(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0088146
(43) 공개일자 2021년07월14일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) *G06N 3/04* (2006.01)
H04L 29/06 (2006.01)
(52) CPC특허분류
G06N 3/08 (2013.01)
G06N 3/04 (2013.01)
(21) 출원번호 10-2020-0001302
(22) 출원일자 2020년01월06일
심사청구일자 2020년01월06일

(71) 출원인
계명대학교 산학협력단
대구광역시 달서구 달구벌대로 1095, 계명대학교 산학협력관 201호(신당동)
(72) 발명자
박기현
대구광역시 달서구 도원로 45(도원동, 대곡강산타운)
이주화
대구광역시 수성구 수성로75길 16, 103동 2201호(수성동1가, 유성푸르나임)
(74) 대리인
특허법인 무한

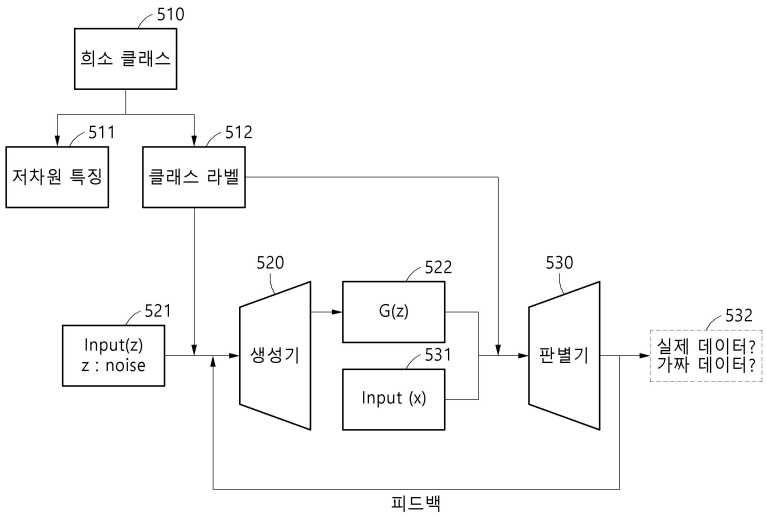
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 **AE-CGAN 모델 기반의 네트워크 침입 탐지 시스템 및 방법**

(57) 요 약

AE-CGAN 모델 기반의 네트워크 침입 탐지 시스템 및 방법이 개시된다. 본 발명의 일실시예에 따른 네트워크 침입 탐지 시스템의 학습 데이터 생성 방법은 원본 학습 데이터 세트로부터 희소 클래스의 특징을 추출하는 단계; 및 상기 희소 클래스의 특징을 데이터 생성 모델에 입력하여 상기 희소 클래스의 특징을 가지는 추가 데이터들을 생성하는 단계를 포함할 수 있다.

대 표 도



(52) CPC특허분류
H04L 63/1408 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 NRF-2018R1D1A1B07043982
부처명 한국연구재단
과제관리(전문)기관명 한국연구재단
연구사업명 기본연구(1년~5년)
연구과제명 접근이 어려운 열악한 환경에서의 사물인터넷 시스템 연구
기 여 율 1/1
과제수행기관명 한국연구재단
연구기간 2019.03.01 ~ 2020.02.29

명 세 서

청구범위

청구항 1

원본 학습 데이터 세트로부터 희소 클래스의 특징을 추출하는 단계; 및

상기 희소 클래스의 특징을 데이터 생성 모델에 입력하여 상기 희소 클래스의 특징을 가지는 추가 데이터들을 생성하는 단계

를 포함하는 학습 데이터 생성 방법.

청구항 2

제1항에 있어서,

상기 희소 클래스의 특징을 추출하는 단계는,

상기 원본 학습 데이터 세트에 포함된 희소 클래스의 특징을 차원 축소하여 데이터 차원이 감소된 희소 클래스의 특징을 추출하는 학습 데이터 생성 방법.

청구항 3

제2항에 있어서,

상기 희소 클래스의 특징을 추출하는 단계는,

오토 인코더를 이용하여 상기 원본 학습 데이터 세트에 포함된 희소 클래스의 특징을 차원 축소하고,

상기 오토 인코더는,

희소 클래스의 특징이 입력된 경우, 상기 오토 인코더의 마지막 코딩층을 희소 클래스의 핵심 특징으로 설정하고, 상기 오토 인코더의 디코더의 출력이 상기 핵심 특징에 대응되도록 학습되는 학습 데이터 생성 방법.

청구항 4

제1항에 있어서,

상기 데이터 생성 모델은,

노이즈와 클래스의 라벨을 이용하여 상기 희소 클래스의 특징으로 판별되도록 하기 위한 가짜 데이터를 생성하도록 학습되는 생성기 및

상기 가짜 데이터와 상기 희소 클래스의 특징, 및 상기 클래스의 라벨을 이용하여 상기 가짜 데이터를 가짜로 판별하도록 학습되는 판별기

를 포함하는 CGAN(Conditional Generative Adversarial Networks) 모델인 학습 데이터 생성 방법.

청구항 5

원본 학습 데이터 세트로부터 희소 클래스의 특징을 추출하는 단계;

상기 희소 클래스의 특징을 데이터 생성 모델에 입력하여 상기 희소 클래스의 특징을 가지는 추가 데이터들을 생성하는 단계; 및

상기 원본 학습 데이터 세트 및 상기 추가 데이터들을 이용하여 침입 탐지 여부를 분류하기 위한 분류기를 학습하는 단계

를 포함하는 침입 탐지 모델 학습 방법.

청구항 6

제5항에 있어서,

상기 희소 클래스의 특징을 추출하는 단계는,

오토 인코더를 이용하여 상기 원본 학습 데이터 세트에 포함된 희소 클래스의 특징을 차원 축소하고,

상기 오토 인코더는,

희소 클래스의 특징이 입력된 경우, 상기 오토 인코더의 마지막 코딩층을 희소 클래스의 핵심 특징으로 설정하고, 상기 오토 인코더의 디코더의 출력이 상기 핵심 특징에 대응되도록 학습되는 침입 탐지 모델 학습 방법.

청구항 7

제5항에 있어서,

상기 데이터 생성 모델은,

노이즈와 클래스의 라벨을 이용하여 상기 희소 클래스의 특징으로 판별되도록 하기 위한 가짜 데이터를 생성하도록 학습되는 생성기 및

상기 가짜 데이터와 상기 희소 클래스의 특징, 및 상기 클래스의 라벨을 이용하여 상기 가짜 데이터를 가짜로 판별하도록 학습되는 판별기

를 포함하는 CGAN(Conditional Generative Adversarial Networks) 모델인 침입 탐지 모델 학습 방법.

청구항 8

네트워크 데이터를 침입 탐지 모델의 분류기에 입력하는 단계; 및

상기 분류기의 출력에 따라 네트워크의 침입 여부를 탐지하는 단계

를 포함하고,

상기 분류기는,

원본 학습 데이터 세트에서 추출한 희소 클래스의 특징을 데이터 생성 모델에 입력하여 생성한 추가 데이터들 및 상기 원본 학습 데이터 세트를 이용하여 학습되는 네트워크 침입 탐지 방법.

청구항 9

제8항에 있어서,

상기 희소 클래스의 특징은,

오토 인코더를 이용하여 상기 원본 학습 데이터 세트에 포함된 특징을 차원 축소하여 추출되고,

상기 오토 인코더는,

희소 클래스의 특징이 입력된 경우, 상기 오토 인코더의 마지막 코딩층을 희소 클래스의 핵심 특징으로 설정하고, 상기 오토 인코더의 디코더의 출력이 상기 핵심 특징에 대응되도록 학습되는 네트워크 침입 탐지 방법.

청구항 10

제8항에 있어서,

상기 데이터 생성 모델은,

노이즈와 클래스의 라벨을 이용하여 상기 희소 클래스의 특징으로 판별되도록 하기 위한 가짜 데이터를 생성하도록 학습되는 생성기 및

상기 가짜 데이터와 상기 희소 클래스의 특징, 및 상기 클래스의 라벨을 이용하여 상기 가짜 데이터를 가짜로 판별하도록 학습되는 판별기

를 포함하는 CGAN(Conditional Generative Adversarial Networks) 모델인 네트워크 침입 탐지 방법.

청구항 11

제1항 내지 제10항 중 어느 한 항의 방법을 실행하기 위한 프로그램이 기록된 컴퓨터에서 판독 가능한 기록 매체.

청구항 12

원본 학습 데이터 세트로부터 희소 클래스의 특징을 추출하는 특징 추출부;

상기 희소 클래스의 특징을 데이터 생성 모델에 입력하여 상기 희소 클래스의 특징을 가지는 추가 데이터들을 생성하는 추가 데이터 생성부; 및

상기 원본 학습 데이터 세트 및 상기 추가 데이터들을 이용하여 침입 탐지 여부를 분류하기 위한 분류기를 학습하는 분류기 학습 장치

를 포함하는 분류기 학습 시스템.

청구항 13

원본 학습 데이터 세트에서 추출한 희소 클래스의 특징을 데이터 생성 모델에 입력하여 추가 데이터들을 생성하는 학습 데이터 생성 장치;

상기 추가 데이터들 및 상기 원본 학습 데이터 세트를 이용하여 침입 탐지 모델의 분류기를 학습하는 분류기 학습 장치; 및

네트워크 데이터를 상기 분류기에 입력하고, 상기 분류기의 출력에 따라 네트워크의 침입 여부를 탐지하는 침입 탐지 장치

를 포함하는 네트워크 침입 탐지 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 네트워크 침입 탐지 시스템 및 방법에 관한 것으로, 보다 구체적으로는 오토 인코더(AutoEncoder) 및 CGAN(Conditional Generative Adversarial Networks) 모델을 사용한 네트워크 침입 탐지 시스템 및 방법에 관한 것이다.

배 경 기 술

[0002] 네트워크 침입 탐지 시스템(NIDS)은 보안 방어 기술이며, 탐지 방식에 따라 오용 탐지 방식과 비정상 행위 탐지 방식 중 하나로 분류될 수 있다. 그러나, 오용 탐지 방식은 전문가에 의하여 만들어진 규칙을 기준으로 침입을 탐지하는 방식이므로, 하루에도 복수의 신형 해킹 방법, 또는 변종이 등장하고 있는 현재로는 침입 탐지 성능에 한계가 있는 실정이다.

[0003] 종래의 비정상 행위 탐지 기법을 사용한 네트워크 침입 탐지 시스템은 네트워크의 정상적이고 평균적인 상태를 기준으로 설정하고, 네트워크의 상태에 상대적으로 급격한 변화가 발생하거나, 확률이 낮은 사건이 발생하면 침입으로 규정하고 있다.

[0004] 그러나, 종래의 비정상 행위 탐지 기법을 사용한 네트워크 침입 탐지 시스템은 높은 차원 및 비선형 특성을 갖는 데이터를 이용하여 침입을 탐지하고 있으므로, 시간 복잡성과 공간 복잡성의 한계를 가지고 있다.

[0005] 또한, 종래의 비정상 행위 탐지 기법을 사용한 네트워크 침입 탐지 시스템은 데이터 불균형으로 인하여 희소 클래스가 잘못 탐지되는 경우가 종종 발생하고 있다.

[0006] 따라서, 데이터 불균형을 해소하고, 시간 복잡성과 공간 복잡성이 낮은 방식으로 네트워크 침입을 탐지하는 방법이 요청되고 있다.

발명의 내용

해결하려는 과제

[0007] 본 발명은 오토 인코더로 원본 학습 데이터 세트에 포함된 희귀 클래스의 저차원 데이터를 추출하고, CGAN 모델로 희귀 클래스의 저차원 데이터를 포함하는 추가 데이터를 생성하는 장치 및 방법을 제공한다.

[0008] 또한, 본 발명은 시간 복잡성과 공간 복잡성이 낮으며, 데이터 불균형이 해소된 데이터를 이용하여 학습된 침입 탐지 모델의 분류기로 네트워크 침입을 탐지함으로써, 네트워크 침입 탐지 성능을 향상시키는 장치 및 방법을 제공한다.

과제의 해결 수단

[0009] 본 발명의 일실시예에 따른 네트워크 침입 탐지 시스템의 학습 데이터 생성 방법은 원본 학습 데이터 세트로부터 희소 클래스의 특징을 추출하는 단계; 및 상기 희소 클래스의 특징을 데이터 생성 모델에 입력하여 상기 희소 클래스의 특징을 가지는 추가 데이터들을 생성하는 단계를 포함할 수 있다.

[0010] 본 발명의 일실시예에 따른 학습 데이터 생성 방법의 상기 희소 클래스의 특징을 추출하는 단계는, 상기 원본 학습 데이터 세트에 포함된 희소 클래스의 특징을 차원 축소하여 데이터 차원이 감소된 희소 클래스의 특징을 추출할 수 있다.

[0011] 본 발명의 일실시예에 따른 학습 데이터 생성 방법의 희소 클래스의 특징을 추출하는 단계는, 오토 인코더를 이용하여 상기 원본 학습 데이터 세트에 포함된 희소 클래스의 특징을 차원 축소하고, 상기 오토 인코더는, 희소 클래스의 특징이 입력된 경우, 상기 오토 인코더의 마지막 코딩층을 희소 클래스의 핵심 특징으로 설정하고, 상기 오토 인코더의 디코더의 출력이 상기 핵심 특징에 대응되도록 학습될 수 있다.

[0012] 본 발명의 일실시예에 따른 학습 데이터 생성 방법의 데이터 생성 모델은, 노이즈와 클래스의 라벨을 이용하여 상기 희소 클래스의 특징으로 판별되도록 하기 위한 가짜 데이터를 생성하도록 학습되는 생성기 및 상기 가짜 데이터와 상기 희소 클래스의 특징, 및 상기 클래스의 라벨을 이용하여 상기 가짜 데이터를 가짜로 판별하도록 학습되는 판별기를 포함하는 CGAN(Conditional Generative Adversarial Networks) 모델일 수 있다.

[0013] 본 발명의 일실시예에 따른 네트워크 침입 탐지 시스템의 침입 탐지 모델 학습 방법은 원본 학습 데이터 세트로부터 희소 클래스의 특징을 추출하는 단계; 상기 희소 클래스의 특징을 데이터 생성 모델에 입력하여 상기 희소 클래스의 특징을 가지는 추가 데이터들을 생성하는 단계; 및 상기 원본 학습 데이터 세트 및 상기 추가 데이터들을 이용하여 침입 탐지 여부를 분류하기 위한 분류기를 학습하는 단계를 포함할 수 있다.

[0014] 본 발명의 일실시예에 따른 침입 탐지 모델 학습 방법은 상기 희소 클래스의 특징을 추출하는 단계는, 오토 인코더를 이용하여 상기 원본 학습 데이터 세트에 포함된 희소 클래스의 특징을 차원 축소하고, 상기 오토 인코더는, 희소 클래스의 특징이 입력된 경우, 상기 오토 인코더의 마지막 코딩층을 희소 클래스의 핵심 특징으로 설정하고, 상기 오토 인코더의 디코더의 출력이 상기 핵심 특징에 대응되도록 학습될 수 있다.

[0015] 본 발명의 일실시예에 따른 침입 탐지 모델 학습 방법의 데이터 생성 모델은, 노이즈와 클래스의 라벨을 이용하여 상기 희소 클래스의 특징으로 판별되도록 하기 위한 가짜 데이터를 생성하도록 학습되는 생성기 및 상기 가짜 데이터와 상기 희소 클래스의 특징, 및 상기 클래스의 라벨을 이용하여 상기 가짜 데이터를 가짜로 판별하도록 학습되는 판별기를 포함하는 CGAN 모델일 수 있다.

[0016] 본 발명의 일실시예에 따른 네트워크 침입 탐지 방법은 네트워크 데이터를 침입 탐지 모델의 분류기에 입력하는 단계; 및 상기 분류기의 출력에 따라 네트워크의 침입 여부를 탐지하는 단계를 포함하고, 상기 분류기는, 원본 학습 데이터 세트에서 추출한 희소 클래스의 특징을 데이터 생성 모델에 입력하여 생성한 추가 데이터들 및 상기 원본 학습 데이터 세트를 이용하여 학습될 수 있다.

[0017] 본 발명의 일실시예에 따른 네트워크 침입 탐지 방법의 희소 클래스의 특징은, 오토 인코더를 이용하여 상기 원본 학습 데이터 세트에 포함된 특징을 차원 축소하여 추출되고, 상기 오토 인코더는, 희소 클래스의 특징이 입력된 경우, 상기 오토 인코더의 마지막 코딩층을 희소 클래스의 핵심 특징으로 설정하고, 상기 오토 인코더의 디코더의 출력이 상기 핵심 특징에 대응되도록 학습될 수 있다.

[0018] 본 발명의 일실시예에 따른 네트워크 침입 탐지 방법의 데이터 생성 모델은, 노이즈와 클래스의 라벨을 이용하여 상기 희소 클래스의 특징으로 판별되도록 하기 위한 가짜 데이터를 생성하도록 학습되는 생성기 및 상기 가짜 데이터와 상기 희소 클래스의 특징, 및 상기 클래스의 라벨을 이용하여 상기 가짜 데이터를 가짜로 판별하도록 학습되는 판별기를 포함하는 CGAN 모델일 수 있다.

[0019] 본 발명의 일실시예에 따른 분류기 학습 장치는 원본 학습 데이터 세트로부터 희소 클래스의 특징을 추출하는

특징 추출부; 상기 희소 클래스의 특징을 데이터 생성 모델에 입력하여 상기 희소 클래스의 특징을 가지는 추가 데이터들을 생성하는 추가 데이터 생성부; 및 상기 원본 학습 데이터 세트 및 상기 추가 데이터들을 이용하여 침입 탐지 여부를 분류하기 위한 분류기를 학습하는 분류기 학습 장치를 포함할 수 있다.

[0020] 본 발명의 일실시예에 따른 네트워크 침입 탐지 시스템은 원본 학습 데이터 세트에서 추출한 희소 클래스의 특징을 데이터 생성 모델에 입력하여 추가 데이터들을 생성하는 학습 데이터 생성 장치; 상기 추가 데이터들 및 상기 원본 학습 데이터 세트를 이용하여 침입 탐지 모델의 분류기를 학습하는 분류기 학습 장치; 및 네트워크 데이터를 상기 분류기에 입력하고, 상기 분류기의 출력에 따라 네트워크의 침입 여부를 탐지하는 침입 탐지 장치를 포함할 수 있다.

발명의 효과

[0021] 본 발명의 일실시예에 의하면, 오토 인코더로 원본 학습 데이터 세트에 포함된 희귀 클래스의 저차원 데이터를 추출하고, CGAN 모델로 희귀 클래스의 저차원 데이터를 포함하는 추가 데이터를 생성함으로써, 시간 복잡성과 공간 복잡성이 낮으며, 데이터 불균형이 해소된 데이터를 침입 탐지 모델의 분류기를 학습시키기 위하여 제공할 수 있다.

[0022] 또한, 본 발명의 일실시예에 의하면, 시간 복잡성과 공간 복잡성이 낮으며, 데이터 불균형이 해소된 데이터를 이용하여 학습된 침입 탐지 모델의 분류기로 네트워크 침입을 탐지함으로써, 네트워크 침입 탐지 성능을 향상시킬 수 있다.

도면의 간단한 설명

[0023] 도 1은 본 발명의 일실시예에 따른 네트워크 침입 탐지 시스템을 도시한 도면이다.

도 2는 본 발명의 일실시예에 따른 학습 데이터 세트 전처리 과정의 일례이다.

도 3은 본 발명의 일실시예에 따른 특징 추출부의 동작 과정의 일례이다.

도 4는 본 발명의 일실시예에 따른 특징 추출부에 포함된 오토 인코더의 일례이다.

도 5는 본 발명의 일실시예에 따른 추가 데이터 생성부에 포함된 CGAN 모델이다.

도 6은 본 발명의 일실시예에 따른 분류기 학습 장치의 동작 일례이다.

도 7은 본 발명의 일실시예에 따른 오토 인코더의 성능 비교 결과의 일례이다.

도 8은 본 발명의 일실시예에 따른 CGAN 모델의 성능 비교 결과의 일례이다.

도 9는 본 발명의 일실시예에 따른 네트워크 침입 탐지 방법을 도시한 플로우차트이다.

발명을 실시하기 위한 구체적인 내용

[0024] 이하에서, 첨부된 도면을 참조하여 실시예들을 상세하게 설명한다. 그러나, 실시예들에는 다양한 변경이 가해질 수 있어서 특허출원의 권리 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 실시예들에 대한 모든 변경, 균등물 내지 대체물이 권리 범위에 포함되는 것으로 이해되어야 한다.

[0025] 실시예에서 사용한 용어는 단지 설명을 목적으로 사용된 것으로, 한정하려는 의도로 해석되어서는 안된다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서 상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0026] 또한, 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 실시예를 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 실시예의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.

[0027] 이하, 본 발명의 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0028] 도 1은 본 발명의 일실시예에 따른 네트워크 침입 탐지 시스템을 도시한 도면이다.

[0029] 네트워크 침입 탐지 시스템(100)는 도 1에 도시된 바와 같이 학습 데이터 생성 장치(100), 분류기 학습 장치

(101), 및 침입 탐지 장치(102)를 포함할 수 있다. 이때, 학습 데이터 생성 장치(100), 분류기 학습 장치(101)는 별도의 분류기 학습 시스템에 포함되고 네트워크 침입 탐지 시스템(100)는 분류기 학습 장치(101)로부터 유선, 또는 무선, 또는 오프라인으로 분류기를 전달받는 침입 탐지 장치(102)로 구성될 수도 있다.

[0030] 학습 데이터 생성 장치(100)는 오토 인코더(AutoEncoder) 및 CGAN(Conditional Generative Adversarial Networks) 모델로 원본 학습 데이터 세트에 포함된 희소 클래스를 오버 샘플링하여 데이터 불균형이 해소된 학습 데이터를 생성할 수 있다.

[0031] 데이터 불균형은 입력 데이터, 또는 원본 학습 데이터 세트에 포함된 클래스들의 개수에 현저한 차이가 발생하는 것을 의미한다.

[0032] 네트워크 데이터에서 침입이 발생하는 경우보다는 침입이 발생하지 않는 경우가 더 많으므로, 데이터에서 침입과 관련된 클래스는 데이터에서 정상 상태와 관련된 클래스보다 개수가 적을 수 있다. 예를 들어, 원본 학습 데이터 세트로 사용될 수 있는 CICIDS 2017 데이터셋에서 정상 상태에 대응하는 클래스는 전체 데이터의 80.3%인 반면, HeartBleed 공격에 대응하는 클래스는 전체 데이터의 0.0004%만을 차지하는 희소 클래스일 수 있다.

[0033] 데이터 불균형이 발생한 비대칭 데이터셋을 사용하여 침입 탐지용 분류기를 학습하는 경우, 분류기의 정확도(Accuracy)는 높지만, 재현율(Recall)이 낮을 수 있다. 예를 들어, 100개의 데이터 안에 침입에 대응하는 공격 데이터가 3개 포함된 데이터 셋에 대하여 침입 탐지용 분류기가 100개 데이터를 모두 정상이라고 예측한 경우, 해당 분류기는 침입을 전혀 감지 못하였음에도 침입 여부에 대한 적중도가 97%로 나타날 수 있다.

[0034] 또한, 앞서 설명한 바와 같이 침입과 관련된 클래스는 정상 상태와 관련된 클래스보다 비율이 적은 희소 클래스일 수 있다. 따라서, 학습 데이터 생성 장치(100)는 원본 학습 데이터 세트에 포함된 희소 클래스를 복제하여 추가하는 오버 샘플링을 수행하여 침입과 관련된 클래스와 정상 상태와 관련된 클래스 간의 비율이 균형을 이루도록 함으로써, 데이터 불균형에 따른 분류기 성능의 감소를 방지할 수 있다.

[0035] 구체적으로, 학습 데이터 생성 장치(100)는 도 1에 도시된 바와 같이 입력부(110), 특징 추출부(120), 및 추가 데이터 생성부(130)를 포함할 수 있다. 이때, 입력부(110)는 입출력 인터페이스이고, 특징 추출부(120), 및 추가 데이터 생성부(130)는 서로 다른 프로세서, 또는 하나의 프로세서에서 수행되는 프로그램에 포함된 각각의 모듈일 수 있다.

[0036] 입력부(110)는 원본 학습 데이터 세트를 입력받을 수 있다. 이때, 원본 학습 데이터는 네트워크 침입 탐지 시스템(100)에서 일정 시간 수집된 데이터일 수도 있고, 침입 탐지를 위한 학습용 데이터들의 집합일 수도 있다. 예를 들어, 입력부는 실제 데이터(PCAP)와 유사한 정상 상태의 데이터 및 최신의 네트워크 공격 방식과 관련된 데이터를 포함하는 CICIDS 2017 데이터 세트를 입력받을 수 있다. 이때, CICIDS 2017 데이터 세트에는 타임 스탬프, 소스 및 대상 IP, 소스 및 대상 Port, 프로토콜 및 공격을 기반으로 흐름이 있는 CICFlowMeter를 사용한 네트워크 트래픽 분석 결과도 포함될 수 있다. 원본 학습 데이터 세트의 구조 및 생성 과정은 이하 도 2를 참조하여 상세히 설명한다.

[0037] 특징 추출부(120)는 입력부(110)가 입력받은 원본 학습 데이터 세트로부터 희소 클래스의 특징을 추출할 수 있다.

[0038] 네트워크 침입과 관련된 학습 데이터는 데이터 차원이 크므로 원본 학습 데이터 세트를 그래도 사용하는 경우, 학습 속도 및 학습 성능이 낮아질 수 있다. 또한, 데이터의 데이터 차원이 증가할수록 데이터 공간의 부피가 증가하여 데이터의 밀도가 감소할 수 있다. 그에 따라 데이터 포인트 간의 거리도 증가하여 학습 모델이 복잡해지고, 오버피팅(overfitting)이 발생할 가능성도 증가하게 된다. 따라서, 특징 추출부(120)는 원본 학습 데이터 세트에 포함된 희소 클래스의 특징을 차원 축소하여 데이터 차원이 감소된 희소 클래스의 저차원 특징을 추출함으로써, 희소 클래스의 저차원 특징을 사용하는 분류기의 학습 속도 및 학습 성능을 향상시킬 수 있다.

[0039] 구체적으로, 특징 추출부(120)는 오토 인코더를 이용하여 원본 학습 데이터 세트에 포함된 희소 클래스의 특징을 차원 축소할 수 있다. 이때, 희소 클래스의 특징이 입력된 경우, 오토 인코더는 오토 인코더의 마지막 코딩층을 희소 클래스의 핵심 특징으로 설정하고, 오토 인코더의 디코더의 출력이 핵심 특징에 대응되도록 학습될 수 있다. 특징 추출부(120)의 동작은 이하 도 3 내지 도 4를 참조하여 상세히 설명한다.

[0040] 추가 데이터 생성부(130)는 특징 추출부(120)에서 추출된 희소 클래스의 저차원 특징을 데이터 생성 모델에 입력하여 희소 클래스의 특징을 가지는 추가 데이터들을 생성함으로써, 오버 샘플링을 수행할 수 있다. 이때, 데이터 생성 모델은 노이즈와 클래스의 라벨을 이용하여 희소 클래스의 특징으로 판별되도록 하기 위한 가짜 데이

터를 생성하도록 학습되는 생성기 및 가짜 데이터와 희소 클래스의 특징, 및 클래스의 라벨을 이용하여 가짜 데이터를 가짜로 판별하도록 학습되는 판별기를 포함하는 CGAN(Conditional Generative Adversarial Networks) 모델일 수 있다. 추가 데이터 생성부(130)의 동작은 이하 도 5를 참조하여 상세히 설명한다.

[0041] 분류기 학습 장치(101)는 입력부(110)가 입력받은 원본 학습 데이터 세트 및 추가 데이터 생성부(130)가 생성한 추가 데이터들을 이용하여 침입 탐지 여부를 분류하기 위한 분류기를 학습시킬 수 있다. 그리고, 분류기 학습 장치(101)는 학습이 종료된 분류기를 침입 탐지 장치(102)로 전송할 수 있다.

[0042] 침입 탐지 장치(102)는 네트워크 데이터를 입력받을 수 있다. 그리고, 침입 탐지 장치(102)는 입력받은 네트워크 데이터를 분류기 학습 장치(101)로부터 수신한 침입 탐지 모델의 분류기에 입력할 수 있다. 다음으로, 침입 탐지 장치(102)는 침입 탐지 모델의 분류기의 출력에 따라 네트워크의 침입 여부를 탐지할 수 있다.

[0043] 본 발명의 일실시예에 따른 네트워크 침입 탐지 시스템은 오토 인코더로 원본 학습 데이터 세트에 포함된 희귀 클래스의 저차원 데이터를 추출하고, CGAN 모델로 희귀 클래스의 저차원 데이터를 포함하는 추가 데이터를 생성함으로써, 시간 복잡성과 공간 복잡성이 낮으며, 데이터 불균형이 해소된 데이터를 침입 탐지 모델의 분류기를 학습시키기 위하여 제공할 수 있다.

[0044] 또한, 본 발명의 일실시예에 따른 네트워크 침입 탐지 시스템은 시간 복잡성과 공간 복잡성이 낮으며, 데이터 불균형이 해소된 데이터를 이용하여 학습된 침입 탐지 모델의 분류기로 네트워크 침입을 탐지함으로써, 네트워크 침입 탐지 성능을 향상시킬 수 있다.

[0045] 도 2는 본 발명의 일실시예에 따른 학습 데이터 세트 전처리 과정의 일례이다.

[0046] 입력부(110)가 입력받은 CICIDS 2017 데이터 세트(210)는 하기 10가지 기준을 만족할 수 있다.

[0047] (1) 완벽한 네트워크 구성: 안전한 네트워크 토폴로지에는 모뎀, 방화벽, 스위치, 라우터 및 Windows, Ubuntu 및 Mac OS X와 같은 다양한 운영체제가 포함될 수 있다..

[0048] (2) 완전한 트래픽: 사용자 프로파일링 에이전트와 12개의 서로 다른 시스템을 피해자 네트워크에 배치하고 네트워크 상에서 실제로 공격을 하여 생성될 수 있다.

[0049] (3) 레이블이 있는 데이터 세트 : 표 1과 같은 양성 및 공격 레이블을 가질 수 있다.

표 1

[0050]

Days	Labels
Monday	Benign
Tuesday	BForce, SFTP, SSH
Wednesday	DoS and Hearbleed Attacks, slowloris, Slowhttptest, Hulk and GoldenEye
Thursday	Web and Infiltration Attacks, Web BForce, XSS and Sql Inject, Infiltration Dropbox Download and Cool disk
Friday	DDoS LOIT, Botnet ARES, PortScans (sS,sT,sF,sX,sN,sP,sV,sU, sO,sA,sW,sR,sL and B)

[0051] (4) 완벽한 상호 작용 : 두 개의 서로 다른 네트워크와 인터넷 통신을 수행하며, 그 과정에서 내부 LAN을 통하여 수집한 데이터로 생성될 수 있다.

[0052] (5) 전체 캡처 : Tapping 시스템과 Mirror 포트를 사용하여 스토리지 서버에서 모든 트래픽이 캡처 및 기록되도록 할 수 있다.

[0053] (6) 사용 가능한 프로토콜 : HTTP, HTTPS, FTP, SSH 및 이메일 프로토콜과 같은 사용 가능한 모든 프로토콜이 포함될 수 있다.

[0054] (7) 공격 다양성 : 2016 McAfee 보고서에 기반한 가장 일반적인 공격이 포함될 수 있다.

[0055] (8) 이기종 : 공격 실행 중 기본 스위치 및 네트워크 덤프의 네트워크 트래픽과 모든 대상 컴퓨터의 시스템 호

출이 캡처 될 수 있다.

[0056] (9) Feature 세트 : CICFlowMeter를 사용하여 생성된 네트워크 트래픽에서 80개 이상의 네트워크 흐름 기능을 추출하고 네트워크 흐름 데이터 세트를 CSV 파일로 구성될 수 있다.

[0057] (10) 메타 데이터 (MetaData) : 시간, 공격, 흐름 및 레이블을 포함하는 데이터 세트일 수 있다.

[0058] 예를 들어, CICIDS 2017 데이터 세트(210)는 표 2에 도시된 바와 같이 정상 트래픽과 14가지 유형의 공격으로 구성될 수 있다. 다만, Brute Force, SQL Injection, XSS로 분류될 수 있는 Web Attack은 각각의 비중이 다른 공격에 비하여 적으므로, Brute Force, SQL Injection, XSS를 동일한 하나의 클래스로 부여할 수 있다. 즉, CICIDS 2017 데이터 세트(210)는 정상 트래픽과 12개의 공격 데이터로 구성될 수 있다.

표 2

[0059]

Classes		Number	Percentage	Label
Benign		2,273,097	80.3004%	0
DDoS		128,027	4.5227%	1
Port Scan		158,930	5.6441%	2
Bot		1,966	0.0695%	3
Infilteration		36	0.0013%	4
Web attack	Brute Force	2,180	0.0770%	5
	SQL Injection			
	XSS			
FTP-Patator		7,938	0.2804%	6
SSH-Patator		5,897	0.2083%	7
DoS GoldenEye		10,293	0.3636%	8
DoS Hulk		231,073	8.1630%	9
DoS Slowhttptest		5,499	0.1943%	10
DoS Slowloris		5,796	0.2048%	11
Heartbleed		11	0.0004%	12

[0060] 이때, CICIDS 2017 데이터 세트(210)에 포함된 특징들은 표 3에 도시된 바와 같이 77개일 수 있다.

표 3

[0061]

Number	Feature	Number	Feature
1	Destination Port	40	Max Packet Length
2	Flow Duration	41	Packet Length Mean
3	Total Fwd Packets	42	Packet Length Std
4	Total Backward Packets	43	Packet Length Variance
5	Total Length of Fwd Packets	44	FIN Flag Count
6	Total Length of Bwd Packets	45	SYN Flag Coun
7	Fwd Packet Length Max	46	RST Flag Count

8	Fwd Packet Length Min	47	PSH Flag Count
9	Fwd Packet Length Mean	48	ACK Flag Count
10	Fwd Packet Length Std	49	URG Flag Count
11	Bwd Packet Length Max	50	CWE Flag Count
12	Bwd Packet Length Min	51	ECE Flag Count
13	Bwd Packet Length Mea	52	Down/Up Ratio
14	Bwd Packet Length Std	53	Average Packet Size
15	Flow Bytes/s	54	Avg Fwd Segment Size
16	Flow Packets/	55	Avg Bwd Segment Size
17	Flow IAT Mean	56	Fwd Avg Bytes/Bulk
18	Flow IAT Std	57	Fwd Avg Packets/Bulk
19	Flow IAT Max	58	Fwd Avg Bulk Rat
20	Flow IAT Min	59	Bwd Avg Bytes/Bulk
21	Fwd IAT Total	60	Bwd Avg Packets/Bulk
22	Fwd IAT Mean	61	Bwd Avg Bulk Rate
23	Fwd IAT Std	62	Subflow Fwd Packets
24	Fwd IAT Max	63	Subflow Fwd Bytes
25	Fwd IAT Min	64	Subflow Bwd Packets
26	Bwd IAT Tota	65	Subflow Bwd Bytes
27	Bwd IAT Mean	66	Init_Win_bytes_forward
28	Bwd IAT Std	67	Init_Win_bytes_backward
29	Bwd IAT Max	68	act_data_pkt_fwd
30	Bwd IAT Min	69	min_seg_size_forward
31	Fwd PSH Flags	70	Active Mean
32	Bwd PSH Flags	71	Active Std
33	Fwd URG Flags	72	Active Max
34	Bwd URG Flags	73	Active Min
35	Fwd Header Length	74	Idle Mea
36	Bwd Header Length	75	Idle Std
37	Fwd Packets/s	76	Idle Mean
38	Bwd Packets/s	77	Idle Min
39	Min Packet Length		

[0062]

[0063] 또한, CICIDS 2017 데이터 세트(210)에 포함된 데이터들은 특징의 특성에 따라 범위가 다르므로, 특징 추출부(120)는 CICIDS 2017 데이터 세트(210)에 정규화를 수행하여 특징 추출에 사용될 학습 데이터 세트(220)를 생성할 수도 있다.

[0064] 예를 들어, 특징 추출부(120)는 수학식 1을 이용한 MinMax Scaler 방식으로 CICIDS 2017 데이터 세트(210)를 정규화할 수 있다.

수학식 1

[0065]
$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

[0066] 이때, X_{max}은 CICIDS 2017 데이터 세트(210)에 포함된 원본 특징 값의 최대값이고, X_{min}은 CICIDS 2017 데이터 세트(210)에 포함된 원본 특징 값의 최소값이며, X_{norm}은 정규화된 특징 값일 수 있다.

[0067] 도 3은 본 발명의 일실시예에 따른 특징 추출부의 동작 과정의 일례이다.

[0068] 단계(310)에서 특징 추출부(120)는 오토 인코더의 인코더로 희소 클래스의 핵심 특징 데이터만 추출하여 학습할 수 있다.

[0069] 단계(320)에서 특징 추출부(120)는 오토 인코더의 디코더의 출력값과 오토 인코더의 인코더에 입력된 값 간의 유사도가 최대가 되도록 오토 인코더의 데이터를 재구축할 수 있다. 구체적으로, 특징 추출부(120)는 오토 인코더의 디코더의 출력값과 오토 인코더의 인코더에 입력된 값 간의 유사도가 최대가 되도록 오토 인코더에 포함된 은닉층의 개수 및 뉴런 수를 조정할 수 있다.

[0070] 단계(330)에서 특징 추출부(120)는 단계(320)에서 재구축된 오토 인코더에 학습 데이터 세트(220)를 입력하여

학습 데이터 세트(220)에 포함된 희소 클래스의 저차원 특징들로 구성된 저차원 데이터 세트(330)를 출력할 수 있다.

[0071] 도 4는 본 발명의 일실시예에 따른 특징 추출부에 포함된 오토 인코더의 일례이다. 특징 추출부(120)는 도 4에 도시된 바와 같은 SSAE(Stacked Sparse AutoEncoder)로 원본 학습 데이터 세트에 포함된 희소 클래스의 특징을 차원 축소하여 데이터 차원이 감소된 희소 클래스의 특징을 추출할 수 있다.

[0072] 오토 인코더는 입력된 데이터의 데이터 차원을 입력층(input layer) 및 은닉층(hidden layer)로 구성된 인코더에서 축소 시킬 수 있다. 이때, 데이터 차원이 축소된 입력 데이터를 코드, 또는 잠재 변수(latent variables)라고 정의할 수 있다.

[0073] 그리고, 오토 인코더는 은닉층(hidden layer) 및 출력층(output layer)으로 구성된 디코더를 이용하여 잠재 변수를 입력 데이터에 대응하는 출력 데이터로 변환할 수 있다.

[0074] 구체적으로, 특징 추출부(120)는 77개의 특징으로 구성된 학습 데이터 세트를 오토 인코더에 입력할 수 있다 이때, 오토 인코더의 인코더는 스파스 파라미터(Sparse parameter)를 이용하여 77개의 특징에서 핵심 특징 데이터만 추출하여 은닉층에서 학습하고, 추출하지 않은 나머지 데이터는 손실 시킬 수 있다. 이때, 특징 추출부(120)는 오토 인코더의 마지막 코딩층을 희소 클래스의 핵심 특징의 저차원 데이터(410)로 추출할 수 있다.

[0075] 그리고, 특징 추출부(120)는 오토 인코더의 디코더의 출력값과 오토 인코더의 인코더에 입력된 값 간의 유사도가 최대가 되도록 하는 오토 인코더의 디코더에 포함된 은닉층 가중치를 검색함으로써, 오토 인코더가 학습 데이터 세트에 포함된 희소 클래스의 특징들 중에서 불필요한 특징을 제외하고, 핵심 특징만을 추출하도록 할 수 있다.

[0076] 특징 추출부(120)는 학습 데이터 세트에 포함된 희소 클래스의 특징들 중에서 핵심 특징의 저차원 데이터만을 추출함으로써, 불필요한 특징의 학습에 소요되는 시간을 절감하여 네트워크 침입 탐지 시간 및 네트워크 침입 탐지 성능을 향상시킬 수 있다.

[0077] 도 5는 본 발명의 일실시예에 따른 추가 데이터 생성부에 포함된 CGAN 모델이다.

[0078] CGAN 모델은 도 5에 도시된 바와 같이 생성기(520), 및 판별기(530)를 포함할 수 있다.

[0079] 특징 추출부(120)는 희소 클래스(510)로부터 저차원 특징(511)을 추출할 수 있다 또한, 추가 데이터 생성부(130)는 입력 데이터(input(z))(521) 및 클래스 라벨(512)을 생성기(Generator)(520)에 입력할 수 있다. 이때, 입력 데이터(input(z))(521)는 노이즈(noise)이고, 클래스 라벨(512)은 희소 클래스(510)에 대응하는 라벨, 및 희소 클래스(510)에 해당하지 않는 다른 클래스의 라벨을 포함할 수 있다.

[0080] 그리고, 생성기(520)는 입력 데이터(input(z))(521) 및 클래스 라벨(512)을 이용하여 판별기(530)에서 희소 클래스의 특징으로 판별되도록 하기 위한 가짜 데이터를 생성하도록 학습될 수 있다.

[0081] 판별기(530)는 생성기(520)가 생성한 가짜 데이터(G(z))(522)와 입력 데이터(input(x))(531) 및 클래스 라벨(512)을 이용하여 가짜 데이터를 가짜로 판별(532)하도록 학습될 수 있다. 이때, 입력 데이터(input(x))(531)는 특징 추출부(120)가 희소 클래스(510)로부터 추출한 저차원 특징(511)일 수 있다.

[0082] 그리고, 생성기(520)는 판별기(530)의 판별 결과를 피드백 받아 이전보다 더 판별기(530)에서 희소 클래스의 특징으로 판별될 가능성이 높은 가짜 데이터를 생성할 수 있다.

[0083] CGAN 모델은 클래스 라벨(512)를 입력받아 학습함으로써, 데이터의 분포에 따른 오버 샘플링의 효과가 저하되는 오리지날(Original) GAN의 단점을 해결하고 최적의 오버 샘플링을 할 수 있다.

[0084] 예를 들어 희소 클래스 중 Bot을 오버 샘플링하는 경우, 추가 데이터 생성부(130)는 임의의 노이즈(521)와 클래스 라벨(512)을 입력하여 생성기(520)가 Bot의 가짜 데이터(522)를 생성하도록 할 수 있다. 이때, 처음으로 생성된 Bot의 가짜 데이터(522)는 Bot의 실제 데이터와는 전혀 다른 모양의 데이터일 수 있다. 그리고, 판별기(530)는 가짜 데이터(522)와 실제 Bot 데이터를 비교하여 가짜 데이터(522)를 가짜로 판별하고, 판별 결과를 생성기(520)에 피드백할 수 있다. 다음으로, 생성기(520)는 피드백을 참조하여 이전에 비하여 실제 Bot 데이터와 더 유사한 가짜 데이터(522)를 생성하는 과정을 반복할 수 있다.

[0085] 도 6은 본 발명의 일실시예에 따른 분류기 학습 장치의 동작 일례이다.

[0086] 분류기 학습 장치(101)는 랜덤 포레스트(RF: Random Forest) (610)와 같은 머신 러닝 알고리즘을 이용하여 분류

기(611)를 학습할 수 있다. 이때, 랜덤 포레스트는 복수의 의사 결정 트리(DT: Decision Tree)들을 생성한 다음, 각각의 트리에서의 예측한 값 중에서 가장 많은 선택을 받은 클래스를 예측할 수 있다.

[0087] 도 7은 본 발명의 일실시예에 따른 오토 인코더의 성능 비교 결과의 일례이다.

[0088] 특징 추출부(120)는 오토 인코더의 인코더에 포함된 은닉층의 개수를 조정하여 희소 클래스에 대응하는 최적의 저차원 특징을 추출할 수 있다. 도 7은 학습 데이터 세트에 77개의 특징이 포함되고, 오토 인코더의 인코더에 포함된 은닉층에 포함된 뉴런의 개수를 각각 (5), (15), (30), (40), (50), (60), (70)으로 압축한 경우의 성능 일례이다. 이때, 성능(performance)은 인코딩으로 압축된 특징을 희소 클래스의 저차원 특징으로 추출하여 랜덤 포레스트 알고리즘으로 분류한 다중 분류(multi-classification)의 평균 성능이며, 정밀도(precision), 재현율(recall), 및 f1 스코어(f1-score)로 표시될 수 있다. 이때, f1 스코어는 정밀도와 재현율의 조화 평균에 따라 결정되며, 데이터가 불균형 경우, 분류기 모델의 성능을 평가하기 위하여 사용되는 지표이다.

[0089] 또한, 은닉층에 포함된 뉴런의 개수가 작으면 학습시간은 감소하나 압축률도 낮으므로, 성능도 감소할 수 있다. 도 7에 도시된 바에 따르면 은닉층에 포함된 뉴런의 개수가 [40]일 때 가장 좋은 성능을 가질 수 있으며, 뉴런의 개수가 [40]을 초과하는 경우, 성능이 감소할 수 있다. 다만, 도 7은 표 2 및 표 3에 따른 학습 데이터에 따른 성능을 나타내고 있으며, 실시예에 따라서는 최적의 성능을 가질 수 있는 은닉층에 포함된 뉴런의 개수가 도 7과 상이할 수도 있다.

[0090] 도 8은 도 7에서 최적의 성능을 가지는 것으로 판단된 저차원 특징을 CGAN 모델에 입력한 결과의 일례이다.

[0091] 이때, 가로축은 원본 학습 데이터 세트에 포함된 class의 실제 개수이고, 세로축은 분류기에서 원본 학습 데이터 세트에 포함된 것으로 예측한 class의 개수이다. 예를 들어, BENIGN class는 실제 데이터 개수는 908,528개이고, 분류기에서 예측한 개수는 907,719이며 2개를 DDoS로 잘못 예측했다는 것을 나타낼 수 있다.

[0092] 또한, 희소 클래스인 Infitration은 실제 데이터수가 10개이고 예측한 수도 10개이기 때문에 Accuracy와 Precision은 100%이다. 그러나 다른 Benign class를 Infiltration라고 잘못 예측한 수가 5이기 때문에 재현율(recall), 및 f1 스코어(f1-score)는 66.67%, 80.00%일 수 있다. 따라서, 희소 클래스의 성능을 측정하기 위해서는 희소 클래스들에 대한 재현율(recall), 및 f1 스코어(f1-score)이 필요할 수 있다.

[0093] 도 9는 본 발명의 일실시예에 따른 네트워크 침입 탐지 방법을 도시한 플로우차트이다.

[0094] 단계(910)에서 입력부(110)는 원본 학습 데이터 세트를 획득할 수 있다.

[0095] 단계(920)에서 특징 추출부(120)는 단계(910)에서 획득한 원본 학습 데이터 세트로부터 희소 클래스의 특징을 추출할 수 있다. 이때, 특징 추출부(120)는 원본 학습 데이터 세트에 포함된 희소 클래스의 특징을 차원 축소하여 데이터 차원이 감소된 희소 클래스의 저차원 특징을 추출할 수 있다.

[0096] 단계(930)에서 추가 데이터 생성부(130)는 단계(920)에서 추출된 희소 클래스의 저차원 특징을 데이터 생성 모델에 입력하여 희소 클래스의 특징을 가지는 추가 데이터들을 생성할 수 있다. 이때, 데이터 생성 모델은 노이즈와 클래스의 라벨을 이용하여 희소 클래스의 특징으로 판별되도록 하기 위한 가짜 데이터를 생성하도록 학습되는 생성기 및 가짜 데이터와 희소 클래스의 특징, 및 클래스의 라벨을 이용하여 가짜 데이터를 가짜로 판별하도록 학습되는 판별기를 포함하는 CGAN(Conditional Generative Adversarial Networks) 모델일 수 있다.

[0097] 단계(940)에서 분류기 학습 장치(101)는 단계(910)에서 획득한 원본 학습 데이터 세트 및 단계(930)에서 생성한 추가 데이터들을 이용하여 침입 탐지 여부를 분류하기 위한 분류기를 학습시킬 수 있다.

[0098] 단계(950)에서 침입 탐지 장치(102)는 네트워크 데이터를 입력받을 수 있다. 그리고, 침입 탐지 장치(102)는 입력받은 네트워크 데이터를 분류기 학습 장치(101)로부터 수신한 침입 탐지 모델의 분류기에 입력할 수 있다. 다음으로, 침입 탐지 장치(102)는 침입 탐지 모델의 분류기의 출력에 따라 네트워크의 침입 여부를 탐지할 수 있다.

[0099] 이때, 학습 데이터 생성 장치(100)의 학습 데이터 생성 방법으로 단계(910) 내지 단계(930)만 수행될 수도 있다. 또한, 학습 데이터 생성 장치(100) 및 분류기 학습 장치(101)를 포함하는 분류기 학습 시스템의 침입 탐지 모델 학습 방법으로 단계(910) 내지 단계(940)만 수행될 수도 있다.

[0100] 본 발명은 오토 인코더로 원본 학습 데이터 세트에 포함된 희귀 클래스의 저차원 데이터를 추출하고, CGAN 모델로 희귀 클래스의 저차원 데이터를 포함하는 추가 데이터를 생성함으로써, 시간 복잡성과 공간 복잡성이 낮으며, 데이터 불균형이 해소된 데이터를 침입 탐지 모델의 분류기를 학습시키기 위하여 제공할 수 있다.

[0101] 또한, 본 발명은 시간 복잡성과 공간 복잡성이 낮으며, 데이터 불균형이 해소된 데이터를 이용하여 학습된 침입 탐지 모델의 분류기로 네트워크 침입을 탐지함으로써, 네트워크 침입 탐지 성능을 향상시킬 수 있다.

[0102] 한편, 본 발명에 따른 네트워크 침입 탐지 시스템 또는 네트워크 침입 탐지 방법은 컴퓨터에서 실행될 수 있는 프로그램으로 작성되어 마그네틱 저장매체, 광학적 판독매체, 디지털 저장매체 등 다양한 기록 매체로도 구현될 수 있다.

[0103] 본 명세서에 설명된 각종 기술들의 구현들은 디지털 전자 회로조직으로, 또는 컴퓨터 하드웨어, 펌웨어, 소프트웨어로, 또는 그들의 조합들로 구현될 수 있다. 구현들은 데이터 처리 장치, 예를 들어 프로그램가능 프로세서, 컴퓨터, 또는 다수의 컴퓨터들의 동작에 의한 처리를 위해, 또는 이 동작을 제어하기 위해, 컴퓨터 프로그램 제품, 예를 들어 기계 판독가능 저장 장치(컴퓨터 판독가능 매체)에서 유형적으로 구체화된 컴퓨터 프로그램으로서 구현될 수 있다. 상술한 컴퓨터 프로그램(들)과 같은 컴퓨터 프로그램은 컴파일된 또는 인터프리트된 언어들을 포함하는 임의의 형태의 프로그래밍 언어로 기록될 수 있고, 독립형 프로그램으로서 또는 모듈, 구성요소, 서브루틴, 또는 컴퓨팅 환경에서의 사용에 적절한 다른 유닛으로서 포함하는 임의의 형태로 전개될 수 있다. 컴퓨터 프로그램은 하나의 사이트에서 하나의 컴퓨터 또는 다수의 컴퓨터들 상에서 처리되도록 또는 다수의 사이트들에 걸쳐 분배되고 통신 네트워크에 의해 상호 연결되도록 전개될 수 있다.

[0104] 컴퓨터 프로그램의 처리에 적절한 프로세서들은 예로서, 범용 및 특수 목적 마이크로프로세서들 둘 다, 및 임의의 종류의 디지털 컴퓨터의 임의의 하나 이상의 프로세서들을 포함한다. 일반적으로, 프로세서는 판독 전용 메모리 또는 랜덤 액세스 메모리 또는 둘 다로부터 명령어들 및 데이터를 수신할 것이다. 컴퓨터의 요소들은 명령어들을 실행하는 적어도 하나의 프로세서 및 명령어들 및 데이터를 저장하는 하나 이상의 메모리 장치들을 포함할 수 있다. 일반적으로, 컴퓨터는 데이터를 저장하는 하나 이상의 대량 저장 장치들, 예를 들어 자기, 자기-광 디스크들, 또는 광 디스크들을 포함할 수 있거나, 이것들로부터 데이터를 수신하거나 이것들에 데이터를 송신하거나 또는 양쪽으로 되도록 결합될 수도 있다. 컴퓨터 프로그램 명령어들 및 데이터를 구체화하는데 적절한 정보 캐리어들은 예로서 반도체 메모리 장치들, 예를 들어, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM(Compact Disk Read Only Memory), DVD(Digital Video Disk)와 같은 광 기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-Optical Media), 롬(ROM, Read Only Memory), 램(RAM, Random Access Memory), 플래시 메모리, EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM) 등을 포함한다. 프로세서 및 메모리는 특수 목적 논리 회로조직에 의해 보충되거나, 이에 포함될 수 있다.

[0105] 또한, 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용매체일 수 있고, 컴퓨터 저장매체를 모두 포함할 수 있다.

[0106] 본 명세서는 다수의 특정한 구현물의 세부사항들을 포함하지만, 이들은 어떠한 발명이나 청구 가능한 것의 범위에 대해서도 제한적인 것으로서 이해되어서는 안되며, 오히려 특정한 발명의 특정한 실시형태에 특유할 수 있는 특징들에 대한 설명으로서 이해되어야 한다. 개별적인 실시형태의 문맥에서 본 명세서에 기술된 특정한 특징들은 단일 실시형태에서 조합하여 구현될 수도 있다. 반대로, 단일 실시형태의 문맥에서 기술한 다양한 특징들 역시 개별적으로 혹은 어떠한 적절한 하위 조합으로도 복수의 실시형태에서 구현 가능하다. 나아가, 특징들이 특정한 조합으로 동작하고 초기에 그와 같이 청구된 바와 같이 묘사될 수 있지만, 청구된 조합으로부터의 하나 이상의 특징들은 일부 경우에 그 조합으로부터 배제될 수 있으며, 그 청구된 조합은 하위 조합이나 하위 조합의 변형물로 변경될 수 있다.

[0107] 마찬가지로, 특정한 순서로 도면에서 동작들을 묘사하고 있지만, 이는 바람직한 결과를 얻기 위하여 도시된 그 특정한 순서나 순차적인 순서대로 그러한 동작들을 수행하여야 한다거나 모든 도시된 동작들이 수행되어야 하는 것으로 이해되어서는 안 된다. 특정한 경우, 멀티태스킹과 병렬 프로세싱이 유리할 수 있다. 또한, 상술한 실시형태의 다양한 장치 컴포넌트의 분리는 그러한 분리를 모든 실시형태에서 요구하는 것으로 이해되어서는 안되며, 설명한 프로그램 컴포넌트와 장치들은 일반적으로 단일의 소프트웨어 제품으로 함께 통합되거나 다중 소프트웨어 제품에 패키징 될 수 있다는 점을 이해하여야 한다.

[0108] 한편, 본 명세서와 도면에 개시된 본 발명의 실시 예들은 이해를 돕기 위해 특정 예를 제시한 것에 지나지 않으며, 본 발명의 범위를 한정하고자 하는 것은 아니다. 여기에 개시된 실시 예들 이외에도 본 발명의 기술적 사상에 바탕을 둔 다른 변형 예들이 실시 가능하다는 것은, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 자명한 것이다.

부호의 설명

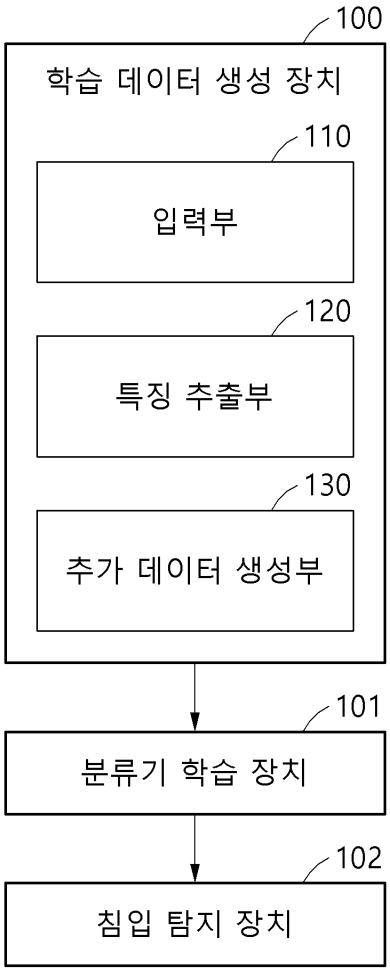
[0109] 100: 학습 데이터 생성 장치

101: 분류기 학습 장치

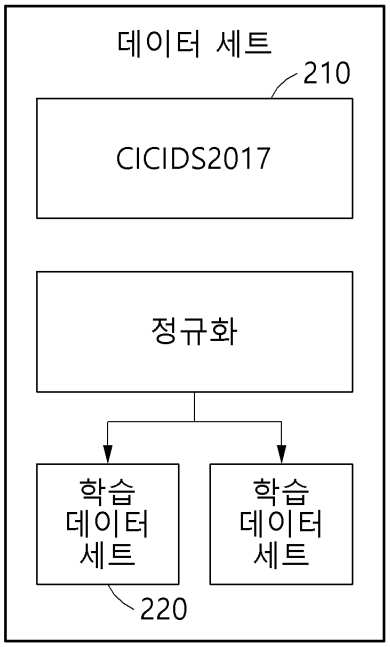
102: 침입 탐지 장치

도면

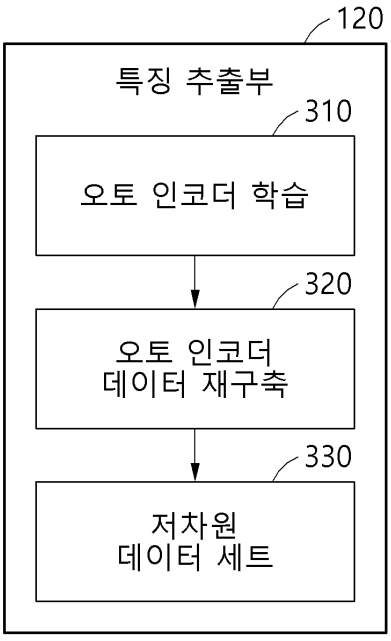
도면1



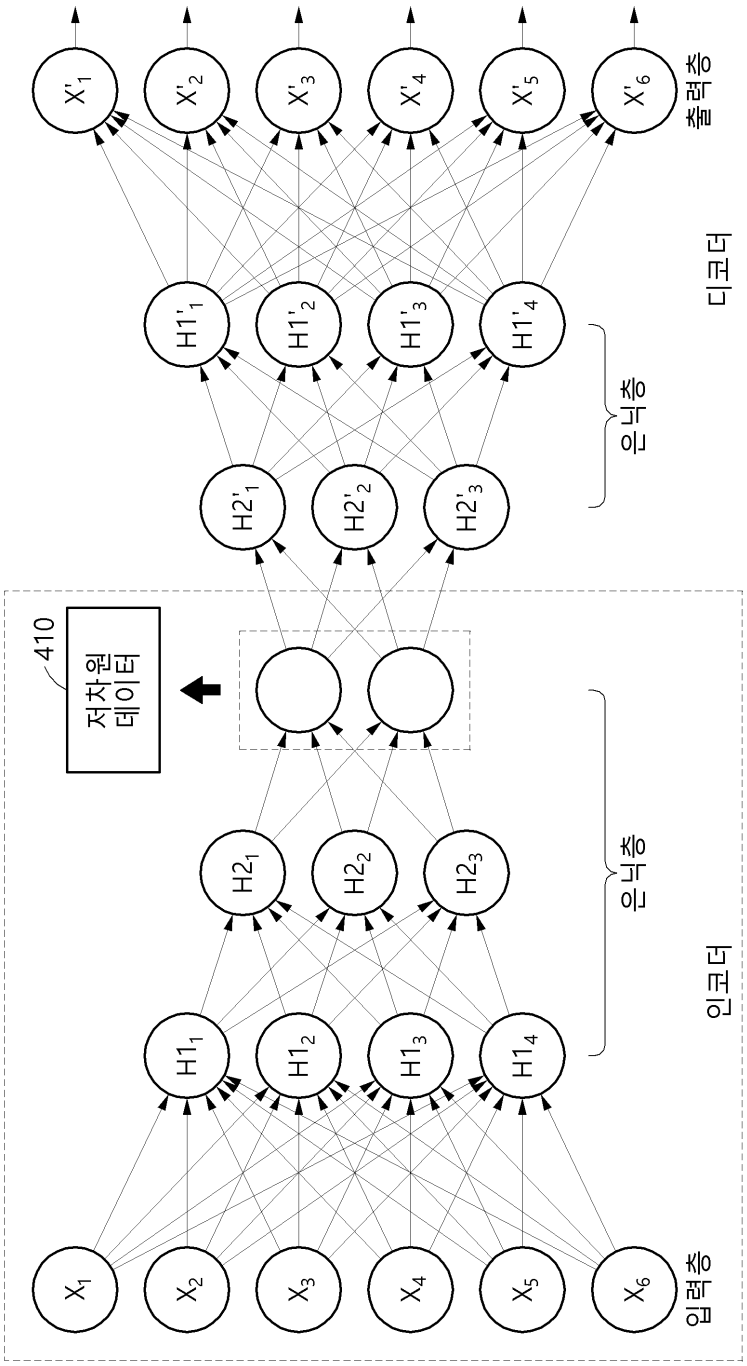
도면2


데이터 세트
210
CICIDS2017
정규화
학습
데이터
세트
학습
데이터
세트
220

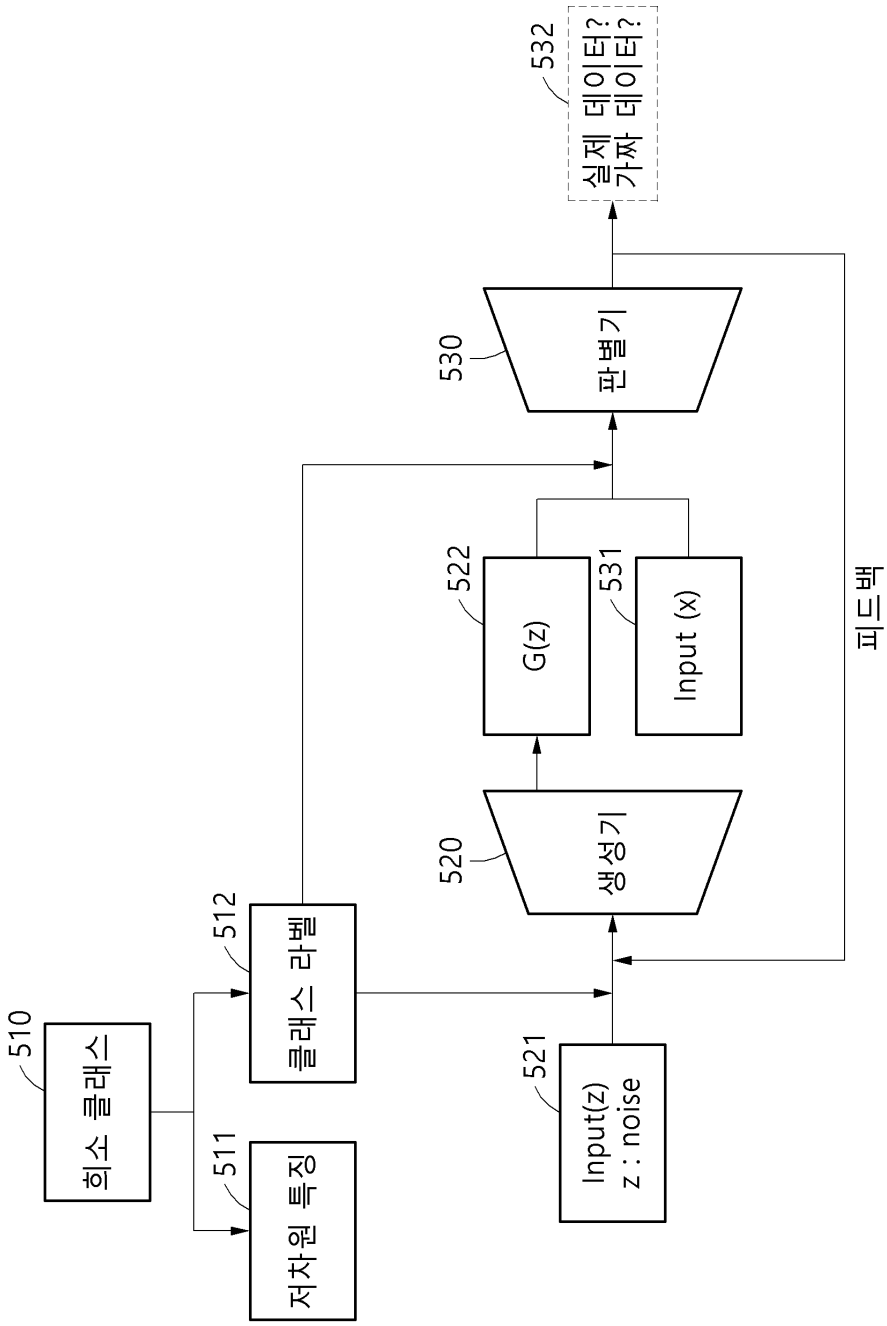
도면3


120
특징 추출부
310
오토 인코더 학습
320
오토 인코더
데이터 재구축
330
저차원
데이터 세트

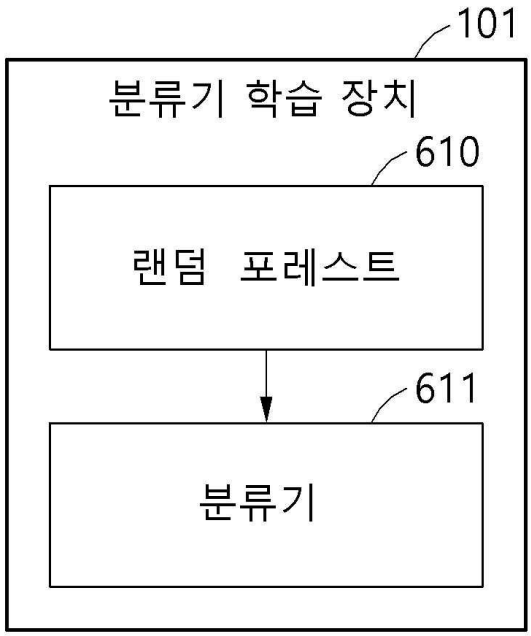
도면4


X'1
X'2
X'3
X'4
X'5
X'6
출력층
H1'1
H1'2
H1'3
H1'4
H2'1
H2'2
H2'3
은닉층
디코더
410
저차원
데이터
H21
H22
H23
H11
H12
H13
H14
은닉층
인코더
X1
X2
X3
X4
X5
X6
입력층

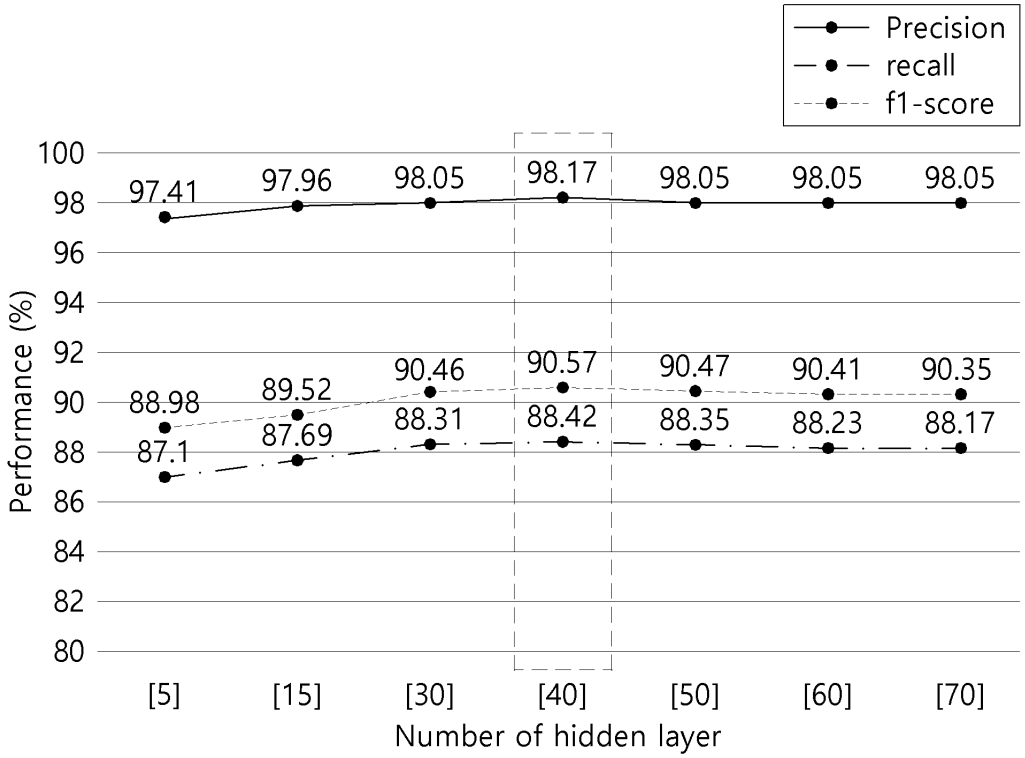
도면5


532
실제 데이터?
가짜 데이터?
530
판별기
522
G(z)
531
Input (x)
520
생성기
510
희소 클래스
512
클래스 라벨
511
저차원 특징
521
Input(z)
z : noise
피드백

도면6


101
분류기 학습 장치
610
랜덤 포레스트
611
분류기

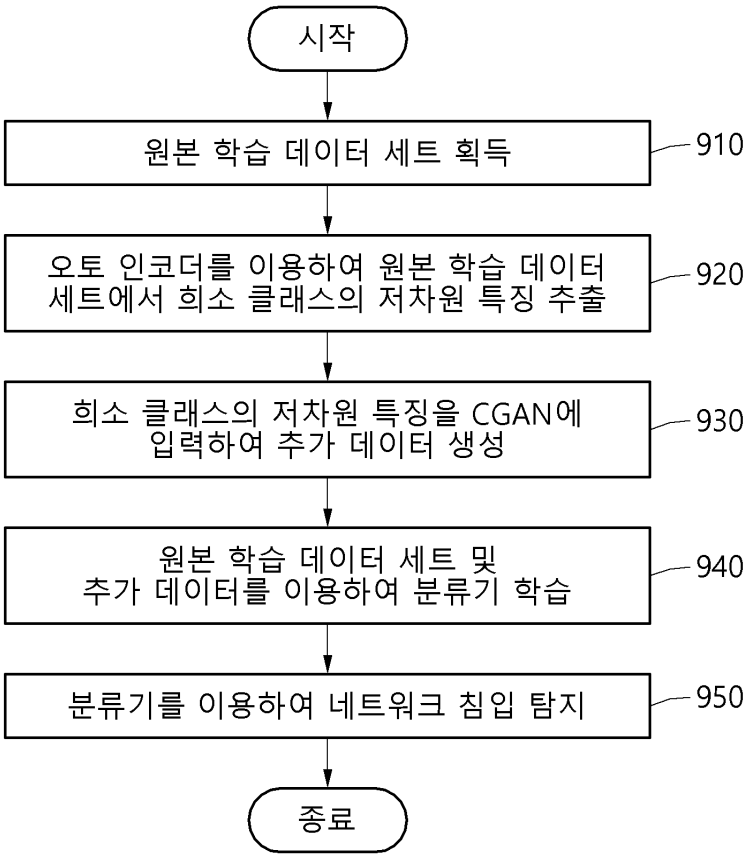
도면7


Precision
recall
f1-score
Performance (%)
100
98
96
94
92
90
88
86
84
82
80
97.41
97.96
98.05
98.17
98.05
98.05
98.05
88.98
89.52
90.46
90.57
90.47
90.41
90.35
87.1
87.69
88.31
88.42
88.35
88.23
88.17
[5]
[15]
[30]
[40]
[50]
[60]
[70]
Number of hidden layer

도면8

Actual
BENIGN
DDoS
PortScans
Bot
Infiltration
Web attack
FTP-Patator
SSH-Patator
DoS GoldenEye
DoS Hulk
DoS Slowhttptes
DoS Slowloris
HeartBleed
Prediction
BENIGN 907719 33 11 357 5 43 5 6 10 234 11 6 0
DDoS 2 51170 0 0 0 0 0 0 0 1 0 0 0
PortScans 390 0 63494 0 0 2 0 0 0 1 1 1 0
Bot 83 0 0 426 0 0 0 0 0 0 0 0 0
Infiltration 0 0 0 0 10 0 0 0 0 0 0 0 0
Web attack 0 0 3 0 0 827 0 0 0 1 0 1 0
FTP-Patator 0 0 0 0 0 0 3169 0 0 0 0 0 0
SSH-Patator 0 0 0 0 0 0 0 2353 0 0 0 0 0
DoS GoldenEye 10 0 0 0 0 0 0 0 4095 11 2 1 0
DoS Hulk 310 0 0 0 0 0 0 0 12 91802 0 0 0
DoS Slowhttptes 14 0 0 0 0 0 0 0 1 0 2177 7 0
DoS Slowloris 0 0 0 0 0 0 0 0 0 0 9 2303 0
HeartBleed 0 0 0 0 0 0 0 0 0 0 0 0 5

도면9


시작
원본 학습 데이터 세트 획득
910
오토 인코더를 이용하여 원본 학습 데이터 세트에서 희소 클래스의 저차원 특징 추출
920
희소 클래스의 저차원 특징을 CGAN에 입력하여 추가 데이터 생성
930
원본 학습 데이터 세트 및 추가 데이터를 이용하여 분류기 학습
940
분류기를 이용하여 네트워크 침입 탐지
950
종료