Negative Selection 알고리즘 기반 이상탐지기를 이용한 이상행위 탐지

김미선* · 서재현*
*목포대학교

Anomaly behavior detection using Negative Selection algorithm based anomaly detector

Mi-Sun Kim* · Jae-Hyeon Seo*
*Mokpo National University
E-mail : misun@mokpo.ac.kr

요 약

인터넷의 급속한 확장으로 인해 네트워크 공격기법의 패러다임의 변화가 시작되었으며 새로운 공격 형태가 나타나고 있으나 대부분의 침입 탐지 기술은 오용 탐지 기술을 기반으로 하는 시스템이 주를 이루고 있어 알려진 공격 유형만을 탐지하고, 새로운 공격에 능동적인 대응이 어려운 실정이다. 이에 새로운 공격 유형에 대한 탐지력을 높이기 위해 인체 면역 메커니즘을 적용하려는 시도들이 나타나고 있다. 본 논문에서는 데이터 마이닝 기법을 이용하여 네트워크 패킷에 대한 정상 행위 프로파일을 생성하고 생성된 프로파일을 자기공간화 하여 인체면역계의 자기, 비자기 구분기능을 이용해 자기 인식 알고리즘을 구현하여 이상행위를 탐지하고자 한다. 자기인식 알고리즘의 하나인 Negative Selelction Algorithm을 기반으로 anomaly detector를 생성하여 자기공간을 모니터하여 변화를 감지하고 이상행위를 검출한다. DARPA Network Dataset을 이용하여 시뮬레이션을 수행하여 침입 탐지율을 통해 알고리즘의 유효성을 검증한다.

ABSTRACT

Change of paradigm of network attack technique was begun by fast extension of the latest Internet and new attack form is appearing. But, Most intrusion detection systems detect informed attack type because is doing based on misuse detection, and active correspondence is difficult in new attack. Therefore, to heighten detection rate for new attack pattern, visibilitys to apply human immunity mechanism are appearing.

In this paper, we create self-file from normal behavior profile about network packet and embody self recognition algorithm to use self-nonself discrimination in the human immune system to detect anomaly behavior. Sense change because monitors self-file creating anomaly detector based on Negative Selelction Algorithm that is self recognition algorithm's one and detects anomaly behavior. And we achieve simulation to use DARPA Network Dataset and verify effectiveness of algorithm through the anomaly detection rate.

키워드
침입탐지, negative selection algorithm, 이상탐지, 인공면역시스템, anomaly detection algorithm

Ⅰ. 서 론

컴퓨터 시스템에 대한 공격의 형태가 매우 다양해지고 변형되어 컴퓨터 시스템 내의 자원이나 정보를 위협함에 따라 이러한 공격들로부터 시스템을 안전하게 보호하는 방법으로 생물계의 면역시스템을 모델링하는 연구가 진행되고 있다. 인체 면

역계는 외부 침입 물질 구분과 자기 복제등의 장치를 이용하여 외부침입으로부터 자신을 보호하는 기능을 갖는다. 이러한 인체 면역계의 특징인 항원 인식 기능은 자기세포의 확실한 인식을 통해서 다른 물질을 구분하는 자기, 비자기 인식 방법을 사용한다.

본 논문에서는 인체 면역 메카니즘을 이용하여 detector의 지속적인 갱신을 통해서 알려지지 않은 새로운 공격에 대한 탐지 방법을 연구하고자 한다. 네트워크 패킷에 대한 정상 행위 프로파일을 생성하고 생성된 프로파일에 Negative Selelction Algorithm을 적용, anomaly detector를 생성하여 자기공간을 모니터하여 변화를 감지하고 이상행위를 검출하는 네트워크 기반의 침입 탐지를 연구하고자 한다. DARPA Network Dataset을 이용하여 시뮬레이션을 수행하여 침입 탐지율을 통해 알고리즘의 유효성을 검증한다.

II. 인공면역시스템

인체 면역계는 외부 유기체나 단백질에 대하여 생명체를 방어할 수 있는 시스템이다. 면역시스템은 항원을 인식할 수 있는 항체의 생성으로부터 시작되며 특별한 항원을 인식할 수 있는 임파구가 활성화되어 항원과의 바인딩을 통해 항원의 제거가 이루어진다. 면역계를 구성하는 기본 요소는 두가지 형태의 임파구로 B 세포와 T 세포이다. B 세포는 특별한 항원을 인식하여 바인딩 함으로써 항원을 제거하는 항체를 생산, 분비하는 역할을 하며 T 세포는 항원을 제거하거나, B 세포의 성장을 억제하거나 도와주는 역할, 또는 항원을 정상적으로 인식하지 못하는 B 세포를 제거하는 역할을 담당한다[4].

2.1 인체 면역 분석 모듈내 구성

· B-cell

Bone Marrow(골수)에서 생산되는 것으로 항원을 바인딩할 수 있는 세포로 혈청속에서 항독 작용 혹은 살균작용을 하는 항체 세포이다.

· T-cell

Thyms(흉선) 에서 생산되며 B-cell이 항원을 바인딩하는것을 도우며 B-cell의 효율성여부를 판단하여 B-cell 을 제거하는 역할을 하며 B-cell 의 성장을 억제하거나 도와준다.

· Antigen(항원)

시스템에 침투하는 항원으로 외부로부터의 이물질이나 세균세포를 의미한다.

· Negative Selection

침입에 대해 정상적으로 반응을 보이지 못하는 cell 을 신속히 제거하는 역할을 하며 항원의 인식에 있어서 자기를 항원으로 인식하는 것을 배제하기 위한 방법이다. 항원으로 MHC 단백질을 인식하지 못하게 하기 위해 면역세포에 MHC 단백질을 결합시켰을 때 항원수용체가 부정적인 선택을 하는 세포만을 선택한다.

· Positive Selection

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

· Clonal Selection

효과적인 탐지능력을 지닌 cell을 복제한다. 항원은 지속적인 변화를 하기 때문에 항원 탐지 효율은 복제 선택을 통한 B-cell 항체의 진화과정에 의해 유지된다. 이러한 메모리 효과 때문에 이전에 구분된 항원은 다음번 침입에 더욱 빠르게 탐색될 수 있다.

인체 면역시스템을 컴퓨터 면역시스템으로 모듈화하기 위해서 각 구성요소를 그림1과 같이 매핑할 수 있다[3].

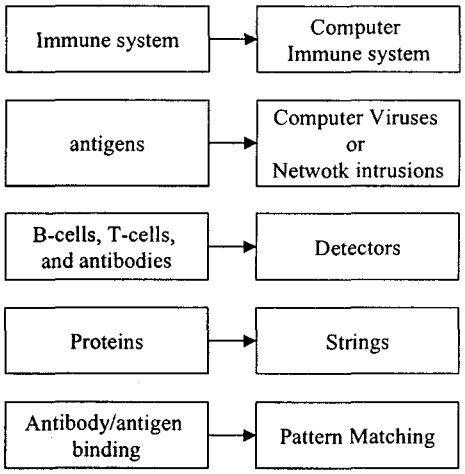

그림 1. 인체 vs 컴퓨터 면역시스템 모듈 매핑

III. 네트워크 데이터의 정상행위 프로파일

네트워크 기반의 침입 탐지는 네트워크 데이터인 패킷 헤더 정보를 이용하여 이상이나 오용침입을 탐지한다. 본 논문에서는 TCP/IP 기반의 서비스에 대한 네트워크의 패킷 헤더 정보를 이용하여 서비스별로 분류하며, 네트워크 서비스별로 정상 행위를 프로파일링하여 이상 침입을 탐지한다. 네트워크의 정상 데이터 패턴을 생성하기 위한 특징 선택을 한다. 본 논문에서는 네트워크 서비스, 세션을 구성하는 패킷수, TCP 통신 절차, 리셋연결, 단방향 통신 그리고 패킷의 플래그의 분포의 12가

지의 특징으로 구성한다. 본 논문에서의 정상 프로파일을 위한 패킷의 특징은 표1과 같이 구성된다.

표 1. 패킷 정상프로파일 특징 구성

특징선택	형식	내용
서비스	Integer	네트워크 서비스 종류
세션크기	Integer	세션의 패킷 갯수
TCP 절차	비트값	TCP 통신절차 수행여부
리셋	비트값	리셋 통신절차 수행여부
단방향 통신	비트값	단방향 통신 수행여부
No Flag	Integer	No Flag의 수
F	Integer	F의 수
S	Integer	S의 수
R	Integer	R의 수
P	Integer	P의 수
ack	Integer	ack의 수
U	Integer	U의 수

본 논문에서는 네트워크 패킷 중에서 FTP, SSH, Telnet, SMTP의 4가지 네트워크 서비스에 대한 정상행위 패턴을 구성하였다. ftp 서비스에 대한 정상행위 패턴을 그림2와 같다.

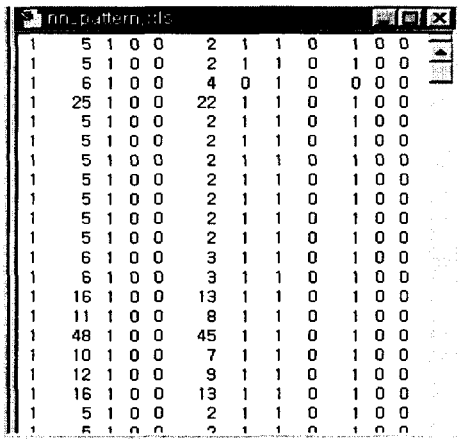

그림 2. ftp 서비스에 대한 정상행위 프로파일

IV. Anomaly Detection Algorithm

Negative Selection은 인공 면역 시스템에서 침입에 대해 정상적으로 반응을 보이지 못하는 cell을 신속히 제거하는 역할을 하며 항원의 인식에 있어서 자기를 항원으로 인식하는 것을 배제하기 위한 방법이다. 생체 면역계의 면역 세포 생성원리 중의 하나인 Negative Selection을 이용하여 자기 인식 알고리즘을 구현한 경우가 Anomaly Detection Algorithm이다[1]. 본 논문에서는 모니터링된 패킷에 대해 정상행위 프로파일을 생성하고, 생성된 프로파일에 대해 Anomaly Detection Algorithm을 적용하여 이상을 탐지하고 지속적인 detector set의 갱신을 통해 새로운 유형의 침입에도 능동적으로 대응할 수 있도록 한다.

생성된 정상행위 프로파일에 대해 전처리 작업을 거쳐 self-file을 생성하고 self-file에 대해서 Negative-Selection기법을 매칭시켜 anomaly detector set을 구성하고 이를 자기-인식 알고리즘에 적용한다. 본 연구에서는 기존의 anomaly detector algorithm에 정상행위 패턴의 자동 생성과 self-file의 자동 갱신 과정을 추가하여 지속적으로 anomlay detector set을 갱신하여 더욱 신뢰성있는 이상 탐지가 되도록 한다. 그림3은 본 논문에서 수정된 Anomaly Detection Algorithm의 수행과정을 보여준다.

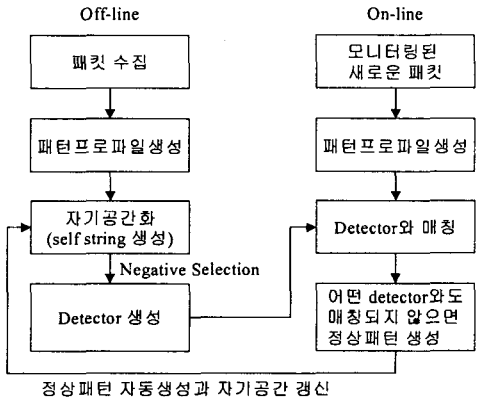

그림 3. 수정된 Anomaly detection algorithm 수행과정

본 논문에서는 모니터링된 패킷에 대해 정상행위 프로파일을 생성하고 이를 2진 스트링열로 변환하여 self-file을 생성, 랜덤하게 생성된 2진 스트링열과 매칭시켜 anomaly detector를 생성한다. 생성된 anomaly detector에 대해 모니터링된 새로운 데이터패턴과 매칭시켜 매칭되지 않으면 정상행위로 간주, 정상패턴을 자동생성, 자기공간을 갱신하고 매칭되는 경우는 이상행위로 간주한다. 그림4와 5는 detector set의 생성과정과 detector set을 이용한 이상탐지 과정을 보인다.

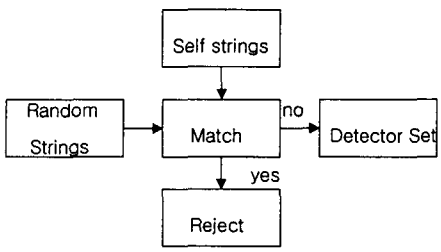

그림 4. 자기공간과의 매칭을 통한 detector 생성

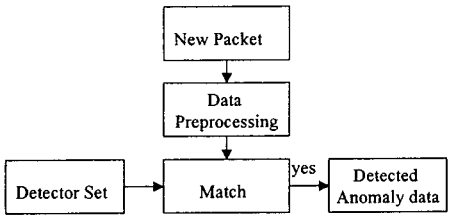

그림 5. detector set에 의한 Anomaly data 검출

본 논문에서는 정상행위 프로파일에 대해 anomaly detection algorithm을 적용하기 위하여 그림6과 같이 2진 스트링열로 구성된 self-file을 생성한다.

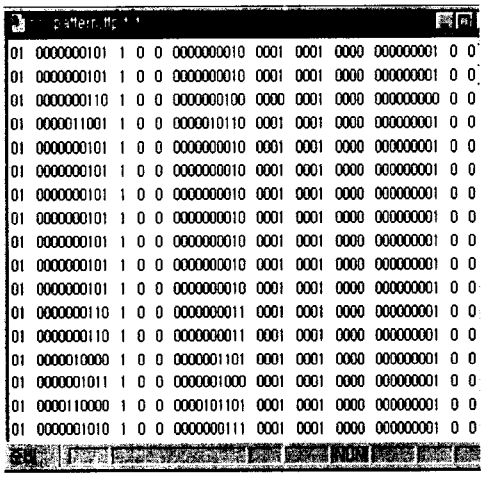

그림 6. ftp 정상행위 프로파일에 대한 self-file

anomaly detector는 self-file과 매칭되지 않는 스트링을 이용하여 구성한다. 두 개의 스트링사이의 완전한 매칭은 각 위치의 값이 일치하는 경우를 의미한다. 그러나 self-file의 크기가 커지게되면 일치하지 않는 스트링을 찾는 것이 어려워진다. 따라서 anomaly detection algorithm에서는 부분매칭기법을 이용해서 연속적으로 인접한 r개의 셀들이 매칭되는 경우를 두 스트링사이의 매칭으로 정의한다[2].

V. 시뮬레이션

본 논문에서는 시뮬레이션을 위하여 DARPA 2000년 NT 데이터 일부를 사용하였으며, 전체 훈련 데이터를 2000개로 샘플링하고 이 중 침입 데이터 유형은 ftp 서비스의 경우 20개의 공격 패킷이 포함 되어있다. self-file의 스트링의 길이를 50bit, self-file의 크기는 650으로 설정하였다. 매칭 한계값 r, anomaly detect의 수 Nr을 변경하여 시뮬레이션을 수행하였다. 시뮬레이션 결과는 표2에 나타냈다. 인식율은 r 값의 변화에 따라 달라질 수 있으며 r값이 커질수록 이상에 더욱 더 민감한 anomaly detector의 생성이 가능하다. 적절한 r값의 선택은 false positive률을 감소하는 신뢰성있는 이상 침입 탐지 효과를 가져올 수 있다.

표 2. ftp 서비스에 대한 공격 탐지 결과

r	Nr	anomaly detection
5	30	44.5
	50	47.3
10	30	31.2
	50	34.7
15	30	25.3
	50	29.2

VI. 결 론

대부분의 침입 탐지 기술은 알려진 공격 유형만을 탐지하고, 새로운 공격에 능동적인 대응이 어려운 실정이다. 본 논문에서는 인체 면역 메카니즘을 이용하여 네트워크 패킷에 대한 정상 행위 프로파일을 생성하고 생성된 프로파일에 Negative Selelction Algorithm을 적용, anomaly detector를 생성하여 자기공간을 모니터하여 변화를 감지하고 이상행위를 검출하는 네트워크 기반의 침입 탐지를 연구하였다. 지속적인 detector의 갱신으로 새로운 공격유형에 대한 탐지가 가능할 것으로 기대된다. 추후 Clonal Selection기법의 활용 등 탐지시간을 단축하여 실시간으로 처리할 수 있는 알고리즘의 보완이 요구된다.

참고문헌

[1] D.Dasgupta(edit), "Artificial Immune Systems and Their Applications", Springer, pp.262-275, 1999.

[2] S.Forrest, A.S.Perelson, L.Allen, R.and Cherukuri, "Self-Nonself Discrimination in a Computer", In Proceeding of the 1994 IEEE Symposium on Research in Security and Privacy, 1994.

[3] P.K.Harmer, P.D.Williams,G.H.Gunsch, G.B. Lamont, "An Artificial Immune System Architecture for Computer Security Applications", IEEE Transaction on Evolutionary Computation, VOL.6, NO.3, pp.252-280, 2002.

[4] J.W.Kim, P.J.Benfley, "The Human Immune System and Network Intrusion Detection", 7th European Congress on Intelligent Techinques and SoftComputing , pp.13-19,1999.