 수 있는 자기 파일의 설정 크기가 커진다. 다음은 r 개의 심벌을 이용하여 매칭을 검사할 때 언제나 항원 인식부의 스트링 A 를 구할 수 있는 자기 파일의 최대 스트링의 개수를 설정한다. 자기 파일 스트링의 최대값 S_R 은

$$\text{If } P(A) = 1, \text{ then } S_R < m^r - A \quad (2)$$

이다. 여기서 A 는 항원 인식부의 개수, m 은 사용한 심벌의 개수, r 은 연속적인 심벌 개수, S_R 은 r 개의 연속적인 심벌을 이용한 매칭에서 자기 파일 스트링의 수를 각각 나타낸다.

표 2는 A, m, r 을 임의로 선정했을 때 자기 파일의 사용 가능한 최대값 S_R 을 나타낸다.

표 2. A, m, r 을 결정했을 때 자기 파일의 사용 가능한 최대값 S_R

Table 2. Usable maximum number S_R of self file string, when A, m, r is selected

m	r	A	S_R
256	1	20	236
256	2	20	65516
256	3	20	16777196

항원 인식부는 랜덤하게 스트링을 하나 만들어 자기 파일과 r 개의 연속적인 심벌이 존재하는지를 검사하는 매칭 연산을 한다. 따라서 구성된 스트링은 자기 파일과는 매칭이

되지 않은 특성을 가지므로 자기 외의 비자기를 구분할 수 있다. 이때 매칭이 되는 스트링은 버리고 또 다른 랜덤한 스트링을 형성해 동일한 연산을 수행한다. 이러한 단계를 거쳐 구성된 항원 인식부로 검사가 필요한 파일을 매칭하여 매칭이 되지 않으면 자기로 인식하고, 매칭이 되면 비자기로 인식하여 자기와 비자기를 구분하게 된다.

매칭의 검사 방법인 r 개의 연속적인 심벌의 존재를 이용하여 두개의 랜덤 스트링이 매칭 될 확률(P_r)을 계산하면 (3)식과 같이 되는데 $m^r \ll 1$ 인 경우에만 식 같이 근사화할 수 있다[3].

$$P_r \approx \frac{(l-r)(m-1)}{m^r} + 1 \quad (3)$$

여기서 l 은 스트링의 길이, m 은 심벌을 개수를 나타낸다. 한편 항원 인식부를 스트링 1개로 구성하여 임의의 파일과 매칭 연산을 수행하였을 때 그 파일이 자기 파일로 인식될 확률 ($P_{A(1)}$)은 (1)식과 같이 되므로,

$$P_{A(1)} = (1 - P_r)^{(S-S_p)} \quad (4)$$

이를 확장하면 항원 인식부를 다수의 스트링으로 구성하여 매칭 연산을 수행할 때 자기 파일로 인식될 확률 (P_A)을 (5)식으로 나타낼 수 있다.

$$P_A \leq (1 - P_r)^{A(S-S_p)} \quad (5)$$

(5)식에서 등호는 항원 인식부가 매칭할 수 있는 집합들이 독립일 때만 가능한 경우이다.

표 3. m, r, l 을 임의로 설정했을 때 두개의 랜덤 스트링이 매칭될 확률 P_r

Table 3. A probability P_r to be matched two random string, when m, r, l is randomly selected

m	l	r	P_r
256	32	2	0.000471
256	16	2	0.000228
256	8	2	0.000106
256	32	8	~0.0
256	16	8	~0.0

표 4. P_r, A, S, S_p 를 임의로 설정했을 때 항원 인식부가 자기로 인식할 확률 P_A

Table 4. A probability P_A that is recognized as self, when P_r, A, S, S_p is randomly selected

P_r	A	S	S_p	P_A
0.000471	10	1600	1440	0.47058
0.000471	20	1600	1440	0.22145
0.000471	20	1600	1520	0.47058
0.000471	30	1600	1520	0.32281
0.000228	10	1600	1440	0.69430
0.000228	20	1600	1440	0.48206

표 3은 심벌의 개수와 스트링의 길이를 결정한 두개의 임의의 스트링을 매칭했을 때 r 개의 연속적인 심벌이 있을 확률을 계산한 예를 나타낸다.

그리고 표 4는 P_r 을 결정한 상태에서 A 개의 항원 인식부 스트링을 이용하여 원래의 자기 파일에서 변환된 파일을 매칭했을 때 자기 파일로 인식할 확률을 나타낸다.

2.2.3 자기-비자기 구분

자기-비자기 구분은 T세포의 두 인식부로 각각 매칭 검사를 하여 얻어진 결과가 자신인지 아닌지를 결정함으로써 이루어진다. 이 경우 어느 하나라도 자신이 아니라는 결과를 가지고 있으면 전체적인 결과는 자신이 아닌 것으로 인식하게 된다. 각각의 매칭 결과를 검사하기 때문에 결국 P_M 과 P_A 는 독립적인 사건이 된다. P_S 는 MHC 인식부와 항원 인식부를 가지고 매칭을 했을 경우에 자신이라고 인식될 확률을 나타내며, 다음 식으로 표현된다.

$$P_S = P_M \cap P_A = P_M \times P_A \quad (6)$$

위의 식에서 P_M 은 (1)식으로 표현되는 자기로 인식할 확률을 나타내는데 자기를 검색할 경우에는 "1"이 되고, 그 외의 경우는 1보다 작게 된다. 따라서 MHC 인식부와 항원 인식부를 이용해서 자기-비자기를 식별할 경우의 P_S 의 값은 항원 인식부만을 가지고 자기-비자기를 구분하는 경우인 P_A 보다 작아지게 되어 자기-비자기를 구분하는 인식 확률이 높아진다.

3. 시뮬레이션 결과

컴퓨터에서 적용이 가능한 자기와 비자기를 구분할 수 있게 모델링한 T세포를 이용하여 자기-비자기를 식별하는 시뮬레이션을 수행하였다.

시뮬레이션에서는 사용 가능한 심벌의 개수 m 을 이진 8bit인 256개로 하였다. 이는 컴퓨터의 기본 입출력 단위가 1byte단위인 Character변수이기 때문이다. 또한 연속적인 심벌 개수 r 을 2, 스트링의 길이 l 을 32, 자기 파일의 스트링의 개수 n 를 1600으로 각각 설정하였으며, T세포의 MHC 인식부 M 과 항원 인식부 A 를 각각 5개, 10개, 20개, 30개로 S 만들어, 자기 파일과의 유사도 S_p/S 를 0.98, 0.95, 0.90로 하였을 때 각각의 경우에 대해 10000번의 시뮬레이션을 통하여 자기-비자기의 식별 가능성을 검사하였다.

상기의 조건에서 시뮬레이션을 수행한 결과를 표 5에 나타낸다.

표 5에서 P_M 은 MHC 인식부만 가지고 있는 경우에 식별한 결과를, P_A 는 항원 인식부만 가지고 있는 경우에 식별한 결과를, P_S 는 MHC와 항원 인식부를 모두 가지고 있는 경우에 식별한 결과를 각각 나타낸다. 표를 보면 알 수 있는 바와 같이 MHC 인식부의 개수가 많아지면 식별 능력이 좋아짐을 알 수 있다. 만약 MHC 인식부의 개수가 많아지면 항원 인식부만을 사용한 경우보다 MHC 인식부만을 사용한 경우가 식별 능력이 우수하고, 이 두 가지의 인식부를 모두 갖는 경우에는 식별 능력이 아주 우수하다는 것을 알 수 있다. 결과적으로 항원 인식부만을 사용한 모델에 비해 MHC 인식부를 같이 사용한 경우의 자기-비자기 인식률이 높아짐을 알 수 있다.

표 5. 자기-비자기 식별 시뮬레이션 결과
Table 5. A simulation result of distinguishing self-nonself

인식부 개 수	S_p/S	P_M		P_A		P_S	
		이론	결과	이론	결과	이론	결과
5	0.98	0.9038	1.00	0.9274	0.9296	0.8382	0.9296
10	0.98	0.8166	1.00	0.8601	0.8569	0.7024	0.8569
20	0.98	0.6659	0.5397	0.7397	0.7337	0.4926	0.3950
30	0.98	0.5424	0.3431	0.6361	0.6206	0.3450	0.2149
5	0.95	0.7735	0.5608	0.8282	0.8224	0.6406	0.4622
10	0.95	0.5978	1.00	0.6860	0.6745	0.4101	0.6745
20	0.95	0.3562	0.1800	0.4706	0.00	0.1676	0.00
30	0.95	0.2112	0.2996	0.3228	0.3145	0.0682	0.0973
5	0.90	0.5901	0.8969	0.6860	0.6799	0.4048	0.6105
10	0.90	0.3476	0.00	0.7405	0.4553	0.1636	0.00
20	0.90	0.1200	0.00	0.2214	0.2085	0.0266	0.00
30	0.90	0.0411	0.001	0.1042	0.00	0.0043	0.00

4. 결 론

본 논문에서는 생체 면역시스템의 자기와 비자기를 구분하여 식별하는 세포독성 T세포의 특징을 이용하여 컴퓨터에서 자기 파일과 비자기 파일을 구분할 수 있는 T세포를 모델링하였다. 모델링 한 T세포는 MHC 인식부와 항원 인식부로 구성되어 있으며, 각각의 인식부를 이용해서 비자기 파일에 적용시켜 자기-비자기 식별이 가능함을 보였다. 결론적으로 항원 식별부만을 갖는 경우보다 본 논문에서 제안한 MHC 인식부만을 갖는 경우에 인식부의 개수가 많아지면 식별 능력이 좋아짐을 알 수 있었고, 특히 이 두 가지의 인식부를 모두 갖는 경우에는 단독으로 사용했을 경우보다 식별 능력이 아주 우수하다는 것을 확인하였다.

참 고 문 헌

- [1] I. Roitt, J. Brostoff, D. Male, Immunology, 4th edition, Mosby, 1996
- [2] 이동욱, 심귀보, "인공 면역 시스템 기반 자율분산로봇 시스템의 협조 전략과 군행동", 한국 퍼지 및 지능시스템학회 논문지 vol. 9, no. 6 pp. 627-633, 1999.
- [3] Stephanie Forrest, Lawrence Allen, Alan S. Perelson, Rajesh Cherkuri "Self-Nonself Discrimination in a Computer" IEEE Symposium on Research in Security and Privacy, 1994.
- [4] Ja Beom Gu, Dong Wook Lee, Kwee-Bo Sim, Se Hyun Park, "An Immunity-based Security Layer Model", Genetic and Evolutionary Computation Conference Workshop Program pp. 47-48, 2000.
- [5] Dipankar Dasgupta, "An Immune Agent Architecture for

Intrusion Detection", Genetic and Evolutionary Computation Conference Workshop Program pp. 42-44, 2000.

- [6] Paul K. Harmer, Gary B. Lamont, "An Agent Based Architecture for a Computer Virus Immune System", *Genetic and Evolutionary Computation Conference Workshop Program* pp. 45-46, 2000.
- [7] 구자범, 이동욱, 박세현, 심귀보, "생체 면역시스템 기반의 새로운 보안 항체 계층 모델", *한국 퍼지 및 지능시스템학회 논문지* vol. 10, no. 2, pp. 122-128, 2000.



심 귀 보 (Kwee-Bo Sim)

1984년 : 중앙대학교 전자공학과 학사
 1986년 : 동 대학원 전자공학과 석사
 1990년 : The University of Tokyo 전자공학과 박사
 1991년~현재 : 중앙대학교 전자전기공학부 교수
 1997년~현재 : 한국퍼지 및 지능시스템학회 논문지 편집위원장
 1997년~현재 : 한국퍼지 및 지능시스템학회 편집이사
 2000년~현재 : 제어자동화시스템공학회 국문논문지 편집위원
 2001년~현재 : 대한전기학회 제어 및 시스템 부문회 학술이사

관심분야 : 인공생명, 진화연산, 지능로봇시스템, 뉴로-퍼지 및 소프트 컴퓨팅, 자율분산시스템, 로봇비전, 진화 하드웨어, 인공 면역계 등

Phone : +82-2-820-5319
 Fax : +82-2-817-0553
 E-mail : kbsim@cau.ac.kr
 URL : <http://rics.cie.cau.ac.kr>

저 자 소 개



선 상 준 (Sang-Joon Sun)

2000년 : 중앙대학교 전자전기공학부 학사
 2000년~현재 : 중앙대학교 전자전기공학부 석사과정

관심분야 : 인공면역계, 인터넷보안 등.



심 재 윤 (Jae-Yoon Sim)

1985년 : 중앙대학교 전자공학과 학사
 1985년~현재 : 우진네트(주)
 2000년~현재 : 중앙대학교 정보산업 대학원 석사과정

관심분야 : 네트워크 보안, 생체인식 등.