(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0081408
(43) 공개일자 2019년07월09일

(51) 국제특허분류(Int. Cl.)
H04L 29/06 (2006.01) *G06N 3/02* (2019.01)
(52) CPC특허분류
H04L 63/1425 (2013.01)
G06N 3/02 (2019.01)
(21) 출원번호 10-2017-0183924
(22) 출원일자 2017년12월29일
심사청구일자 2017년12월29일

(71) 출원인
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)
(72) 발명자
니샤트 모우라
서울특별시 서대문구 이화여대길 52, 이화여대 아산공학관 324호
도인실
인천광역시 연수구 센트럴로 194, 201동 2705호 (더샵센트럴파크2단지아파트)
채기준
서울특별시 강남구 강남대로136길 34
(74) 대리인
윤귀상

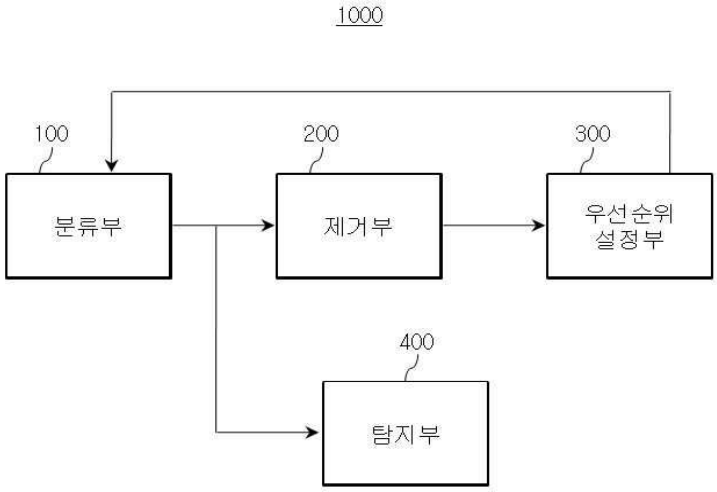
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **네트워크 침입 탐지 시스템 및 방법, 이를 수행하기 위한 기록매체**

(57) 요 약

본 발명에 따른 네트워크 침입 탐지 방법은 인공 신경망을 이용하여 네트워크로 입력되는 트래픽을 두 개의 클래스 중 어느 하나의 클래스로 분류하고, 분류된 클래스의 결과값에 따라 어느 하나의 클래스에 포함된 인스턴스를 삭제한 후 나머지 인스턴스에 대하여 2-클래스 분류 과정을 반복적으로 수행하여 단계적으로 네트워크 공격 트래픽의 종류를 탐지함으로써, 분석에 소요되는 트래픽의 양을 단계적으로 줄일 수 있으며, 분석 정확도 및 분석에 소모되는 시간을 단축시킬 수 있다.

대 표 도 - 도1



이 발명을 지원한 국가연구개발사업
과제고유번호 2016R1A2B4015899
부처명 과학기술정보통신부
연구관리전문기관 한국연구재단
연구사업명 기초연구사업(학술진흥)-중견연구자지원사업
연구과제명 CPS환경에서의신뢰성있는서비스제공을위한보안기법및공격대응방안
기 여 율 1/1
주관기관 이화여자대학교 산학협력단
연구기간 2016.06.01 ~ 2019.05.31

명 세 서

청구범위

청구항 1

두 개의 결과값을 출력하는 인공 신경망 기반의 분류 모델을 이용하여 네트워크로 입력되는 트래픽을 정상 클래스와 공격 클래스로 분류하는 분류 단계;

상기 정상 클래스에 대한 제1 결과값과 상기 공격 클래스에 대한 제2 결과값을 비교하여, 상기 제1 결과값이 상기 제2 결과값보다 작은 경우 상기 정상 클래스에 포함된 인스턴스를 제거하는 제거 단계;

미리 정해진 우선순위 정보에 따라 상기 공격 클래스에 포함된 인스턴스를 제1 공격 유형 및 상기 제1 공격 유형이 아닌 공격 유형으로 설정하고, 상기 제1 공격 유형에 대한 제1 클래스와 상기 제1 공격 유형이 아닌 공격 유형에 대한 제2 클래스를 생성하는 우선순위 설정 단계; 및

상기 분류 단계, 상기 제거 단계 및 상기 우선순위 설정 단계를 반복 수행하여 탐지된 공격 트래픽의 종류를 단계적으로 구분하는 탐지 단계를 포함하는, 네트워크 침입 탐지 방법.

청구항 2

제1항에 있어서,

상기 분류 모델은, 다층 퍼셉트론 기반의 인공 신경망을 기초로 모델링되고,

상기 분류 단계는, 상기 트래픽에 대한 특징 벡터를 추출하고, 복수의 상기 특징 벡터를 상기 인공 신경망에 입력하여 서로 다른 두 개의 출력값을 산출하는 것을 특징으로 하는, 네트워크 침입 탐지 방법.

청구항 3

제1항에 있어서,

상기 탐지 단계는,

상기 분류 모델을 이용하여 상기 공격 클래스에 포함된 인스턴스를 상기 제1 클래스와 상기 제2 클래스로 구분하고, 상기 제1 클래스에 대한 출력값이 상기 제2 클래스에 대한 출력값보다 작은 것으로 확인되면 상기 제1 클래스에 포함된 제1 공격 유형 인스턴스를 제거하고, 상기 우선순위 정보에 따라 상기 제2 클래스에 포함된 인스턴스를 제2 공격 유형 및 상기 제1 공격 유형과 상기 제2 공격 유형이 아닌 공격 유형으로 설정하며, 상기 제2 공격 유형에 대한 제3 클래스와 상기 제1 공격 유형과 상기 제2 공격 유형이 아닌 공격 유형에 대한 제4 클래스를 생성하는 것을 특징으로 하는, 네트워크 침입 탐지 방법.

청구항 4

제1항에 있어서,

상기 탐지 단계는,

상기 우선순위의 선순위에 대한 클래스의 결과값이 후순위에 대한 클래스의 결과값보다 큰 것으로 확인될 때까지 상기 분류 단계, 상기 제거 단계 및 상기 우선순위 설정 단계를 순차적으로 반복 수행하는 것을 특징으로 하는, 네트워크 침입 탐지 방법.

청구항 5

제1항에 있어서,

상기 우선순위 정보는,

반복 수행되는 상기 탐지 단계의 특정 단계에서 공격 트래픽의 종류를 단계적으로 구분하도록 서로 다른 공격 트래픽의 종류가 순차적으로 설정된 정보인 것을 특징으로 하는, 네트워크 침입 탐지 방법.

청구항 6

제5항에 있어서,

상기 우선순위 정보는,

DoS, Probe, U2R 및 R2L 순서로 설정되는 것을 특징으로 하는, 네트워크 침입 탐지 방법.

청구항 7

제1항 내지 제6항 중 어느 하나의 항에 따른 악 네트워크 침입 탐지 방법을 수행하기 위한, 컴퓨터 프로그램이 기록된 컴퓨터로 판독 가능한 기록매체.

청구항 8

두 개의 결과값을 출력하는 인공 신경망 기반의 분류 모델을 이용하여 네트워크로 입력되는 트래픽을 정상 클래스와 공격 클래스로 분류하는 분류부;

상기 정상 클래스에 대한 제1 결과값과 상기 공격 클래스에 대한 제2 결과값을 비교하여, 상기 제1 결과값이 상기 제2 결과값보다 작은 경우 상기 정상 클래스에 포함된 인스턴스를 제거하는 제거부;

미리 정해진 우선순위 정보에 따라 상기 공격 클래스에 포함된 인스턴스를 제1 공격 유형 및 상기 제1 공격 유형이 아닌 공격 유형으로 설정하고, 상기 제1 공격 유형에 대한 제1 클래스와 상기 제1 공격 유형이 아닌 공격 유형에 대한 제2 클래스를 생성하는 우선순위 설정부; 및

상기 분류부, 상기 제거부 및 상기 우선순위 설정부의 동작이 반복 수행되는 과정에서 탐지된 공격 트래픽의 종류를 단계적으로 구분하는 탐지부를 포함하는, 네트워크 침입 탐지 시스템.

청구항 9

제8항에 있어서,

상기 탐지부는,

상기 분류 모델을 이용하여 상기 공격 클래스에 포함된 인스턴스를 상기 제1 클래스와 상기 제2 클래스로 구분하고, 상기 제1 클래스에 대한 출력값이 상기 제2 클래스에 대한 출력값보다 작은 것으로 확인되면 상기 제1 클래스에 포함된 제1 공격 유형 인스턴스를 제거하고, 상기 우선순위 정보에 따라 상기 제2 클래스에 포함된 인스턴스를 제2 공격 유형 및 상기 제1 공격 유형과 상기 제2 공격 유형이 아닌 공격 유형으로 설정하며, 상기 제2 공격 유형에 대한 제3 클래스와 상기 제1 공격 유형과 상기 제2 공격 유형이 아닌 공격 유형에 대한 제4 클래스를 생성하는, 네트워크 침입 탐지 시스템.

청구항 10

제8항에 있어서,

상기 탐지부는,

상기 우선순위의 선순위에 대한 클래스의 결과값이 후순위에 대한 클래스의 결과값보다 큰 것으로 확인될 때까지 상기 분류 단계, 상기 제거 단계 및 상기 우선순위 설정 단계를 순차적으로 반복 수행하는, 네트워크 침입 탐지 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 네트워크 침입 탐지 시스템 및 방법, 이를 수행하기 위한 기록매체에 관한 것으로서, 더욱 상세하게는 인공신경망을 이용하여 네트워크 공격 트래픽을 효율적으로 탐지하는 네트워크 침입 탐지 시스템 및 방법, 이를 수행하기 위한 기록매체에 관한 것이다.

배 경 기 술

[0003] 네트워크와 관련된 분야의 기술이 발전하면서, 네트워크를 공격하는 트래픽을 감지하는 침입 탐지 시스템(Intrusion Detection System, IDS)의 중요성이 대두되고 있다.

[0004] 이러한 침입 탐지 시스템을 발전시기키 위해 다양한 형태의 머신 러닝 기술이 접목되고 있으며, 딥러닝이 사회적 이슈로 부각되면서 인공 신경망을 이용하여 네트워크 공격 트래픽을 감지하는 기술들이 공개되고 있다. 인공신경망은 입력 데이터를 복수의 계층으로 구성된 레이어로 입력시켜 이에 따른 결과를 이용하거나 학습하여 공격 데이터를 감지할 수 있다는 점에서 각광받고 있다.

[0005] 하지만, 종래 기술에 따른 침입 탐지 시스템은 고속 네트워크 환경에서 트래픽을 처리하는 데 어려움을 겪고 있다. 특히, 공격자는 공격을 수행하는 동안 의미없는 패킷을 전송함으로써 침입 탐지 시스템의 과부하를 유발하여 공격 의도를 숨길 수 있다. 따라서, 대용량의 트래픽을 빠른 시간에 내에 처리하면서도 공격 트래픽을 정확하게 구분할 수 있는 기술의 필요성이 부각되고 있는 실정이다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 한국등록특허 제10-0738537호

(특허문헌 0002) 한국등록특허 제10-0351306호

발명의 내용

해결하려는 과제

[0008] 본 발명의 일측면은 네트워크로 입력되는 트래픽을 분석하여 네트워크 침입을 탐지함에 있어, 처리가 요구되는 트래픽의 양을 감소시기면서도 분석 속도 및 정확도는 향상될 수 있는 네트워크 침입 탐지 시스템 및 방법, 이를 수행하기 위한 기록매체를 제공한다.

[0009] 본 발명의 기술적 과제는 이상에서 언급한 기술적 과제로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0011] 본 발명의 일 실시예에 따른 네트워크 침입 탐지 방법은, 두 개의 결과값을 출력하는 인공 신경망 기반의 분류 모델을 이용하여 네트워크로 입력되는 트래픽을 정상 클래스와 공격 클래스로 분류하는 분류 단계, 상기 정상

클래스에 대한 제1 결과값과 상기 공격 클래스에 대한 제2 결과값을 비교하여, 상기 제1 결과값이 상기 제2 결과값보다 작은 경우 상기 정상 클래스에 포함된 인스턴스를 제거하는 제거 단계, 미리 정해진 우선순위 정보에 따라 상기 공격 클래스에 포함된 인스턴스를 제1 공격 유형 및 상기 제1 공격 유형이 아닌 공격 유형으로 설정하고, 상기 제1 공격 유형에 대한 제1 클래스와 상기 제1 공격 유형이 아닌 공격 유형에 대한 제2 클래스를 생성하는 우선순위 설정 단계 및 상기 분류 단계, 상기 제거 단계 및 상기 우선순위 설정 단계를 반복 수행하여 탐지된 공격 트래픽의 종류를 단계적으로 구분하는 탐지 단계를 포함한다.

[0012] 상기 분류 모델은, 다층 퍼셉트론 기반의 인공 신경망을 기초로 모델링되고, 상기 분류 단계는, 상기 트래픽에 대한 특징 벡터를 추출하고, 복수의 상기 특징 벡터를 상기 인공 신경망에 입력하여 서로 다른 두 개의 출력값을 산출하는 것을 특징으로 할 수 있다.

[0013] 상기 탐지 단계는, 상기 분류 모델을 이용하여 상기 공격 클래스에 포함된 인스턴스를 상기 제1 클래스와 상기 제2 클래스로 구분하고, 상기 제1 클래스에 대한 출력값이 상기 제2 클래스에 대한 출력값보다 작은 것으로 확인되면 상기 제1 클래스에 포함된 제1 공격 유형 인스턴스를 제거하고, 상기 우선순위 정보에 따라 상기 제2 클래스에 포함된 인스턴스를 제2 공격 유형 및 상기 제1 공격 유형과 상기 제2 공격 유형이 아닌 공격 유형으로 설정하며, 상기 제2 공격 유형에 대한 제3 클래스와 상기 제1 공격 유형과 상기 제2 공격 유형이 아닌 공격 유형에 대한 제4 클래스를 생성하는 것을 특징으로 할 수 있다.

[0014] 상기 탐지 단계는, 상기 우선순위의 선순위에 대한 클래스의 결과값이 후순위에 대한 클래스의 결과값보다 큰 것으로 확인될 때까지 상기 분류 단계, 상기 제거 단계 및 상기 우선순위 설정 단계를 순차적으로 반복 수행하는 것을 특징으로 할 수 있다.

[0015] 상기 우선순위 정보는, 반복 수행되는 상기 탐지 단계의 특정 단계에서 공격 트래픽의 종류를 단계적으로 구분하도록 서로 다른 공격 트래픽의 종류가 순차적으로 설정된 정보인 것을 특징으로 할 수 있다.

[0016] 상기 우선순위 정보는, DoS, Probe, U2R 및 R2L 순서로 설정되는 것을 특징으로 할 수 있다.

[0017] 또한, 본 발명에 따른 네트워크 침입 탐지 방법을 수행하기 위한, 컴퓨터 프로그램이 기록된 컴퓨터로 판독 가능한 기록매체일 수 있다.

[0018] 또한, 본 발명의 일 실시예에 따른 네트워크 침입 탐지 시스템은, 두 개의 결과값을 출력하는 인공 신경망 기반의 분류 모델을 이용하여 네트워크로 입력되는 트래픽을 정상 클래스와 공격 클래스로 분류하는 분류부, 상기 정상 클래스에 대한 제1 결과값과 상기 공격 클래스에 대한 제2 결과값을 비교하여, 상기 제1 결과값이 상기 제2 결과값보다 작은 경우 상기 정상 클래스에 포함된 인스턴스를 제거하는 제거부, 미리 정해진 우선순위 정보에 따라 상기 공격 클래스에 포함된 인스턴스를 제1 공격 유형 및 상기 제1 공격 유형이 아닌 공격 유형으로 설정하고, 상기 제1 공격 유형에 대한 제1 클래스와 상기 제1 공격 유형이 아닌 공격 유형에 대한 제2 클래스를 생성하는 우선순위 설정부 및 상기 분류부, 상기 제거부 및 상기 우선순위 설정부의 동작이 반복 수행되는 과정에서 탐지된 공격 트래픽의 종류를 단계적으로 구분하는 탐지부를 포함한다.

[0019] 상기 탐지부는, 상기 분류 모델을 이용하여 상기 공격 클래스에 포함된 인스턴스를 상기 제1 클래스와 상기 제2 클래스로 구분하고, 상기 제1 클래스에 대한 출력값이 상기 제2 클래스에 대한 출력값보다 작은 것으로 확인되면 상기 제1 클래스에 포함된 제1 공격 유형 인스턴스를 제거하고, 상기 우선순위 정보에 따라 상기 제2 클래스에 포함된 인스턴스를 제2 공격 유형 및 상기 제1 공격 유형과 상기 제2 공격 유형이 아닌 공격 유형으로 설정하며, 상기 제2 공격 유형에 대한 제3 클래스와 상기 제1 공격 유형과 상기 제2 공격 유형이 아닌 공격 유형에 대한 제4 클래스를 생성할 수 있다.

[0020] 상기 탐지부는, 상기 우선순위의 선순위에 대한 클래스의 결과값이 후순위에 대한 클래스의 결과값보다 큰 것으로 확인될 때까지 상기 분류 단계, 상기 제거 단계 및 상기 우선순위 설정 단계를 순차적으로 반복 수행할 수 있다.

발명의 효과

[0022] 상술한 본 발명의 일측면에 따르면, 네트워크로 입력되는 트래픽이 공격 트래픽인 것으로 확인되는 경우, 공격 트래픽의 종류를 단계적으로 판단함으로써 분석에 소요되는 트래픽의 양을 단계적으로 줄일 수 있으며, 이에 따라 분석 정확도 및 분석에 소모되는 시간을 단축시킬 수 있다.

도면의 간단한 설명

[0024] 도 1은 본 발명에 따른 악성코드 검출 장치의 개략적인 구성이 도시된 블록도이다.

도 2는 도 1의 네트워크 침입 탐지 시스템에서 이용되는 분류 모델의 일 예가 도시된 도면이다.

도 3 내지 도 5는 도 1의 네트워크 침입 탐지 시스템을 이용하여 공격 트래픽의 종류를 단계적으로 구분하는 일 예가 도시된 도면이다.

도 6 내지 도 8은 도 1의 네트워크 침입 탐지 시스템의 성능을 평가한 일 예가 도시된 도면이다.

도 9는 본 발명의 일 실시예에 따른 네트워크 탐지 방법의 개략적인 흐름이 도시된 도면이다.

발명을 실시하기 위한 구체적인 내용

[0025] 후술하는 본 발명에 대한 상세한 설명은, 본 발명이 실시될 수 있는 특정 실시예를 예시로서 도시하는 첨부 도면을 참조한다. 이들 실시예는 당업자가 본 발명을 실시할 수 있기에 충분하도록 상세히 설명된다. 본 발명의 다양한 실시예는 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 여기에 기재되어 있는 특정 형상, 구조 및 특성은 일 실시예와 관련하여 본 발명의 정신 및 범위를 벗어나지 않으면서 다른 실시예로 구현될 수 있다. 또한, 각각의 개시된 실시예 내의 개별 구성요소의 위치 또는 배치는 본 발명의 정신 및 범위를 벗어나지 않으면서 변경될 수 있음이 이해되어야 한다. 따라서, 후술하는 상세한 설명은 한정적인 의미로서 취하려는 것이 아니며, 본 발명의 범위는, 적절하게 설명된다면, 그 청구항들이 주장하는 것과 균등한 모든 범위와 더불어 첨부된 청구항에 의해서만 한정된다. 도면에서 유사한 참조부호는 여러 측면에 걸쳐서 동일하거나 유사한 기능을 지칭한다.

[0026] 이하, 도면들을 참조하여 본 발명의 바람직한 실시예들을 보다 상세하게 설명하기로 한다.

[0027] 도 1은 본 발명의 일 실시예에 따른 네트워크 침입 탐지 시스템(1000)의 개략적인 구성이 도시된 도면이다.

[0028] 본 발명에 따른 네트워크 침입 탐지 시스템(1000)은 네트워크 기반의 침입 탐지 시스템(Network-Based Intrusion Detection System, NIDS)로, 네크워크 상에 구현되어 독립된 시스템으로 운용될 수 있다.

[0029] 구체적으로, 본 발명의 일 실시예에 따른 네트워크 침입 탐지 시스템(1000)은 분류부(100), 제거부(200), 우선순위 설정부(300) 및 탐지부(400)를 포함한다.

[0030] 분류부(100)는 분류 모델을 이용하여 네트워크로 입력되는 트래픽을 두 개의 클래스로 분류할 수 있다. 분류 모델은 다층 퍼셉트론(Multi-Layer Perceptron, MLP) 기반의 인공 신경망을 이용하여 모델링 될 수 있다. 이와 관련하여, 도 2를 함께 참조하여 설명하기로 한다.

[0031] 도 2는 본 발명에 따른 네트워크 침입 탐지 시스템(1000)에서 이용되는 분류 모델의 일 예가 도시된 도면이다.

[0032] 다층 퍼셉트론 기반의 인공 신경망은 입력 계층, 은닉 계층 및 출력 계층으로 구성될 수 있다. 입력 계층은 인공 신경망을 통해 분석하고자 하는 데이터를 수신하는 계층이다. 은닉 계층은 중개 계층으로도 불리우며, 입력 계층의 입력 노드로부터 수신되는 특징값에 가중치(wij, wjk)가 반영된 값을 수신하는 은닉 노드들로 구성될 수 있다. 출력 계층은 은닉 계층으로부터 전달되는 값들을 출력하는 계층으로, 본 실시예에서는 두 개의 출력 노드를 포함할 수 있다. 이에 따라, 분류 모델은 입력 데이터가 A일 확률을 나타내는 제1 결과값과, 입력 데이터가 B일 확률을 나타내는 제2 결과값을 출력할 수 있다.

[0033] 분류부(100)는 이와 같은 분류 모델을 이용하여 네트워크로 입력되는 트래픽을 구분할 수 있다. 구체적으로, 분류부(100)는 분류 모델에 입력 데이터를 입력시켜 이에 따른 결과값이 출력되도록 제어할 수 있다. 여기서, 입력 데이터는 트래픽에 대한 특징 벡터로 정의될 수 있다. 분류부(100)는 트래픽을 전처리하여 트래픽의 특성을 나타내는 특징 벡터를 추출할 수 있다. 본 실시예에서, 분류부(100)는 KDD 데이터셋에서 제공하는 특징들을 이용하여 특징 벡터를 추출하는 것으로 가정하여 설명하기로 한다. 하지만, 분류부(100)는 기 공지된 다양한 데이터 특징 추출 기법을 이용하여 특징 벡터를 추출할 수 있다.

[0034] 분류부(100)는 추출된 복수의 특징 벡터(x_i)를 입력 계층의 입력값으로 설정할 수 있다. 즉, 분류부(100)는 각각의 특징 벡터(x_i)를 입력 계층을 구성하는 각각의 입력 노드로 입력시킬 수 있다. 이후, 분류부(100)는 인공

신경망 기반의 분류 모델을 통해 출력되는 결과값에 따라 트래픽을 구성하는 어느 하나의 인스턴스를 두 개의 클래스 중 어느 하나의 클래스로 분류할 수 있다. 분류부(100)는 트래픽을 구성하는 모든 인스턴스에 대하여 상술한 과정을 반복 수행함으로써, 트래픽의 모든 인스턴스들을 제1 클래스 또는 제2 클래스로 그룹화할 수 있다. 이때, 제1 클래스와 제2 클래스는 서로 다른 우선순위를 가질 수 있다. 제1 클래스는 미리 설정된 우선순위 정보에 따라 설정될 수 있으며, 제2 클래스는 제1 클래스가 아닐 확률에 대한 클래스로 정의될 수 있다. 이에 대한 구체적인 내용은 후술하기로 한다.

[0035] 제거부(200)는 서로 다른 우선순위를 갖는 두 개의 클래스로 분류된 트래픽 중 어느 하나의 클래스에 포함된 인스턴스를 제거할 수 있다. 구체적으로, 제거부(200)는 제1 클래스에 대한 제1 결과값과, 제2 클래스에 대한 제2 결과값을 비교할 수 있다. 이때, 제거부(200)는 제1 결과값이 제2 결과값보다 큰 것으로 확인되면 해당 트래픽을 제1 클래스인 것으로 판단할 수 있다. 예를 들어, 제1 클래스가 정상 클래스이고, 제2 클래스가 공격 클래스로 설정된 경우, 제거부(200)는 정상 클래스에 대한 제1 결과값이 공격 클래스에 대한 제2 결과값보다 큰 것으로 확인되면 해당 트래픽을 정상 트래픽인 것으로 판단할 수 있다.

[0036] 반면, 제거부(200)는 제1 결과값이 제2 결과값보다 작은 것으로 확인되면, 해당 트래픽을 네트워크를 공격하는 트래픽인 것으로 판단할 수 있다. 이때, 제거부(200)는 제1 클래스, 즉 정상 클래스인 것으로 분류된 인스턴스를 이후의 탐지 과정에서 제외시킬 수 있다. 즉, 제거부(200)는 트래픽을 구성하는 복수의 인스턴스들 중 정상으로 분류된 인스턴스를 이후의 분석 과정에서 제외시킴으로써 분석 속도 및 분석할 데이터의 용량을 감소시킬 수 있다.

[0037] 우선순위 설정부(300)는 우선순위에 따라 설정된 제1 클래스의 결과값이 제2 클래스의 결과값보다 낮은 경우, 미리 설정된 우선순위 정보에 따라 제2 클래스에 포함된 인스턴스를 어떤 종류의 공격 트래픽으로 구분할지를 2-클래스 분류 방법으로 설정할 수 있다. 구체적인 일 예로, 우선순위 설정부(300)는 우선순위 정보를 참조하여 제거된 클래스에 포함된 공격 종류가 1순위로 설정된 공격 트래픽인 경우, 제거된 공격 트래픽의 다음 순위인 2순위로 설정된 공격 종류를 제1 클래스로 재설정할 수 있다. 그리고, 우선순위 설정부(300)는 제2 클래스를 2순위로 설정된 공격 종류와 다른 유형의 공격 종류(예컨대 제3, 4 순위)에 대한 클래스로 재설정할 수 있다.

[0038] 이후, 분류부(100)는 새롭게 설정된 제1 클래스와 제2 클래스에 대한 분류 과정을 다시 수행하고, 제거부(200)는 재설정된 제1 클래스와 제2 클래스에 대한 결과값을 비교하여 제2 클래스에 대한 결과값이 더 큰 것으로 확인되면, 해당 트래픽이 2순위로 설정된 공격 종류를 제1 클래스로 또 다시 설정하고, 제2 순위에 따른 공격 종류와 다른 유형의 공격 종류를 제2 클래스로 다시 설정하여 상술한 2-클래스 분류 과정을 반복 수행할 수 있다. 이와 같은 과정은 우선순위 설정부(300)에 의해 설정된 제1 클래스의 결과값이 제2 클래스의 결과값보다 큰 것으로 확인될 때까지 순차적으로 반복되어 수행될 수 있다.

[0039] 탐지부(400)는 분류부(100), 제거부(200) 및 우선순위 설정부(300)의 동작이 반복 수행되는 과정에서, 탐지된 공격 트래픽의 종류를 단계적으로 구분할 수 있다. 이와 관련하여, 도 3 및 도 4를 함께 참조하여 설명하기로 한다.

[0040] 도 3 내지 도 5는 도 1의 네트워크 침입 탐지 시스템(1000)을 이용하여 공격 트래픽의 종류를 단계적으로 구분하는 일 예가 도시된 도면이다. 도 3 및 도 5는 침입 탐지 시스템(1000)에서 공격 트래픽을 2-클래스로 구분하여 탐지하는 일 예가 도시된 개념도이고, 도 4는 도 3 및 도 5의 탐지 알고리즘을 코드화한 일 예가 도시된 도면이다.

[0041] 최초에, 네트워크 침입 탐지 시스템(1000)은 네트워크로 입력되는 트래픽을 정상 클래스(normal)와 공격 클래스(attack)로 분류할 수 있다. 즉, 분류부(100)는 트래픽을 구성하는 각각의 인스턴스에 대한 특징 벡터를 추출하고, 특징 벡터를 분류 모델에 입력하여 해당 인스턴스가 정상 인스턴스(normal)인지 혹은 비정상 인스턴스(공격 트래픽, attack)인지를 판단할 수 있다. 최초 단계에서, 네트워크 침입 탐지 시스템(1000)은 비정상 인스턴스가 어떠한 종류의 공격 트래픽인지에 대한 분류 과정은 무시한 채, 오직 해당 인스턴스가 정상인지 혹은 비정상인지 여부만을 2분법으로 분류할 수 있다.

[0042] 네트워크 침입 탐지 시스템(1000)은 공격 트래픽일 확률에 대한 결과값이 정상 트래픽일 확률에 대한 결과값보다 큰 경우(if c!=normal), 공격 트래픽으로 분류된 클래스에 대한 세부적인 분류 과정을 수행할 수 있다. 이때, 네트워크 침입 탐지 시스템(1000)은 정상 트래픽으로 판단된 인스턴스들(normal)은 이후의 탐지 대상에서 제외시키고, 공격 트래픽으로 분류된 인스턴스들(attack)에 대해서만 추가적인 탐지 과정을 수행할 수 있다.

[0043] 네트워크 침입 탐지 시스템(1000)은 미리 설정된 우선순위 정보에 따라 공격 트래픽을 이분법적으로 구분하기

위한 2-클래스를 설정할 수 있다.

[0044] 우선순위 정보는 네트워크를 공격하는 공격 트래픽을 종류별로 구분하고, 우선 순위로 검출되어야 할 공격 트래픽 순서가 설정된 정보일 수 있다. 본 실시예에서, 우선순위 정보는 DoS, Probe, U2R 및 R2L 순서로 설정될 수 있으나, 이에 한정되는 것은 아니며 DoS, Probe, U2R 및 R2L를 서로 다른 순서로 설정하거나 다른 공격 트래픽 종류를 더 반영하여 설정될 수도 있음은 물론이다.

[0045] Dos(Denial of Service attack)는 서비스 거부 공격으로도 불리우며, 비정상적으로 컴퓨터의 리소스를 고갈시켜 사용자가 인터넷 상에서 평소 잘 이용하던 자원에 대한 서비스를 더 이상 받지 못하게 하는 공격 유형을 일컫는다. 관련 기능에서는 소스 바이트, 패킷 속도 등이 포함될 수 있다. Probe는 네트워크에 대한 지식을 얻기 위해 무의미한 패킷을 전송함으로써 수행되는 공격 유형으로 정의된다. Probe 공격은 연결 지속 시간 또는 소스 바이트와 같은 특징으로 감지될 수 있다. R2L은 공격자가 원격 시스템에 대한 접근을 시도하는 원격 액세스 공격 유형이다. 관련 기능에는 연결 지속 시간, 서비스 요청 또는 실패한 로그인 시도가 포함될 수 있다. U2R은 공격자가 일반 계정에 로그인 한 다음 루트 관리자 액세스 권한을 획득하는 공격 유형으로, 생성된 파일 개수나 호출되는 쉘(shell) 프롬프트의 수와 같은 특징으로 식별될 수 있다.

[0046] 즉, 네트워크 침입 탐지 시스템(1000)은 미리 설정된 우선순위 정보에 따라 공격 트래픽이 1순위로 설정된 Dos일 확률과 Dos가 아닐 확률에 대한 두 개의 클래스를 설정할 수 있다.

[0047] 이후, 네트워크 침입 탐지 시스템(1000)은 분류 모델에 의해 특징 벡터로부터 분류되는 공격 트래픽이 Dos로 분류되는 제1 클래스에 대한 결과값과 공격 트래픽이 Dos가 아닌 다른 유형의 공격 트래픽으로 분류되는 제2 클래스(otherattack_1)에 대한 결과값을 산출할 수 있다. 네트워크 침입 탐지 시스템(1000)은 제2 클래스에 대한 결과값이 더 큰 것으로 확인되는 경우 Dos로 분류된 제1 클래스의 인스턴스들을 이후의 탐지 과정에서 제외시키고, 제2 클래스(otherattack_1)에 포함된 인스턴스들에 대하여 새로운 2-클래스 분류 과정을 다시 수행할 수 있다. 이 경우, 네트워크 침입 탐지 시스템(1000)은 제2 클래스(otherattack_1)에 포함된 인스턴스들이 우선순위 정보의 제2 순위인 Probe일 확률을 나타내는 새로운 제1 클래스와, Dos 및 Probe가 아닌 다른 유형의 공격 트래픽일 확률을 나타내는 새로운 제2 클래스(otherattack_2)로 클래스를 재구성할 수 있다.

[0048] 이와 같은 방법으로, 네트워크 침입 탐지 시스템(1000)은 상술한 과정들을 순차적으로 반복 수행하여 우선순위 정보에 따른 네트워크 공격 트래픽의 유형에 대한 분류 과정을 단계적으로 수행할 수 있다. 즉, 네트워크 침입 탐지 시스템(1000)은 미리 설정된 우선순위 정보에 따른 클래스의 결과값이 다른 클래스의 결과값보다 큰 것으로 확인될 때까지 상술한 과정들을 순차적으로 반복 수행할 수 있다.

[0049] 이 과정에서, 탐지부(400)는 트래픽으로부터 탐지되는 공격 트래픽의 종류를 단계적으로 구분할 수 있다. 탐지부(400)는 최초 단계에서 제외되는 정상 클래스(normal)에 포함된 인스턴스들을 정상 트래픽으로 판단할 수 있다. 또한, 탐지부(400)는 두 번째 단계에서 제외되는 제1 클래스에 포함된 인스턴스들을 Dos 트래픽으로 판단할 수 있다. 한편, 탐지부(400)는 우선순위에 따른 클래스에 대한 결과값이 다른 클래스에 대한 결과값보다 큰 것으로 확인되면, 두 클래스에 포함된 모든 인스턴스들을 우선순위에 따른 네트워크 공격 트래픽으로 판단할 수 있다. 예를 들어, 상술한 두 번째 반복 단계에서 DoS 클래스에 대한 결과값이 otherattack_1 클래스에 대한 결과값보다 큰 것으로 확인되면, 탐지부(400)는DoS 클래스 및 otherattack_1 클래스에 포함된 모든 인스턴스들을 Dos공격 트래픽인 것으로 판단할 수 있다.

[0050] 도 6 내지 도 8은 네트워크 침입 탐지 시스템(1000)의 성능 평가 결과를 나타내는 그래프이다. 구체적으로, 도 6은 클래스의 개수에 따른 학습 소요 시간을 나타내는 그래프이고, 도 7은 네트워크 공격 트래픽의 종류 개수에 따른 학습 소요 시간을 비교하는 그래프이고, 도 8은 본 실시예에 따른 네트워크 침입 탐지 시스템(1000)의 정확도를 종래 기술과 비교한 그래프이다.

[0051] 성능 평가를 위해, KDD99 데이터 세트에 포함된 41가지 특징 벡터를 사용하여 다양한 공격 유형에 대한 네트워크 침입 탐지 시스템(1000)의 학습 시간 및 탐지 정확도를 평가하였다. 탐지 대상으로, 정상적인 네트워크 트래픽 및 서로 다른 네 종류의 네트워크 공격 트래픽을 포함하는 1200개의 데이터 인스턴스를 사용했다. 또한, 네 가지 주요 네트워크 공격 트래픽의 모든 하위 클래스에 대한 샘플을 사용하였다. 표 1은 성능 평가를 위해 사용된 네 가지 주요 네트워크 공격 트래픽의 모든 하위 유형을 나타내는 도표이다.

표 1

[0052]

Attack class	Attack Types
DoS	Back, Land, Neptune, Pod, Smurf, Teardrop
Probe	Satan, Ipsweep, Nmap, Portsweep
R2L	Guess_Password, Ftp_write, Imap, Phf, Warezmaster
U2R	Loadmodule

[0054] 도 6에 도시된 바와 같이, 분류되는 클래스의 개수를 n개에서 2개로 줄일수록 학습 시간도 크게 감소함을 알 수 있다. 이는, 많은 수의 클래스를 하위 타입을 상위 레벨 타입으로 줄임으로써 학습 시간이 더욱 단축될 수 있음을 의미한다.

[0055] 도 7에 도시된 case1은 5개의 트래픽 종류(normal, dos, probe, R2L, U2R)가 포함된 경우이고, case2는 4개의 트래픽 종류(dos, probe, R2L, U2R)가 포함된 경우이고, case3은 3개의 트래픽 종류(probe, R2L, U2R)가 포함된 경우이고, case4는 2개의 트래픽 종류(R2L, U2R)가 포함된 경우, 트래픽을 2개의 클래스로 구분하는 본 발명에 따른 네트워크 침입 탐지 시스템(1000)과 트래픽을 n개의 클래스로 구분하는 비교 대상 시스템의 학습 시간을 비교한 그래프이다. 또한, 본 발명에 따른 네트워크 침입 탐지 시스템(1000)에서, case1은 정상 및 공격 클래스를 의미하고, case2는 Dos와 기타 공격 클래스를 의미하고, case3은 Probe와 기타 공격 클래스를 의미하며, case4는 R2L과 U2R 클래스를 의미한다. 도시된 바와 같이, 본 발명에 따른 네트워크 침입 탐지 시스템(1000)은 일반적인 n-클래스 매커니즘에 학습 속도 시간이 감소되는 장점이 있으며, 트래픽의 초기 개수를 증가시키는 경우 이러한 학습 시간의 차이는 더욱 큰 것을 확인할 수 있다. 이는 인스턴스 제거로 인한 데이터 탐색 시간이 감소됨에 따라 기인한 것이다.

[0056] 그리고, 도 8에 도시된 바와 같이 본 발명에 따른 네트워크 침입 탐지 시스템(1000)은 모든 case에 대하여 일반적인 n 클래스 분류 기술에 비해 정확도가 높은 것을 알 수 있다.

[0057] 도 9는 본 발명의 일 실시예에 따른 네트워크 탐지 방법의 개략적인 흐름이 도시된 도면이다.

[0058] 최초 단계에서, 네트워크 침입 탐지 시스템(1000)은 두 개의 결과값을 출력하는 인공 신경망 기반의 분류 모델을 이용하여 네트워크로 입력되는 트래픽을 정상 클래스와 공격 클래스로 분류할 수 있다(810).

[0059] 네트워크 침입 탐지 시스템(1000)은 정상 클래스에 대한 제1 결과값과 공격 클래스에 대한 제2 결과값을 비교하여(820), 제1 결과값이 제2 결과값보다 작은 경우(820의 YES) 정상 클래스에 포함된 인스턴스를 제거할 수 있다(830).

[0060] 그리고, 네트워크 침입 탐지 시스템(1000)은 제1 결과값이 제2 결과값보다 작은 것으로 확인되면, 제거된 인스턴스를 제외한 나머지 인스턴스들에 대하여, 미리 정해진 우선순위 정보에 따라 해당 인스턴스를 제1 공격 유형(Dos) 및 상기 제1 공격 유형이 아닌 공격 유형으로 설정하고, 상기 제1 공격 유형에 대한 제1 클래스와 상기 제1 공격 유형이 아닌 공격 유형에 대한 제2 클래스를 생성할 수 있다(830).

[0061] 네트워크 침입 탐지 시스템(1000)은 제1 클래스와 제2 클래스의 결과값을 서로 비교하여(840), 제1 클래스에 대한 제1 결과값이 제2 클래스에 대한 제2 결과값보다 작은 것으로 확인되면(840의 YES), Dos로 분류된 인스턴스들을 이후 비교 과정에서 제외시키고, 나머지 인스턴스들을 2순위로 설정된 Probe일 가능성으로 분류되는 새로운 제1 클래스 또는 Probe가 아닐 가능성으로 분류되는 새로운 제2 클래스로 설정하여(850), 각각의 클래스에 대한 결과값을 다시 비교할 수 있다(860).

[0062] 네트워크 침입 탐지 시스템(1000)은 새롭게 설정된 제1 클래스에 대한 제1 결과값이 새롭게 설정된 제2 클래스

에 대한 제2 결과값보다 작은 경우(860의 YES), Probe로 분류된 제1 클래스에 포함된 인스턴스들을 제거한 후 나머지 인스턴스들에 대하여 2-클래스 분류 과정을 재수행할 수 있다(870). 네트워크 침입 탐지 시스템(1000)은 재수행되는 결과값 비교 과정(880)에서 제1 결과값이 제2 결과값보다 작은 것으로 판단되면, 해당 트래픽을 U2R 트래픽으로 판단하고(890), 제1 결과값이 제2 결과값보다 큰 것으로 판단되면, 해당 트래픽을 R2L 트래픽으로 판단할 수 있다(881).

[0063] 이때, 상술한 처리 단계들을 수행하는 과정에서, 탐지된 공격 트래픽의 종류를 단계적으로 구분할 수 있다. 예를 들어, 820과정에서 제1 결과값이 제2 결과값보다 큰 경우(820의 NO), 네트워크 침입 탐지 시스템(1000)은 정상 클래스 및 공격 클래스에 포함된 모든 인스턴스들을 정상 트래픽인 것으로 판단할 수 있다(821). 또한, 네트워크 침입 탐지 시스템(1000)은 840단계에서 제1 결과값이 제2 결과값보다 큰 경우(830의 NO), 제1 클래스 및 제2 클래스에 포함된 모든 인스턴스들을 Dos 트래픽인 것으로 판단할 수 있다(841). 이와 유사하게, 네트워크 침입 탐지 시스템(1000)은 860단계에서 제1 결과값이 제2 결과값보다 큰 경우(860의 NO), 제1 클래스 및 제2 클래스에 포함된 모든 인스턴스들을 Probe 트래픽인 것으로 판단할 수 있다(841).

[0064] 이와 같은, 네트워크 침입 탐지 방법을 제공하는 기술은 애플리케이션으로 구현되거나 다양한 컴퓨터 구성요소를 통하여 수행될 수 있는 프로그램 명령어의 형태로 구현되어 컴퓨터 판독 가능한 기록 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능한 기록 매체는 프로그램 명령어, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다.

[0065] 상기 컴퓨터 판독 가능한 기록 매체에 기록되는 프로그램 명령어는 본 발명을 위하여 특별히 설계되고 구성된 것들이거니와 컴퓨터 소프트웨어 분야의 당업자에게 공지되어 사용 가능한 것일 수도 있다.

[0066] 컴퓨터 판독 가능한 기록 매체의 예에는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM, DVD 와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 ROM, RAM, 플래시 메모리 등과 같은 프로그램 명령어를 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다.

[0067] 프로그램 명령어의 예에는, 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드도 포함된다. 상기 하드웨어 장치는 본 발명에 따른 처리를 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0068] 이상에서는 실시예들을 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

부호의 설명

[0070] 1000: 네트워크 침입 탐지 시스템

100: 분류부

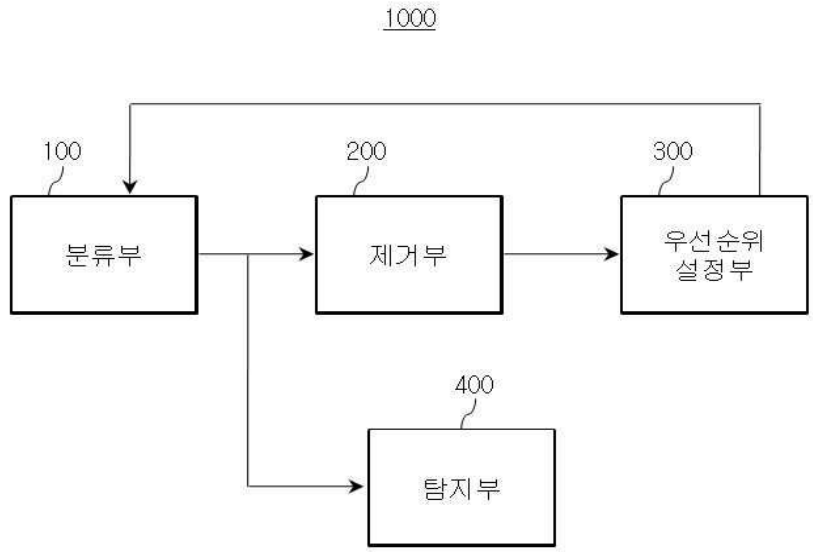
200: 제거부

300: 우선순위 설정부

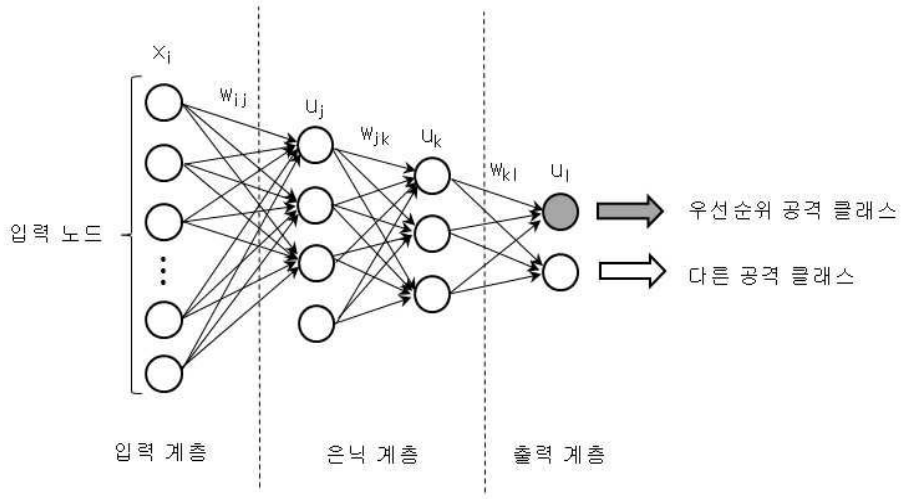
400: 탐지부

도면

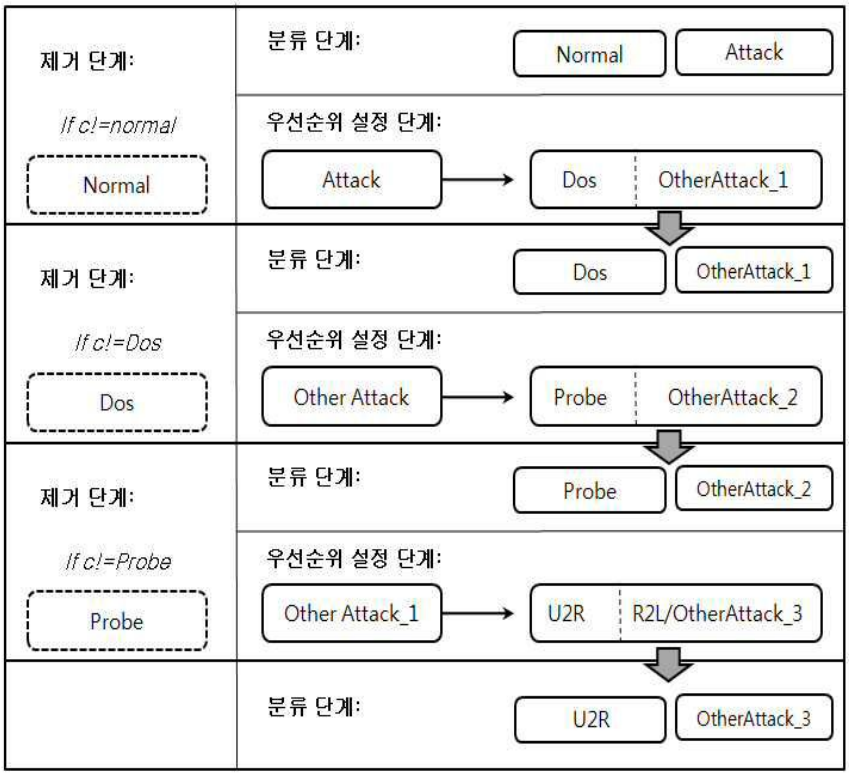
도면1


1000
100
200
300
분류부
제거부
우선순위
설정부
400
탐지부

도면2


x_i
w_{ij}
u_j
w_{jk}
u_k
w_{kl}
u_l
입력 노드
우선순위 공격 클래스
다른 공격 클래스
입력 계층
은닉 계층
출력 계층

도면3

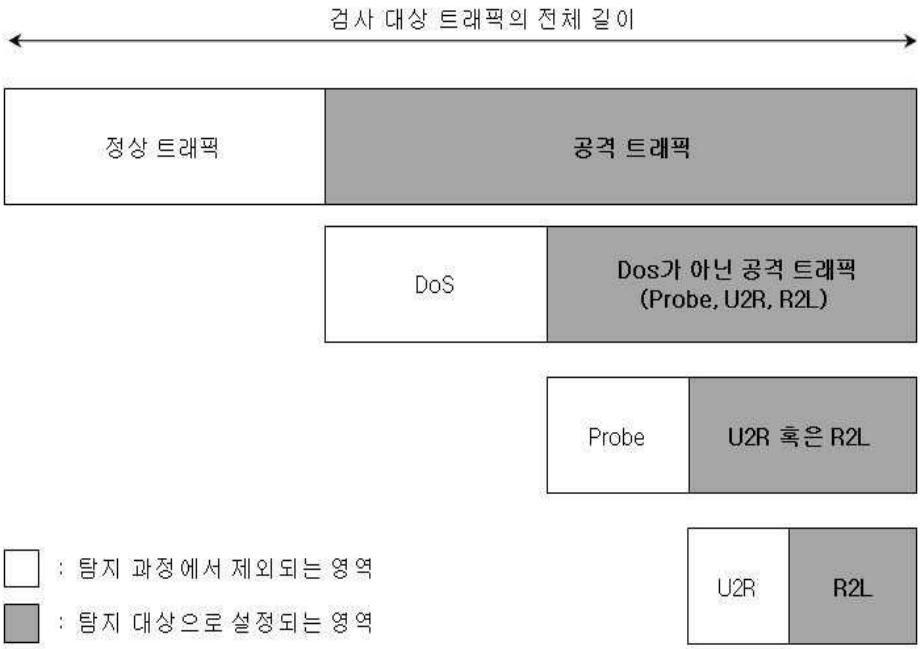


도면4

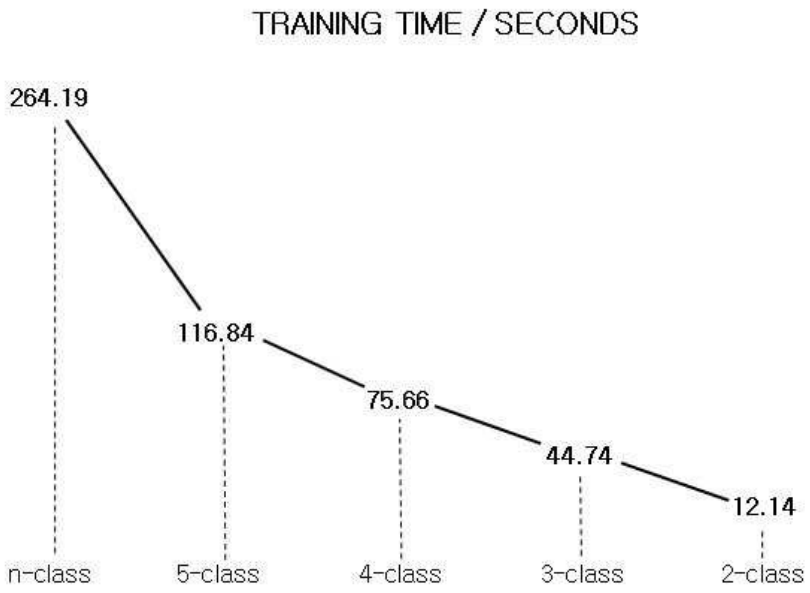
Algorithm 1 Evolving ANN Intrusion Detection System

```
class =sum of 2 or more classes;
normal = 1;
attack =0;
i = 1;
c= classification result;
run classification test;
if c == 0 && class ==TRUE then
    remove normal instances;
    dos = 1;
    other_attack_i = 0;
    if c == 0 && class ==TRUE then
        remove dos instances;
        probe =1;
        other_attack_{i+1} =0;
        if c == 0 && class==TRUE then
            remove probe instances;
            U2R =1;
            R2L/other_attack_{i+2} =0;
            if c == 0 && class ==TRUE then
                traffic is R2L;
            else
                traffic is U2R;
        else
            traffic is probe;
    else
        traffic is dos;
else
    traffic is normal;
```

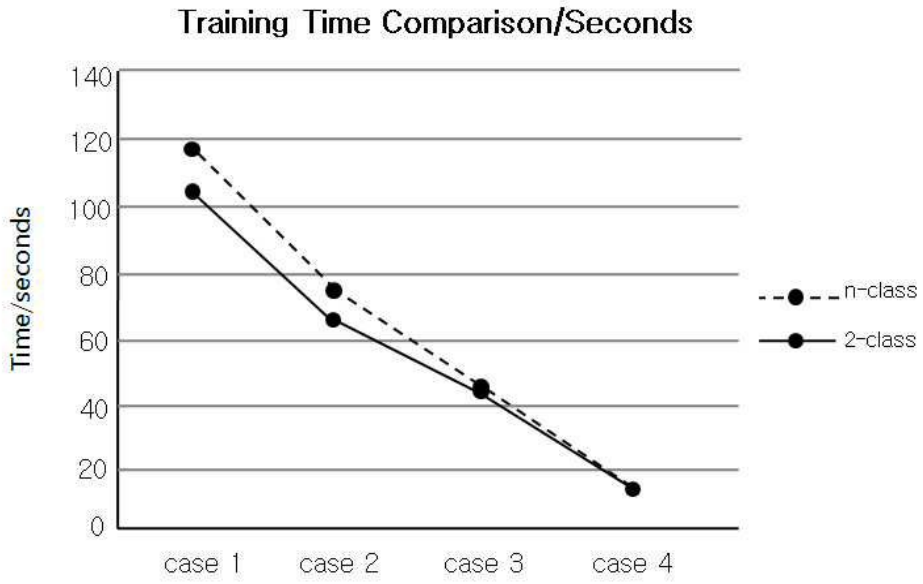
도면5


검사 대상 트래픽의 전체 길이
정상 트래픽
공격 트래픽
DoS
Dos가 아닌 공격 트래픽
(Probe, U2R, R2L)
Probe
U2R 혹은 R2L
: 탐지 과정에서 제외되는 영역
: 탐지 대상으로 설정되는 영역
U2R
R2L

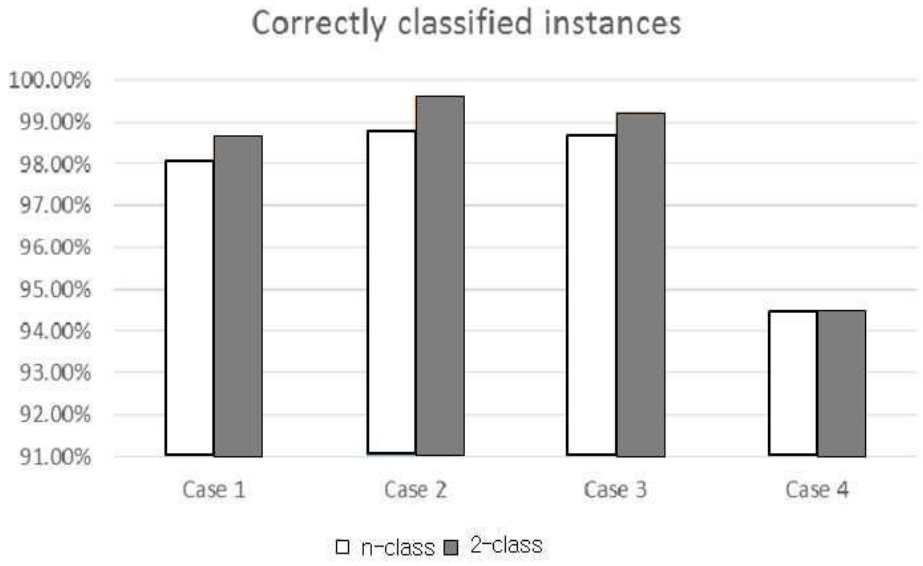
도면6


TRAINING TIME / SECONDS
264.19
116.84
75.66
44.74
12.14
n-class
5-class
4-class
3-class
2-class

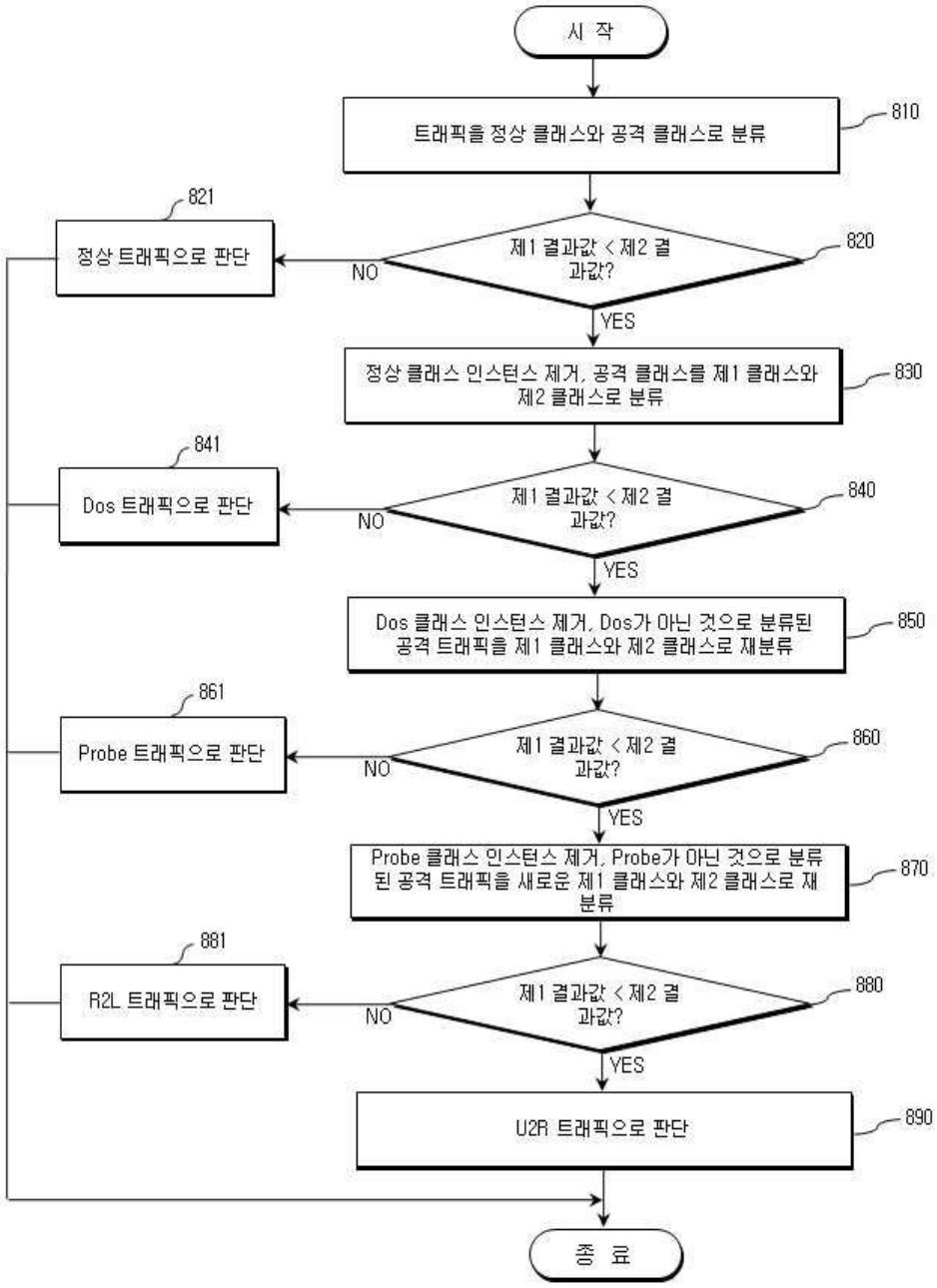
도면7


Training Time Comparison/Seconds
Time/seconds
140
120
100
80
60
40
20
0
case 1
case 2
case 3
case 4
n-class
2-class

도면8


Correctly classified instances
100.00%
99.00%
98.00%
97.00%
96.00%
95.00%
94.00%
93.00%
92.00%
91.00%
Case 1
Case 2
Case 3
Case 4
n-class
2-class

도면9


시 작
트래픽을 정상 클래스와 공격 클래스로 분류
810
제1 결과값 < 제2 결과값?
820
NO
정상 트래픽으로 판단
821
YES
정상 클래스 인스턴스 제거, 공격 클래스를 제1 클래스와 제2 클래스로 분류
830
제1 결과값 < 제2 결과값?
840
NO
Dos 트래픽으로 판단
841
YES
Dos 클래스 인스턴스 제거, Dos가 아닌 것으로 분류된 공격 트래픽을 제1 클래스와 제2 클래스로 재분류
850
제1 결과값 < 제2 결과값?
860
NO
Probe 트래픽으로 판단
861
YES
Probe 클래스 인스턴스 제거, Probe가 아닌 것으로 분류된 공격 트래픽을 새로운 제1 클래스와 제2 클래스로 재분류
870
제1 결과값 < 제2 결과값?
880
NO
R2L 트래픽으로 판단
881
YES
U2R 트래픽으로 판단
890
종 료