(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0105082
(43) 공개일자 2024년07월05일

(51) 국제특허분류(Int. Cl.)
H04L 9/40 (2022.01) *G06N 3/044* (2023.01)
(52) CPC특허분류
H04L 63/1416 (2013.01)
G06N 3/044 (2023.01)
(21) 출원번호 10-2022-0187932
(22) 출원일자 2022년12월28일
심사청구일자 2022년12월28일

(71) 출원인
포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)
한동대학교 산학협력단
경상북도 포항시 북구 흥해읍 한동로 558
(72) 발명자
홍원기
경상북도 포항시 남구 청암로 77
유재형
경상북도 포항시 남구 청암로 77
(뒷면에 계속)
(74) 대리인
특허법인이상

전체 청구항 수 : 총 12 항

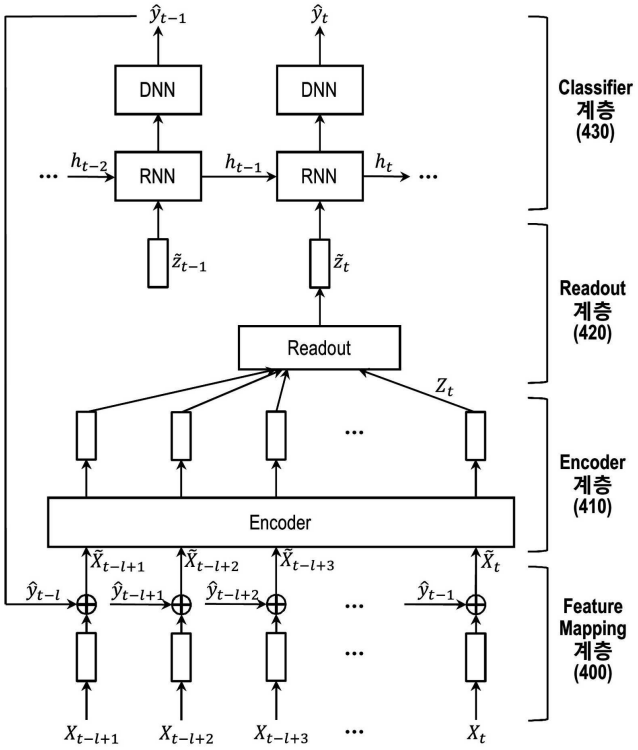
(54) 발명의 명칭 **머신러닝 기반 네트워크 공격 및 침입 탐지 방법 및 장치**

(57) 요 약

인공신경망 모델에 의해 SDN/NFV 환경에서 네트워크 공격 및 침입 상태를 탐지함에 있어, 실제 네트워크 환경에서 획득한 데이터를 토대로 학습 데이터셋을 구축하여 모델의 학습에 사용하고, 상기 인공신경망 모델이 네트워크 트래픽 데이터의 시계열 특성을 고려하도록 함으로써, 실제 네트워크 환경에서 높은 탐지 정확도를 시현할 수

(뒷면에 계속)

대 표 도 - 도2



있는 네트워크 공격 및 침입 탐지 방법 및 장치를 제공한다. 네트워크 공격 및 침입 탐지 방법은, 네트워크를 통해 전달되는 데이터 플로우 또는 패킷으로부터 트래픽 데이터를 획득하고, 획득한 네트워크 트래픽 데이터를 토대로 학습용 데이터셋을 생성하는 단계; 상기 학습용 데이터셋을 사용하여 소정의 이상상태 탐지 신경망 모델을 학습시키는 단계; 상기 네트워크로부터 공격 또는 침입 상태 판정이 필요한 트래픽 데이터를 받아들이고, 시계열 트래픽 데이터를 저장하는 단계; 및 학습된 이상상태 탐지 신경망에 의한 추론에 의해 상기 시계열 트래픽 데이터에 공격 또는 침입 징후가 있는지 결정하는 단계를 포함한다.

(72) 발명자

홍지범

경상북도 포항시 남구 청암로 77

최희열

경상북도 포항시 북구 흥해읍 초곡지구로 102, 102동 904호

허동녕

경상북도 포항시 북구 천마로46번길 16-1, 302호

이청준

서울특별시 송파구 성내천로18길 16

이 발명을 지원한 국가연구개발사업

과제고유번호	1711152960
과제번호	2018-0-00749-005
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	방송통신산업기술개발(R&D,정보화)
연구과제명	인공지능 기반 가상 네트워크 관리기술 개발
기 여 율	1/1
과제수행기관명	포항공과대학교 산학협력단
연구기간	2022.01.01 ~ 2022.12.31

명 세 서

청구범위

청구항 1

네트워크를 통해 전달되는 데이터 플로우 또는 패킷으로부터 트래픽 데이터를 획득하고, 획득한 네트워크 트래픽 데이터를 토대로 학습용 데이터셋을 생성하는 단계;

상기 학습용 데이터셋을 사용하여 소정의 이상상태 탐지 신경망 모델을 학습시키는 단계;

상기 네트워크로부터 공격 또는 침입 상태 판정이 필요한 트래픽 데이터를 받아들이고, 시계열 트래픽 데이터를 저장하는 단계; 및

학습된 이상상태 탐지 신경망에 의한 추론에 의해 상기 시계열 트래픽 데이터에 공격 또는 침입 징후가 있는지 결정하는 단계;

를 포함하는,

네트워크 공격 및 침입 탐지 방법.

청구항 2

청구항 1에 있어서, 상기 학습용 데이터셋을 생성하는 단계가

상기 획득한 네트워크 트래픽 데이터에 결함을 주입하는 단계; 및

결함이 주입된 데이터를 토대로 상기 학습용 데이터셋을 생성하는 단계;

를 포함하는,

네트워크 공격 및 침입 탐지 방법.

청구항 3

청구항 1에 있어서, 상기 시계열 트래픽 데이터에 공격 또는 침입 징후가 있는지 결정하는 단계가

상기 시계열 트래픽 데이터에 대하여 스케일링을 포함한 전처리를 수행하는 단계; 및

전처리가 완료된 데이터에서 상기 공격 또는 침입 징후를 탐지하는 단계;

를 포함하는,

네트워크 공격 및 침입 탐지 방법.

청구항 4

청구항 3에 있어서, 상기 이상상태 탐지 신경망이

입력 데이터 시퀀스를 소정의 차원을 가지는 특징 벡터로 벡터화하는 특징 맵핑 계층;

상기 특징 벡터의 크기를 감소시키는 인코더 계층;

상기 인코더 계층의 출력 벡터를 풀링 또는 셀프-어텐션에 의하여 하나의 벡터로 축약시키는 판독 계층; 및

상기 판독 계층에 의해 축약된 벡터를 토대로, RNN을 포함하는 신경망에 의하여 분류해서, 상기 공격 또는 침입 징후가 있는지를 탐지하는 분류기 계층;

을 포함하는,

네트워크 공격 및 침입 탐지 방법.

청구항 5

청구항 4에 있어서,

상기 특징 맵핑 계층이, 상기 분류기 계층로부터 이전 시간 단계에서의 분류 결과를 받아들이고 상기 입력 데이터 시퀀스 또는 상기 특징 벡터에 상기 분류 결과를 결합시키는 단계;

를 더 포함하는,

네트워크 공격 및 침입 탐지 방법.

청구항 6

청구항 5에 있어서, 상기 특징 벡터에 상기 분류 결과를 결합시키는 단계는

상기 특징 맵핑 계층이, 소정의 시간 단계들에 상응한 상기 분류기 계층의 분류 결과들을 받아들이는 단계;

상기 입력 데이터 시퀀스 또는 상기 특징 벡터에 상기 분류기 계층으로부터의 분류 결과들을 직합 연산하여 결합시키는,

네트워크 공격 및 침입 탐지 방법.

청구항 7

프로그램 명령들을 저장하는 메모리와; 상기 메모리에 접속되고 상기 메모리에 저장된 상기 프로그램 명령들을 실행하는 프로세서;를 구비하며,

상기 프로그램 명령들은 상기 프로세서에 의해 실행될 때 상기 프로세서로 하여금:

네트워크를 통해 전달되는 데이터 플로우 또는 패킷으로부터 트래픽 데이터를 획득하고, 획득한 네트워크 트래픽 데이터를 토대로 학습용 데이터셋을 생성하며;

상기 학습용 데이터셋을 사용하여 소정의 이상상태 탐지 신경망 모델을 학습시키고;

상기 네트워크로부터 공격 또는 침입 상태 판정이 필요한 트래픽 데이터를 받아들이고, 시계열 트래픽 데이터를 저장하며;

학습된 이상상태 탐지 신경망에 의한 추론에 의해 상기 시계열 트래픽 데이터에 공격 또는 침입 징후가 있는지 결정하게 하는,

네트워크 공격 및 침입 탐지 장치.

청구항 8

청구항 7에 있어서, 상기 프로세서로 하여금 상기 학습용 데이터셋을 생성하게 하는 프로그램 명령들은 상기 프로세서로 하여금:

상기 획득한 네트워크 트래픽 데이터에 결함을 주입하고;

결함이 주입된 데이터를 토대로 상기 학습용 데이터셋을 생성하게 하는,

네트워크 공격 및 침입 탐지 장치.

청구항 9

청구항 7에 있어서, 상기 프로세서로 하여금 상기 시계열 트래픽 데이터에 공격 또는 침입 징후가 있는지 결정하게 하는 프로그램 명령들은 상기 프로세서로 하여금:

상기 시계열 트래픽 데이터에 대하여 스케일링을 포함한 전처리를 수행하고;

전처리가 완료된 데이터에서 상기 공격 또는 침입 징후를 탐지하게 하는,

네트워크 공격 및 침입 탐지 장치.

청구항 10

청구항 9에 있어서, 상기 이상상태 탐지 신경망이

입력 데이터 시퀀스를 소정의 차원을 가지는 특징 벡터로 벡터화하는 특징 맵핑 계층;

상기 특징 벡터의 크기를 감소시키는 인코더 계층;

상기 인코더 계층의 출력 벡터를 풀링 또는 셀프-어텐션에 의하여 하나의 벡터로 축약시키는 판독 계층; 및

상기 판독 계층에 의해 축약된 벡터를 토대로, RNN을 포함하는 신경망에 의하여 분류해서, 상기 공격 또는 침입 징후가 있는지를 탐지하는 분류기 계층;

을 포함하는,

네트워크 공격 및 침입 탐지 장치.

청구항 11

청구항 10에 있어서, 상기 프로그램 명령들은 상기 프로세서로 하여금:

상기 특징 맵핑 계층이, 상기 분류기 계층로부터 이전 시간 단계에서의 분류 결과를 받아들이고 상기 입력 데이터 시퀀스 또는 상기 특징 벡터에 상기 분류 결과를 결합시키게 하는,

네트워크 공격 및 침입 탐지 장치.

청구항 12

청구항 11에 있어서, 상기 프로세서로 하여금 상기 특징 벡터에 상기 분류 결과를 결합시키게 하는 프로그램 명령들은 상기 프로세서로 하여금:

상기 특징 맵핑 계층이, 소정의 시간 단계들에 상응한 상기 분류기 계층의 분류 결과들을 받아들이고;

상기 입력 데이터 시퀀스 또는 상기 특징 벡터에 상기 분류기 계층으로부터의 분류 결과들을 직합 연산하여 결합시키도록 하는,

네트워크 공격 및 침입 탐지 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 네트워크 보안을 위한 방법 및 장치에 관한 것으로서, 보다 상세하게는 네트워크 공격이나 침입과 같은 비정상적인 트래픽을 탐지하는 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 소프트웨어 정의 네트워킹(SDN: Software-Defined Networking) 및 네트워크 기능 가상화(NFV: Network Function Virtualization) 기술의 급속한 발전으로 통신사업자와 클라우드 데이터 센터 사업자들이 네트워크 기능을 가상화한 가상 네트워크 기능(VNF: Virtualized Network Function)를 도입하여 운용하는 예가 증가하고 있다. VNF의 규모와 복잡성이 증가함에 따라 SDN/NFV 환경의 자원 할당(resource allocation)과 성능 관리, 장애 관리(fault management), 및 보안(security) 등과 같은 관리 문제가 중요해지고 있다. 특히, 안정적인 네트워크 관리를 위해서는 DoS/DDoS와 같은 네트워크 공격이나 침입과 같은 비정상적인 트래픽을 초기에 탐지하고 예방하는 것이 중요하다. 이를 위해서는 SDN/NFV 환경에서 트래픽 정보를 실시간으로 파악하고 분석해야 한다.

[0003] 네트워크 공격 및 침입 탐지는 데이터 센터 내부에서 운용되는 물리 네트워크 및 자원은 물론, 가상 머신(Virtual Machine, VM) 및 VNF와 같이 SDN/NFV 환경에서 동작하는 가상 자원 및 가상 네트워크 관리와 보안의 중요한 요소이다. 서비스 제공자 및 네트워크 관리자는 SDN/NFV 환경에서 제공되는 서비스들이 정상적으로 동작하고 있는지 파악하고, 비정상적인 상황 발생 시 그에 맞는 조치 및 정책을 실행하기 위해 네트워크 공격 및 침입 탐지 방법을 사용한다.

[0004] 네트워크 공격 및 침입 탐지 방법의 일 예로서, 페이로드에서 포트 번호 또는 시그너처(signature)를 사용하여 공격 및 침입 상태를 탐지하는 방법을 들 수 있다. 예를 들어, 시그너처 기반 접근 방식에서는, 시스템에 공격 시그너처가 사전에 설치되고, 시스템에 설치된 시그너처에 대한 트래픽 패턴의 일치 여부를 판단하여 네트워크 공격 및 침입을 탐지한다. 그런데, 이와 같이 IP 패킷의 페이로드에 포함된 포트 번호 또는 시그너처를 사용하

여 트래픽을 분류하고 공격 및 침입 상태를 탐지하는 방법은 새로운 공격 방법에 대하여 취약할 수 있기 때문에, 네트워크 공격 방법이 다양화됨에 따라 더 큰 한계를 보일 수 있다.

[0005] 네트워크 공격 및 침입 탐지 방법의 다른 예로서, 트래픽의 통계 정보로부터 공격 및 침입 행동이나 특성을 분석하는 통계 기반 방법이 있다. 이 방법에 따르면, 동작이나 메트릭 값이 규칙적 또는 주기적인 패턴을 벗어날 때 특정 네트워크 트래픽을 공격 및 침입으로 간주한다. 그렇지만, 이 방법은 공격 및 침입으로 인한 패턴 변화가 현저하지 않으면, 정상적인 탐지가 어려울 수 있다.

[0006] 최근에는, 통계 정보를 기반으로 특성을 분석하여 비정상 트래픽을 탐지하는 통계 기반 방법에 머신러닝(machine learning) 내지 딥러닝과 같은 인공지능 기술을 접목시켜서, 사람의 개입없이 공격 및 침입 트래픽을 탐지하고 네트워크를 관리하려는 시도도 증가하고 있다. 이러한 방식에서 사용되는 모델들은 예컨대 KDD 99, NSL-KDD 등과 같은 공개된 데이터셋을 사용하여 지도학습(supervised learning) 및/또는 비지도학습(unsupervised learning)으로 학습되어, 공격 및 침입 트래픽을 탐지한다. 그렇지만, 이러한 모델들은 일반적으로 학습이 완료된 후 실제 네트워크에 적용되었을 때에 학습이나 학습 성능 검증 시에 비하여 성능이 떨어진다는 문제가 있다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 등록특허공보 10-2454075호

발명의 내용

해결하려는 과제

[0008] 본 출원의 발명자들은 인공신경망 모델들의 위와 같은 학습 시의 성능과 실제 추론 시의 성능 차이의 가장 큰 원인이, 시뮬레이션 환경에서 생성된 공개된 데이터셋을 사용하여 모델을 학습시킨다는 것과, 모델 학습 과정에서 네트워크 트래픽 데이터의 시계열 특성을 충분히 고려하지 않는 것이라고 보고 있다.

[0009] 이와 같은 문제점을 해결하기 위하여, 본 발명은 인공신경망 모델에 의해 SDN/NFV 환경에서 네트워크 공격 및 침입 상태를 탐지함에 있어, 실제 네트워크 환경에서 획득한 데이터를 토대로 학습 데이터셋을 구축하여 모델의 학습에 사용하고, 상기 인공신경망 모델이 네트워크 트래픽 데이터의 시계열 특성을 고려하도록 함으로써, 실제 네트워크 환경에서 높은 탐지 정확도를 시현할 수 있는 네트워크 공격 및 침입 탐지 방법 및 장치를 제공하는 것을 기술적 과제로 한다.

과제의 해결 수단

[0010] 예시적인 실시예의 일 측면에 따른 네트워크 공격 및 침입 탐지 방법은, 네트워크를 통해 전달되는 데이터 플로우 또는 패킷으로부터 트래픽 데이터를 획득하고, 획득한 네트워크 트래픽 데이터를 토대로 학습용 데이터셋을 생성하는 단계; 상기 학습용 데이터셋을 사용하여 소정의 이상상태 탐지 신경망 모델을 학습시키는 단계; 상기 네트워크로부터 공격 또는 침입 상태 판정이 필요한 트래픽 데이터를 받아들이고, 시계열 트래픽 데이터를 저장하는 단계; 및 학습된 이상상태 탐지 신경망에 의한 추론에 의해 상기 시계열 트래픽 데이터에 공격 또는 침입 징후가 있는지 결정하는 단계;를 포함한다.

[0011] 상기 학습용 데이터셋을 생성하는 단계는 상기 획득한 네트워크 트래픽 데이터에 결함을 주입하는 단계; 및 결함이 주입된 데이터를 토대로 상기 학습용 데이터셋을 생성하는 단계;를 포함할 수 있다.

[0012] 상기 시계열 트래픽 데이터에 공격 또는 침입 징후가 있는지 결정하는 단계는 상기 시계열 트래픽 데이터에 대하여 스케일링을 포함한 전처리를 수행하는 단계; 및 전처리가 완료된 데이터에서 상기 공격 또는 침입 징후를 탐지하는 단계;를 포함할 수 있다.

[0013] 상기 이상상태 탐지 신경망은 입력 데이터 시퀀스를 소정의 차원을 가지는 특징 벡터로 벡터화하는 특징 맵핑 계층; 상기 특징 벡터의 크기를 감소시키는 인코더 계층; 상기 인코더 계층의 출력 벡터를 풀링 또는 셀프-어텐션에 의하여 하나의 벡터로 축약시키는 판독 계층; 및 상기 판독 계층에 의해 축약된 벡터를 토대로, RNN을 포함하는 신경망에 의하여 분류해서, 상기 공격 또는 침입 징후가 있는지를 탐지하는 분류기 계층;을 포함할 수

있다.

[0014] 상기 특징 맵핑 계층은 상기 분류기 계층로부터 이전 시간 단계에서의 분류 결과를 받아들이고 상기 입력 데이터 시퀀스 또는 상기 특징 벡터에 상기 분류 결과를 결합시킬 수 있다.

[0015] 상기 특징 벡터에 상기 분류 결과를 결합시키는 단계는 상기 특징 맵핑 계층이, 소정의 시간 단계들에 상응한 상기 분류기 계층의 분류 결과들을 받아들이는 단계; 상기 입력 데이터 시퀀스 또는 상기 특징 벡터에 상기 분류기 계층으로부터의 분류 결과들을 직합 연산하여 결합시킬 수 있다.

[0016] 예시적인 실시예의 다른 측면에 따른 네트워크 공격 및 침입 탐지 장치는 프로그램 명령들을 저장하는 메모리와; 상기 메모리에 접속되고 상기 메모리에 저장된 상기 프로그램 명령들을 실행하는 프로세서;를 구비한다. 상기 프로그램 명령들은 상기 프로세서에 의해 실행될 때 상기 프로세서로 하여금: 네트워크를 통해 전달되는 데이터 플로우 또는 패킷으로부터 트래픽 데이터를 획득하고, 획득한 네트워크 트래픽 데이터를 토대로 학습용 데이터셋을 생성하며; 상기 학습용 데이터셋을 사용하여 소정의 이상상태 탐지 신경망 모델을 학습시키고; 상기 네트워크로부터 공격 또는 침입 상태 판정이 필요한 트래픽 데이터를 받아들이고, 시계열 트래픽 데이터를 저장하며; 학습된 이상상태 탐지 신경망에 의한 추론에 의해 상기 시계열 트래픽 데이터에 공격 또는 침입 징후가 있는지 결정하게 한다.

[0017] 상기 프로세서로 하여금 상기 학습용 데이터셋을 생성하게 하는 프로그램 명령들은 상기 프로세서로 하여금: 상기 획득한 네트워크 트래픽 데이터에 결함을 주입하고; 결함이 주입된 데이터를 토대로 상기 학습용 데이터셋을 생성하게 할 수 있다.

[0018] 상기 프로세서로 하여금 상기 시계열 트래픽 데이터에 공격 또는 침입 징후가 있는지 결정하게 하는 프로그램 명령들은 상기 프로세서로 하여금: 상기 시계열 트래픽 데이터에 대하여 스케일링을 포함한 전처리를 수행하고; 전처리가 완료된 데이터에서 상기 공격 또는 침입 징후를 탐지하게 할 수 있다.

[0019] 상기 이상상태 탐지 신경망은 입력 데이터 시퀀스를 소정의 차원을 가지는 특징 벡터로 벡터화하는 특징 맵핑 계층; 상기 특징 벡터의 크기를 감소시키는 인코더 계층; 상기 인코더 계층의 출력 벡터를 풀링 또는 셀프-어텐션에 의하여 하나의 벡터로 축약시키는 판독 계층; 및 상기 판독 계층에 의해 축약된 벡터를 토대로, RNN을 포함하는 신경망에 의하여 분류해서, 상기 공격 또는 침입 징후가 있는지를 탐지하는 분류기 계층;을 포함하도록 구성될 수 있다.

[0020] 상기 프로그램 명령들은 상기 특징 맵핑 계층이, 상기 분류기 계층로부터 이전 시간 단계에서의 분류 결과를 받아들이고 상기 입력 데이터 시퀀스 또는 상기 특징 벡터에 상기 분류 결과를 결합시키도록 할 수 있다.

[0021] 상기 프로세서로 하여금 상기 특징 벡터에 상기 분류 결과를 결합시키게 하는 프로그램 명령들은 상기 프로세서로 하여금: 상기 특징 맵핑 계층이, 소정의 시간 단계들에 상응한 상기 분류기 계층의 분류 결과들을 받아들이고; 상기 입력 데이터 시퀀스 또는 상기 특징 벡터에 상기 분류기 계층으로부터의 분류 결과들을 직합 연산하여 결합시키도록 할 수 있다.

발명의 효과

[0022] 본 발명의 일 실시예에 따르면, 인공신경망 모델에 의해 SDN/NFV 환경에서 네트워크 공격 및 침입 상태를 탐지함에 있어서, 실제 네트워크 환경에서 획득한 데이터를 토대로 학습 데이터셋을 구축하여 모델의 학습에 사용하고, 상기 인공신경망 모델이 네트워크 트래픽 데이터의 시계열 특성이 충분히 고려되도록 한다. 이에 따라, 실제 네트워크 환경에서 높은 탐지 정확도를 나타낼 수 있다. 이와 같은 인공신경망 모델을 토대로 네트워크 공격/침입 탐지 장치를 구현하는 경우, 높은 분류 정확도를 제공함으로써, 오탐지를 막고, 보다 정밀하게 네트워크 공격 및 침입 시도를 감지해낼 수 있다.

도면의 간단한 설명

[0023] 도 1은 본 발명의 예시적 실시예에 따른 네트워크 공격 및 침입 탐지 시스템의 블록도이다.

도 2는 도 1에 도시된 공격/침입 탐지 모델의 일 실시예의 블록도이다.

도 3은 도 2의 공격/침입 탐지 모델의 학습 과정의 일 예를 보여주는 흐름도이다.

도 4a 및 4b는 도 2의 이상상태 탐지 모델이 사용하는 데이터 속성들의 목록의 일 예를 보여주는 표이다.

도 5는 도 1의 네트워크 공격 및 침입 탐지 시스템에서의 탐지 동작의 일 예를 보여주는 흐름도이다.

도 6은 예시적 실시예에 따른 네트워크 공격/침입 탐지 장치의 물리적 구성을 보여주는 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0024] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 구성요소에 대해서는 유사한 참조부호를 사용하였다.

[0025] 제1, 제2, 등의 서수가 다양한 구성요소들을 설명하는 데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. "및/또는"이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0026] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[0027] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 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.

[0028] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0029] 이하, 본 발명에 따른 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다. 본 발명을 설명함에 있어 전체적인 이해를 용이하게 하기 위하여 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.

[0030] 도 1은 본 발명의 예시적 실시예에 따른 네트워크 공격 및 침입 탐지 시스템의 블록도이다.

[0031] 도 1의 네트워크 공격 및 침입 탐지 시스템은 모니터링 대상이 되는 네트워크(10, 20)와, 상기 네트워크(10, 20)에 대한 공격이나 침입을 탐지하는 네트워크 공격/침입 탐지 장치(100)를 포함한다.

[0032] 모니터링 대상인 네트워크 중 적어도 일부는 네트워크 기능이 가상화(Network Function Virtualization, NFV)되어 전용 하드웨어 의존성이 큰 물리 네트워크(10)와 별도의 가상 네트워크(20)를 구성할 수 있다. 가상 네트워크920)는 다양한 네트워크 소프트웨어 기능을 상용 하드웨어를 사용하여 유연하게 구현될 수 있다. 이러한 가상 네트워크(20)에서는, 네트워크 서비스를 구현하기 위한 하나 이상의 VNF(Virtualized Network Function)가 정의될 수 있다. 각각의 네트워크 서비스를 구현할 때 필요한 VNF에 대해, 물리적/가상적 네트워크 자원이 자동으로 할당될 수 있다. 예를 들어, MANO(Management and Orchestration, 50)는 다양한 요인에 따라 자동으로 물리 네트워크(10)와 가상 네트워크(20)에 대하여 컴퓨팅 자원을 할당할 수 있다.

[0033] MANO(50)는 정책 관리 모듈(60)과 오케스트레이터(Orchestrator, 70)를 포함할 수 있다. 정책 관리 모듈(60)은 관리자가 결정한 네트워크 관리 정책을 받아들이고 유지하며, 상기 관리 정책에 기초하여 물리 네트워크(10)와 가상 네트워크(20)를 관리한다. 오케스트레이터(70)는 정책 관리 모듈(60)이 유지하는 정책에 따라서, 네트워크(10, 20)를 자동으로 관리한다. 예컨대, 오케스트레이터(70)는 네트워크 서비스의 요구사항, 컴퓨팅 자원의 최대 성능 및 용량, 네트워크 사업자의 컴퓨팅 자원 관리 정책, 또는 네트워크 서비스 및 컴퓨팅 자원의 실시간 상황 변동 등의 요인에 따라 자동으로 물리 네트워크(10)와 가상 네트워크(20)에 대하여 컴퓨팅 자원을

할당할 수 있다. 도 1에는 물리 네트워크(10)와 가상 네트워크(20)가 모두 도시되어 있지만, 변형된 실시예에서는 모니터링 대상 네트워크가 물리 네트워크(10)와 가상 네트워크(20) 중 어느 하나만을 포함할 수도 있다.

[0034] 특히, 본 발명의 예시적 실시예에 따르면, 정책 관리 모듈(60)은 보안 모니터링 정책을 유지하고 관리하며, 상기 보안 모니터링 정책에 기초하여 오케스트레이터(70)의 보안 모니터링의 관리를 조율할 수 있다. 또한, 정책 관리 모듈(60)은 보안 규칙을 업데이트하거나 탐지 결과를 기반으로 서비스 품질 또는 보안 수준을 유지하기 위한 새로운 관리 정책을 생성할 수 있다. 오케스트레이터(70)는 정책 관리 모듈(60)이 제공하는 정책을 네트워크에 반영하며 네트워크를 관리한다. 특히, 오케스트레이터(70)는 정책 관리 모듈(60)의 보안 모니터링 정책 또는 명령에 따라서 특정 리소스의 할당 축소 또는 제한, 클라이언트의 접근 통제, 트래픽의 제한 등의 조치를 취할 수 있다.

[0035] 본 발명의 예시적 실시예에 따르면, 네트워크 공격 및 침입을 탐지하기 위하여, 물리 네트워크(10) 및/또는 가상 네트워크(20)에는 해당 네트워크에 대한 트래픽 정보를 감지하기 위한 모니터링 에이전트(미도시)가 설치될 수 있다. 모니터링 에이전트는 하드웨어, 소프트웨어, 또는 이들의 결합에 의해 구현될 수 있으며, 플로우 또는 패킷 단위로 트래픽 데이터를 수집하여 네트워크 공격/침입 탐지 장치(100)에 제공할 수 있다.

[0036] 일 실시예에 있어서, 모니터링 에이전트가 수집하여 네트워크 공격/침입 탐지 장치(100)에 제공하는 트래픽 정보는 네트워크 페이로드 패킷에 대한 5 튜플(tuple) 정보 즉, 송신자 IP, 송신 포트 번호, 수신자 IP, 수신 포트 번호, 및 사용 프로토콜 정보와 아울러, 대역폭 정보를 포함할 수 있다. 이하의 설명에서, 위와 같이 모니터링 에이전트가 수집하여 제공하는 5 튜플 정보 및 대역폭 정보를 '네트워크 트래픽 데이터'로 칭하기로 한다. 그렇지만, 모니터링 에이전트가 수집하는 정보의 종류가 이에 한정되는 것은 아니며, 후술하는 바와 같이 수집되는 네트워크 데이터의 모니터링 메트릭은 80가지를 넘을 수도 있음을 유의해야 한다.

[0037] 네트워크 공격/침입 탐지 장치(100)는 데이터 수집부(200)와 데이터 분석부(300)를 포함할 수 있다.

[0038] 데이터 수집부(200)는 물리 네트워크(10) 및/또는 가상 네트워크(20) 내에서 동작하는 모니터링 에이전트로부터 네트워크 트래픽 데이터를 받아들이고, 이를 데이터베이스(220)에 저장하고, 관리자에게 표출하며, 데이터 분석부(300)에 제공할 수 있다. 구체적으로, 데이터 수집부(200)에 있어서, 모니터링 모듈(210)은 모니터링 에이전트로부터 수신되는 네트워크 트래픽 데이터를 데이터베이스(220)에 제공하고, 대쉬보드(230)를 통해서 관리자가 인지할 수 있는 형태로 제공할 수 있다. 데이터베이스(220)는 모니터링 모듈(210)에 의해 수집되는 시계열 네트워크 트래픽 데이터를 저장하여, 데이터 분석부(300)가 이용할 수 있게 해준다.

[0039] 데이터 분석부(300)는 데이터 수집부(200)가 수집한 원시 데이터를 실시간으로 받아들이고, 원시 데이터로부터 공격 또는 침입의 징후를 탐지하며, 공격 또는 침입의 징후가 탐지된 경우 경보를 발생하여 MANO(50)에 제공할 수 있다. 구체적으로, 데이터 분석부(300)는 데이터 수집부(200)를 통해 수집된 원시 데이터를 가공하여 학습용 데이터셋을 생성할 수 있고, 생성된 데이터셋을 사용하여 공격/침입 탐지 모델을 지도학습시킬 수 있다. 공격/침입 탐지 모델의 학습이 된 상태에서, 데이터 분석부(300)는 원시 데이터로부터 공격 또는 침입의 징후를 추론에 의해 탐지하고, 공격 또는 침입의 징후가 탐지된 경우 네트워크 관리자에게 통보하거나 경보를 발생할 수 있다.

[0040] 데이터 분석부(300)는 전처리기(310), 이상상태 탐지 모델(320), 및 경보생성 및 관리 모듈(330)를 포함할 수 있다. 전처리기(310)는 데이터 수집부(200)를 통해 수신한 원시 데이터를 스케일링하는 과정과, 트래픽을 정상 상태 및 비정상 상태로 구분하는 레이블링(labeling) 과정에 의해 전처리한다. 이상상태 탐지 모델(320)은 지도학습에 의해 학습될 수 있으며, 학습된 상태에서 원시 데이터로부터 공격 또는 침입의 징후를 추론에 의해 탐지할 수 있다. 일 실시예에 있어서, 이상상태 탐지 모델(320)은 시계열 데이터의 특성을 반영하여 분(classification) 정확도를 높일 수 있도록 RNN(Recurrent Neural Network) 구조를 토대로 구성될 수 있다. 경보생성 및 관리 모듈(330)은 공격 또는 침입의 징후가 탐지된 경우 경보를 발생할 수 있고, 특히 MANO(50)에 경보 정보를 제공하여 MANO(50)가 보안을 위해 필요한 조치를 취할 수 있도록 한다.

[0041] 도 2는 도 1에 도시된 이상상태 탐지 모델(320)의 일 실시예의 블록도이다.

[0042] 예시적인 실시예에 따르면, 이상상태 탐지 모델(320)은 네트워크 트래픽 데이터의 시계열 특성을 보다 잘 파악하기 위해 한 시점의 데이터 값만을 기준으로 학습 및 분류를 하는 것이 아니라 특정 기간의 입력 데이터 시퀀스(sequence)를 사용하는 RNN 구조를 포함할 수 있다. 네트워크 공격 및 침입 탐지하는 분류 문제는 출력 $\hat{y}_t$ 를 생성하는 매개변수 함수 $\hat{y}_t = f(X_t; \theta)$ 로 볼 수 있다. 여기서 출력 $\hat{y}_t$ 는 이진 분류의 경우 0과 1 중 어느

하나의 값을 가지고, 다중 클래스 분류의 경우 숫자 값을 가질 수 있다. 여기서 t와 Θ는 시간 인덱스와 모델의 매개변수를 각각 나타낸다. 또한 $X_t \in \mathrm{R}^{\mathrm{D}_{\mathrm{input}}}$ 는 입력 데이터 시퀀스이고, D_{input}은 네트워크 트래픽 데이터셋을 구성하는 속성의 개수이다. 이상상태 탐지 모델(320)은 네트워크 트래픽 데이터의 시계열 특성을 보다 잘 파악할 수 있도록 RNN 구조를 기반으로 하기 때문에 현재 시간 인덱스의 입력과 일정 개수(예컨대, l 개)의 이전 시간의 입력을 고려한다. 따라서 입력은 $X_{t-l+1}^{t} = [X_{t-l+1}, X_{t-l+2}, \cdots, X_t]$ 이다. 여기서, l 은 입력 데이터 시퀀스의 길이 즉, 특정 기간 동안 데이터 인스턴스의 수를 나타낸다.

[0043] 위와 같은 문제 정의를 기반으로 하는 이상상태 탐지 모델(320)은 도 2에 도시된 바와 같이 4개 계층 즉, 특징 맵핑(Feature mapping) 계층(400), 인코더 계층(410), 판독(Readout) 계층(420), 및 분류기(Classifier) 계층(430)을 포함하여 구성된다. 이중 특징 맵핑 계층(400), 인코더 계층(410), 판독 계층(420)의 3개 계층은 가변 길이 l 을 갖는 네트워크 트래픽 데이터 시퀀스를 통해 받아들임으로써 시계열 데이터의 연속적인(sequential) 특성을 반영하게 된다.

[0044] 먼저, 특징 맵핑 계층(400)은 활성화(activation) 함수가 없는 완전 연결 계층(fully connected layer)이다. 이 계층은 입력 데이터 시퀀스 X_t 를 $\tilde{X}_t \in R^{D_{input}}$ 에 매핑한다. 이에 따라, 수집된 네트워크 트래픽 데이터는 이상상태 탐지 모델(320)이 활용할 수 있도록 벡터화된다. 특징 맵핑 계층(400)의 프로세스는 수학식 1과 같이 정의될 수 있다.

수학식 1

[0045] $$\tilde{X}_t = feature\_mapping(X_t; \theta_{map})$$

[0046] 인코더 계층(410)은 특징 벡터를 구성하는 특징의 개수를 축소시켜서 특정 개수의 특징을 가지는 벡터로 인코딩하는 계층이다. 일 실시예에 있어서, 인코더 계층(410)에 자연언어처리(NLP) 및 컴퓨터 비전 분야에서 연속적인 패턴의 모델링에 최근 우수한 성능을 보이고 있는 트랜스포머(Transformer) 모델을 사용할 수 있다. 한편, 변형된 실시예에서는, 인코더 계층(410)이 RNN이나 Bi-RNN 모델을 이용하여 구현될 수 있다. 인코더 계층(410)은 트랜스포머 모델을 사용하여 입력 시퀀스 $\tilde{X}_t$ 를 전달받고 이를 처리하여 Z_t 를 출력한다. 인코더 계층(410)의 프로세스는 수학식 2로 표현될 수 있다.

수학식 2

[0047] $$Z_t = encoder(\tilde{X}_t; \theta_{encoder})$$

[0048] 판독 계층(420)은 인코더 계층(410)에서 전달받은 벡터 Z_t 를 하나의 벡터 $\tilde{z}_t$ 로 더욱 축약한다. 이 과정에서 사용되는 함수는 max, mean, 또는 self-attention이다. Max 및 mean 함수의 경우 최댓값 또는 평균값을 계산하는 함수를 통해 다수의 입력 벡터를 결합하는 풀링(pooling) 기법이다. 그리고 self-attention은 매개변수가 있는 시퀀스를 기반으로 가중치(weight)를 계산할 수 있는 가중치 합(weighted sum)을 통해 벡터 집합을 요약 및 압축하여 하나의 벡터로 나타낸다. 이 과정을 통해 특정 시점의 트래픽 데이터를 정상 트래픽 또는 비정상 트래픽(즉, 네트워크 공격 및 침입 트래픽)으로 분류하기 위한 입력 시퀀스 데이터 $X_{t-l+1}^{t} = [X_{t-l+1}, X_{t-l+2}, \cdots, X_t]$ 는 하나의 벡터 $\tilde{z}_t$ 로 축약되며, Readout 계층의 프로세스는 수학식 3으로 표현될 수 있다.

수학식 3

[0049] $$\tilde{z}_t = readout(Z_t; \theta_{attention})$$

[0050] 분류기 계층(440)은 RNN과 DNN 모델로 구성될 수 있다. 이 계층에서는 시간 단계(t)에 따른 입력 데이터 시퀀스를 축약한 벡터 $\tilde{z}_t$ 를 통해 트래픽의 정상 및 비정상을 판단하여 분류 결과 $\hat{y}_t$를 출력한다. 먼저 RNN 모델은 입력 시퀀스 $[\tilde{z}_{t-l+1}, \tilde{z}_{t-l+2}, \cdots, \tilde{z}_t]$ 를 받아들이고 은닉 상태(hidden state) $[h_{t-l+1}, h_{t-l+2}, \cdots, h_t]$ 를 출력한다. 그 후 DNN 모델은 RNN의 은닉 상태 h_t 를 입력으로 받아들이고, 탐지 결과 $\hat{y}_t$를 출력한다. 분류기(440) 계층의 프로세스는 수학식 4와 수학식 5로 표현될 수 있으며, 이때 DNN 모델에서 분류를 위한 목적함수(J)는 수학식 6과 같이 정의될 수 있다.

수학식 4

[0051] $$h_{t-l+1}^{t} = RNN(\tilde{z}_{t-l+1}^{t}; \theta_{RNN})$$

수학식 5

[0052] $$\hat{y}_t = DNN(h_t; \theta_{DNN})$$

수학식 6

[0053] $$J(y_t, \hat{y}_t) = -y_t \log(\hat{y}_t) - (1 - y_t)\log(1 - \hat{y}_t)$$

[0054] 여기서, 훈련 가능한 파라미터는 Θ_{map}, $\Theta_{encoder}$, $\Theta_{attention}$, Θ_{RNN}, Θ_{DNN}을 포함한다. 이 파라미터들은 교차 엔트로피인 목적함수 J를 최소화할 수 있도록 경사하강법을 사용하여 최적화될 수 있다.

[0055] 또한, 예시적 실시예에서 사용하는 모델은 분류 성능의 향상을 위해 이전 시간 단계의 분류 결과를 사용할 수 있다. 네트워크 공격 및 침입과 같은 비정상적 사건은 이전 시간 단계의 이벤트와 높은 상관관계가 있을 수 있으므로, 선행 시간 단계의 분류 결과를 함께 활용하면 현재 시점의 분류 정확도 향상에 도움이 될 수 있다. 이를 위해 현재 입력 X_t 을 위시한 최근의 l 개의 입력에 직전 시간 단계의 분류 결과 $\hat{y}_{t-1}$를 위시한 최근의 l 개의 분류 결과를 직합에 의해 결합할 수 있다. 일 실시예에 있어서, 직합 연산은 각 시간 단계의 입력에 대하여 이루어질 수 있다. 예컨대 입력 X_t 는 수학식 7과 같이 수정될 수 있다. 여기서 $\oplus$는 직합 (direct sum)을 의미한다. 변형된 실시예에서는, 직합 연산이 특징 추출 및 매칭 결과에 대해서 이루어질 수도 있다.

수학식 7

[0056] $$X_t = \hat{y}_{t-1} \oplus X_t$$

[0057] 도 3은 도 2의 이상상태 탐지 모델(320)의 학습 과정의 일 예를 보여주는 흐름도이다. 이상상태 탐지 모델(320)의 학습 과정은 학습용 데이터셋을 생성하는 과정(제500단계 내지 제522단계)와, 이상상태 탐지 모델(320)을 학습시키는 단계(제530단계)와, 이상상태 탐지 모델(320)의 성능을 확인하고 수정하는 단계(제540단계 내지 제550단계)를 포함할 수 있다.

[0058] 먼저 제500단계에서는, 이상상태 탐지 모델(320)을 학습시키는데 사용할 학습용 데이터셋을 생성하기 위하여 SDN/NFV 환경에서 오가는 네트워크 트래픽 데이터를 획득한다(제500단계). 네트워크 트래픽 데이터의 획득은 모니터링 에이전트와 모니터링 모듈(210)에 의해 수행될 수 있다. 모니터링 에이전트는 물리 네트워크(10)의 네트워크 요소의 동작 상태를 주기적/비주기적으로 수집할 수 있다. 또한 가상 네트워크(20)에서 동작하는 각 가상머신의 자원 사용 상태를 주기적으로 수집할 수 있다. 일 실시예에서, 모니터링 에이전트에 의해 수집되는 네트워크 데이터의 속성 내지 모니터링 메트릭은 네트워크 트래픽의 5 튜플 정보와 트래픽 양(bandwidth) 이외

에, 패킷/플로우별 길이(Size), 시간(duration) 등 83개 세부항목을 포함할 수 있다. 도 4a 및 4b는 이상상태 탐지 모델(320)이 사용하는 데이터 속성들의 일부 예들을 보여준다. 모니터링 에이전트는 데이터를 모니터링 모듈(210)로 전달하고, 모니터링 모듈은 수집된 데이터를 시계열 데이터 데이터베이스(220)에 저장할 수 있다.

[0059] 이어서, 획득된 네트워크 트래픽 데이터에 결함(fault)을 주입(injection)시킬 수 있다(제510단계). 결함 주입(fault injection)은 실제 운용 환경에서 정상 트래픽 대비 매우 드물게 일어나는 네트워크 공격 및 침입 트래픽의 발생 빈도를 제어하기 위해 사용될 수 있다. 결함이 주입된 데이터는 오프라인 상태에서의 학습을 위한 데이터셋으로 사용될 수도 있고, 실제로 네트워크에 인가되어 네트워크 공격 및 침입 상태를 발생시킴으로써 이상상태 탐지 모델(320)을 학습시키는데 사용될 수도 있다. 결함 주입된 데이터를 네트워크 공격 및 침입 상태를 발생시키는 데에는 네트워크 공격툴(tool)이 필요하다. 이와 같은 공격 툴을 통해 Scan 공격 등과 같은 간단한 공격에서 DoS/DDoS와 같은 서비스 거부 공격 등 다양한 최신 네트워크 공격에 대한 트래픽을 생성하고, 이를 통해 실제 네트워크 공격 및 침입 상황을 모사할 수 있다.

[0060] 그 다음, 결함이 주입된 데이터 또는 결함이 주입된 상태로 실제 네트워크에 인가된 후 감지된 데이터에 대하여 전처리 과정을 수행하여, 학습을 위한 데이터셋 형태로 변환할 수 있다(제520단계, 제522단계)된다. 전처리 단계(제520단계)는 스케일링(scaling) 단계와 레이블링(labeling) 단계를 포함할 수 있다. 스케일링 단계는 모니터링을 통해 수집된 메트릭 값들 중, 데이터의 측정 단위(scale)가 다를 경우 큰 단위의 값을 갖는 특성이 작은 단위의 값을 갖는 특성과 중첩되어 머신러닝 알고리즘의 학습에 영향을 미치는 것을 피하기 위해 특성들의 값을 일정한 수준으로 맞추는 것이다. 이때 스케일링 과정은 표준화(standardization)와 정규화(normalization) 과정을 포함할 수 있다. 표준화는 특성들이 가지는 값에 대한 정규 분포를 평균이 0이고 분산이 1인 표준정규분포 값으로 변환하는 것을 의미한다. 그리고 정규화는 특성들이 가지는 값을 0과 1 사이의 값으로 변환하는 것을 의미한다. 데이터 레이블링 단계(제522단계)는 데이터를 지도학습 기반의 머신러닝 알고리즘에 사용할 수 있도록 하기 위해 각 시점의 트래픽 데이터를 정상 및 비정상(즉, 공격 및 침입 상태)로 분류하는 단계이다. 비정상 트래픽은 결함 주입으로 네트워크 공격 및 트래픽을 발생시킨 시점을 기준으로 정의하고, 나머지를 정상으로 레이블링하여 데이터셋을 생성한다.

[0061] 제530단계에서는, 학습용 데이터셋을 사용하여 이상상태 탐지 모델(320)을 학습시킨다. 이상상태 탐지 모델(320)의 학습은 위와 같이 생성된 레이블링된 데이터셋을 사용하여 지도학습으로 이루어질 수 있다. 이때, 학습용 데이터셋 중 일부를 사용하여 이상상태 탐지 모델(320)의 성능 즉 분류 정확도를 평가하면서(제540단계), 성능평가 결과를 피드백시켜 이상상태 탐지 모델(320)의 매개변수를 수정할 수 있다(제550단계).

[0062] 도 5는 도 1의 네트워크 공격 및 침입 탐지 시스템에서의 탐지 동작의 일 예를 보여주는 흐름도이다.

[0063] 먼저, 물리 네트워크(10)와 가상 네트워크(20)에 설치된 모니터링 에이전트는 해당 네트워크의 트래픽을 모니터링하고, 네트워크 트래픽 데이터를 수집하여, 데이터 수집부(200)의 모니터링 모듈(210)에 제공한다(제600단계).

[0064] 트래픽 모니터링은 모니터링 에이전트와, 모니터링 모듈(210), 대시보드 (dashboard)/REST AP에 의해 이루어질 수 있다. 모니터링 에이전트는 물리 네트워크(10)의 네트워크 요소의 동작 상태를 주기적/비주기적으로 수집할 수 있다. 또한 가상 네트워크(20)에서 동작하는 각 가상머신의 자원 사용 상태를 주기적으로 수집할 수 있다. 앞에서 언급한 바와 같이, 일 실시예에서, 모니터링 에이전트에 의해 수집되는 네트워크 데이터의 속성 내지 모니터링 메트릭은 네트워크 트래픽의 5 튜플 정보와 트래픽 양(bandwidth) 이외에, 패킷/플로우별 길이(Size), 크기(size), 시간(duration) 등 세부항목을 포함하여 83개 항목을 포함할 수 있다. 모니터링 에이전트는 데이터를 모니터링 모듈(210)로 전달하고, 모니터링 모듈은 수집된 데이터를 시계열 데이터 데이터베이스(220)에 저장할 수 있다. 또한, 모니터링 상황은 대시보드를 통해 그래프, 표 등과 같이 사용자가 원하는 시각화 형태로 디스플레이될 수 있다.

[0065] 이어서, 모니터링 모듈(210)에 의해 획득된 데이터에 대하여 전처리 과정이 수행될 수 있다(제610단계). 추론을 위한 전처리(제610단계)는 스케일링(scaling) 단계를 포함할 수 있다. 스케일링 단계는 모니터링을 통해 수집된 메트릭 값들 중, 데이터의 측정 단위(scale)가 다를 경우 큰 단위의 값을 갖는 특성이 작은 단위의 값을 갖는 특성과 중첩되어 머신러닝 알고리즘의 학습에 영향을 미치는 것을 피하기 위해 특성들의 값을 일정한 수준으로 맞추는 것이다. 이때 스케일링 과정은 표준화(standardization)와 정규화(normalization) 과정을 포함할 수 있다. 표준화는 특성들이 가지는 값에 대한 정규 분포를 평균이 0이고 분산이 1인 표준정규분포 값으로 변환하는 것을 의미한다. 그리고 정규화는 특성들이 가지는 값을 0과 1 사이의 값으로 변환하는 것을 의미한다. 한편, 변형된 실시예에서는, 이와 같은 전처리 과정이 추론에 의한 이상 상황 탐지 동작의 수행과정에서 생략될

수도 있다.

[0066] 제620단계에서는, 모니터링 모듈(210)에 의해 획득된 네트워크 데이터 또는 전처리된 데이터를 이상상태 탐지 모델(320)에 인가하여, 이상상태 탐지 모델(320)이 추론에 의해 이상상태 발생 여부를 결정하도록 한다. 이상상태 탐지 모델(320)은 전처리된 데이터를 입력으로 사용하여 탐지 결과를 생성한다. 탐지 결과에 따라 트래픽 데이터는 정상 또는 네트워크 공격/침입으로 분류된다. 이상상태 즉, 네트워크 공격 또는 침입의 징후가 탐지된 경우, 경보생성 및 관리 모듈(330)은 경보를 발생할 수 있고, 특히 MANO(50)에 경보 정보를 제공하여 MANO(50)가 보안을 위해 필요한 조치를 취할 수 있도록 한다(제630단계).

[0067] 도 6은 예시적 실시예에 따른 네트워크 공격/침입 탐지 장치(100)의 물리적 구성을 보여주는 블록도이다.

[0068] 네트워크 공격/침입 탐지 장치(100)는 프로세서(400), 메모리(402), 저장 장치(404), 및 통신 인터페이스(406)를 포함할 수 있다. 또한, 네트워크 공격/침입 탐지 장치(100)는 입력 인터페이스 장치(410) 및 출력 인터페이스 장치(412)를 더 포함할 수 있다. 네트워크 공격/침입 탐지 장치(100)에 포함된 각각의 구성 요소들은 버스에 의해 연결되어 서로 통신할 수 있다.

[0069] 프로세서(400)는 메모리(402) 및/또는 저장 장치(404)에 저장된 프로그램 명령을 실행할 수 있다. 프로세서(400)는 적어도 하나의 중앙 처리 장치(central processing unit, CPU)나 그래픽 처리 장치(graphics processing unit, GPU)에 의해 구현될 수 있으며, 그밖에 본 발명에 따른 방법을 수행할 수 있는 여타의 프로세싱 디바이스일 수 있다. 프로세서(400)는 본 발명에 의한 전파채널 시뮬레이션 방법을 구현하기 위한 프로그램 명령들을 실행할 수 있다.

[0070] 메모리(402)는 예컨대 RAM(Random Access Memory)와 같은 휘발성 메모리와, ROM(Read Only Memory)과 같은 비휘발성 메모리를 포함할 수 있다. 메모리(402)는 저장 장치(404)에 저장된 프로그램 명령을 로드하여, 프로세서(400)에 제공함으로써 프로세서(400)가 이를 실행할 수 있도록 할 수 있다. 메모리(402)는 프로그램 명령 이외에, 네트워크 공격 및 침입 탐지 방법을 구현하기 위한 프로그램 수행 과정에서 발생하는 데이터를 임시 저장할 수 있다.

[0071] 저장 장치(404)는 프로그램 명령과 데이터를 저장하기에 적합한 기록매체로서, 예컨대 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), 플래시 메모리나 EPROM(Erasable Programmable ROM) 또는 이들을 기반으로 제작되는 SSD와 같은 반도체 메모리를 포함할 수 있다. 저장 장치(404)는 본 발명에 의한 네트워크 공격 및 침입 탐지 방법을 구현하기 위한 프로그램과 데이터베이스(220)를 저장할 수 있다.

[0072] 통신 인터페이스(406)는 네트워크 어댑터, WLAN 인터페이스, PLC 모듈, 4G LTE나 5G NR 인터페이스 또는 이와 유사한 통신 인터페이스 중 하나 이상을 포함할 수 있으며, 네트워크 공격/침입 탐지 장치(100)가 모니터링 에이전트 및/또는 외부 장치와 통신할 수 있게 해준다. 입력 인터페이스 장치(410)는 관리자가 조작이나 명령을 입력할 수 있게 해주고, 출력 인터페이스 장치(412)는 장치의 동작 상태 및 동작 결과를 디스플레이할 수 있다.

[0073] 상기 네트워크 공격 및 침입 탐지 모델의 성능을 레이블링된 데이터를 75%, 25%의 학습 데이터셋 (training dataset)과 테스트 데이터셋 (test dataset)으로 나누고, 학습 데이터셋을 통해 학습된 모델의 성능을 5겹 교차검증(5-fold cross validation) 방법으로 평가하였다. 모델의 평가를 위한 항목으로는 정확도(accuracy), 정밀도(precision), 재현율(recall), F-Measure(F1 score) 등을 사용하였다. 그 후, 모델 학습에 관여하지 않은 테스트 데이터셋을 통해 최종적으로 모델의 성능을 평가하였다.

[0074] 기존 연구들과의 성능 비교를 위한 공개 데이터셋을 통한 성능 검증 결과, 기존 연구들은 80~90%의 분류 정확도를 보이고 그 대상이 특정 데이터셋으로 제한되어 있지만 본 발명에서 제시하는 모델은 NSL-KDD, UNSW-NB15, CICIDS 2017 등 서로 다른 3개의 공개 데이터셋에서 모두 93% 이상의 높은 분류 정확도 (F1 Score)를 보였다. 따라서, 예시적 실시예에 따른 공격 및 침입 탐지 모델은 오탐지를 막는데 보다 적합하다고 할 수 있다.

[0075] 위에서 언급한 바와 같이 본 발명의 실시예에 따른 장치와 방법은 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 프로그램 또는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의해 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산 방식으로 컴퓨터로 읽을 수 있는 프로그램 또는 코드가 저장되고 실행될 수 있다.

[0076] 상기 컴퓨터가 읽을 수 있는 기록매체는 롬(rom), 램(ram), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함할 수 있다. 프로그램 명령은 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다.

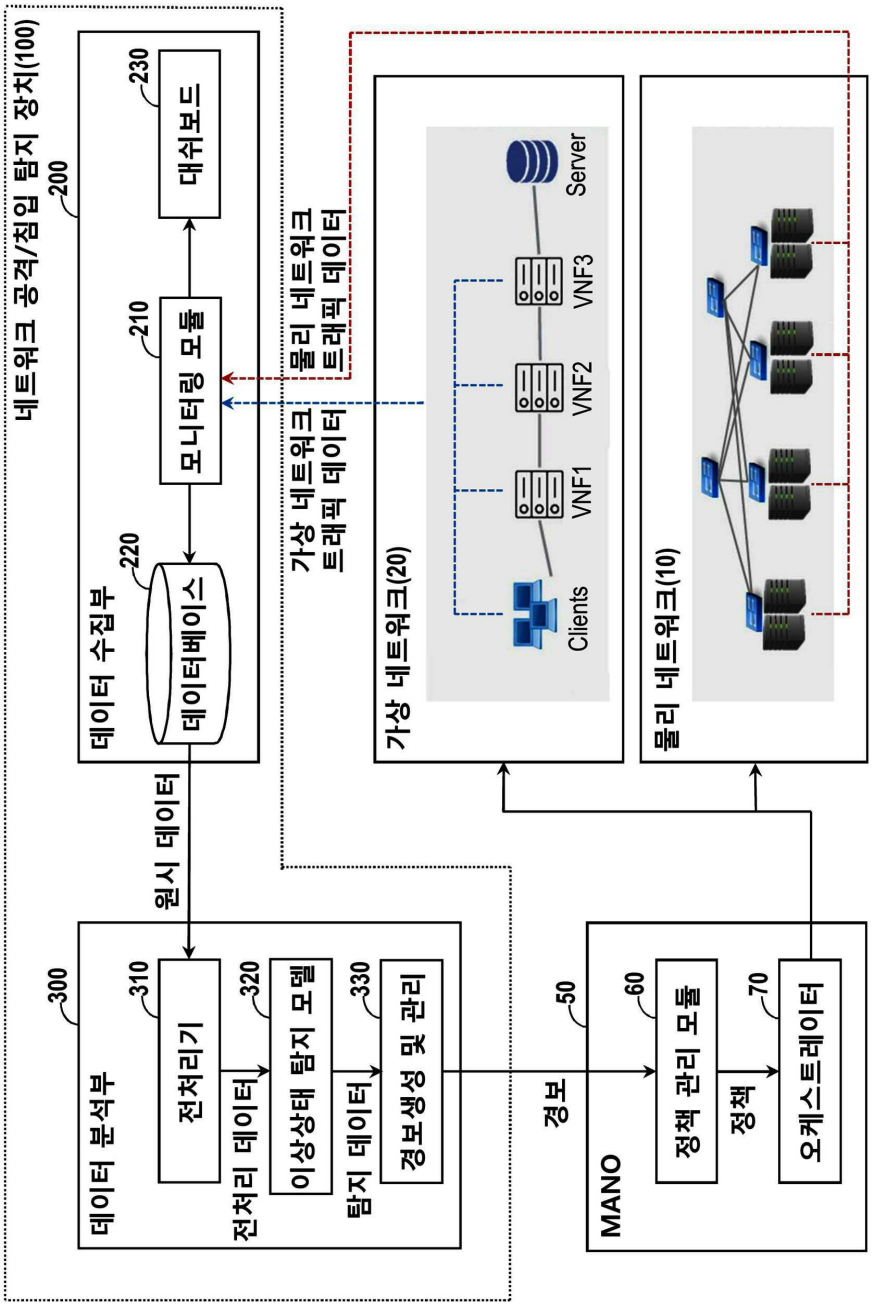
[0077] 본 발명의 일부 측면들은 장치의 문맥에서 설명되었으나, 그것은 상응하는 방법에 따른 설명 또한 나타낼 수 있고, 여기서 블록 또는 장치는 방법 단계 또는 방법 단계의 특징에 상응한다. 유사하게, 방법의 문맥에서 설명된 측면들은 또한 상응하는 블록 또는 아이템 또는 상응하는 장치의 특징으로 나타낼 수 있다. 방법 단계들의 몇몇 또는 전부는 예를 들어, 마이크로프로세서, 프로그램 가능한 컴퓨터 또는 전자 회로와 같은 하드웨어 장치에 의해(또는 이용하여) 수행될 수 있다. 몇몇의 실시예에서, 가장 중요한 방법 단계들의 하나 이상은 이와 같은 장치에 의해 수행될 수 있다.

[0078] 실시예들에서, 프로그램 가능한 로직 장치(예를 들어, 필드 프로그래머블 게이트 어레이)가 여기서 설명된 방법들의 기능의 일부 또는 전부를 수행하기 위해 사용될 수 있다. 실시예들에서, 필드 프로그래머블 게이트 어레이는 여기서 설명된 방법들 중 하나를 수행하기 위한 마이크로프로세서와 함께 작동할 수 있다. 일반적으로, 방법들은 어떤 하드웨어 장치에 의해 수행되는 것이 바람직하다.

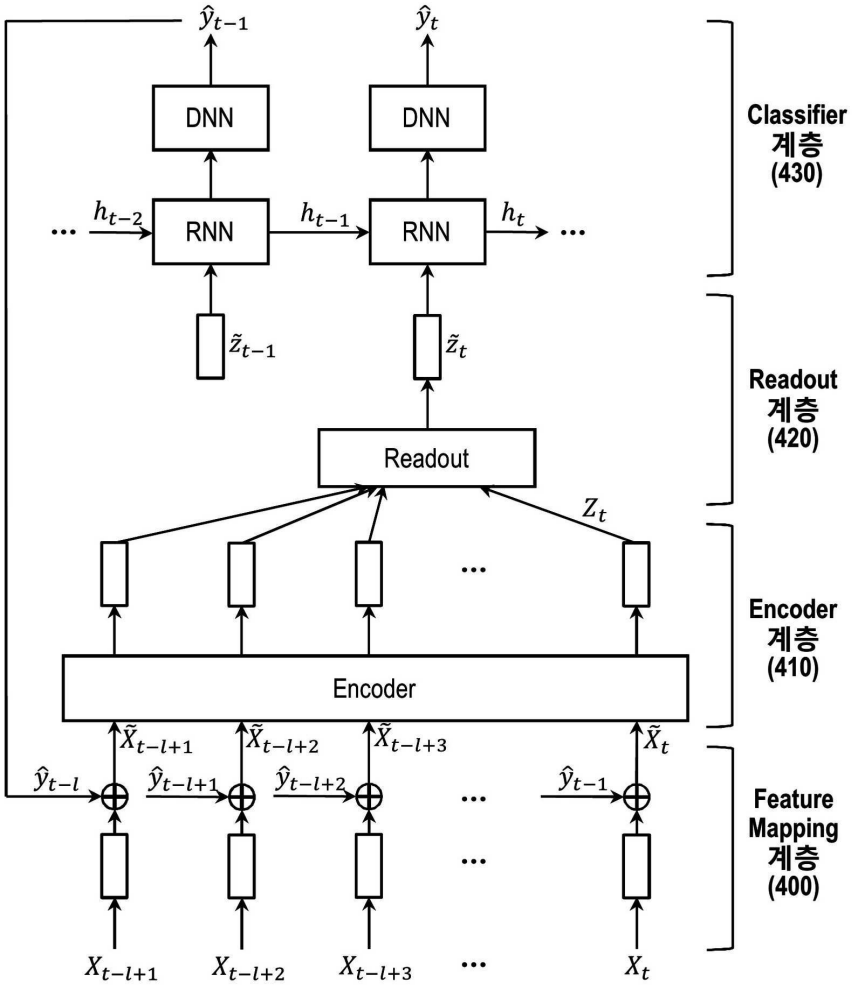
[0079] 위에서는 본 발명의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

도면

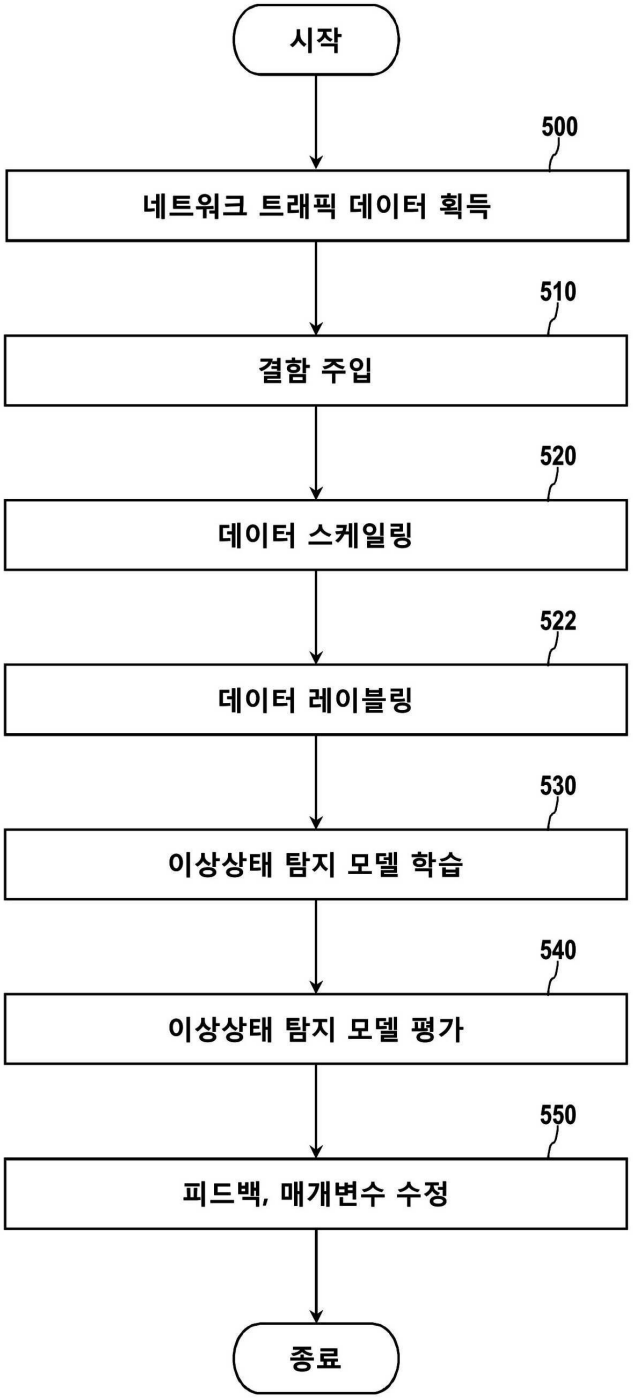
도면1


네트워크 공격/침입 탐지 장치(100)
200
데이터 수집부
220
데이터베이스
210
모니터링 모듈
230
대쉬보드
원시 데이터
가상 네트워크 트래픽 데이터
물리 네트워크 트래픽 데이터
가상 네트워크(20)
Server
VNF3
VNF2
VNF1
Clients
물리 네트워크(10)
300
데이터 분석부
310
전처리기
320
이상상태 탐지 모델
330
경보생성 및 관리
전처리 데이터
탐지 데이터
50
경보
MANO
60
정책 관리 모듈
70
오케스트레이터
정책

도면2


$\hat{y}_{t-1}$
$\hat{y}_t$
DNN
DNN
Classifier 계층 (430)
h_{t-2}
RNN
h_{t-1}
RNN
h_t
$\tilde{z}_{t-1}$
$\tilde{z}_t$
Readout
Readout 계층 (420)
Z_t
Encoder
Encoder 계층 (410)
$\tilde{X}_{t-l+1}$
$\tilde{X}_{t-l+2}$
$\tilde{X}_{t-l+3}$
$\tilde{X}_t$
$\hat{y}_{t-l}$
$\hat{y}_{t-l+1}$
$\hat{y}_{t-l+2}$
$\hat{y}_{t-1}$
Feature Mapping 계층 (400)
X_{t-l+1}
X_{t-l+2}
X_{t-l+3}
X_t

도면3


시작
500
네트워크 트래픽 데이터 획득
510
결함 주입
520
데이터 스케일링
522
데이터 레이블링
530
이상상태 탐지 모델 학습
540
이상상태 탐지 모델 평가
550
피드백, 매개변수 수정
종료

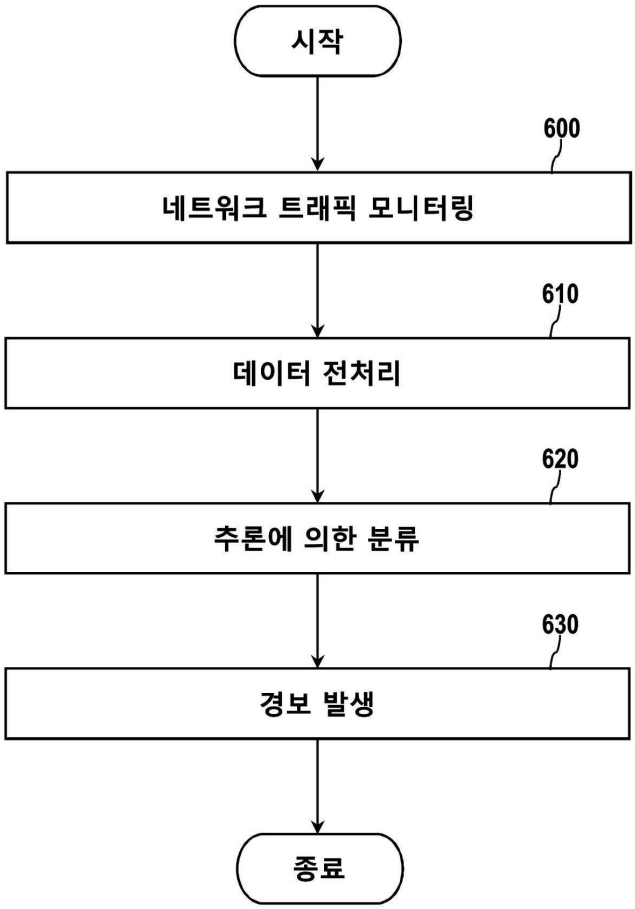
도면4a

특성	설명	특성	설명
Src IP	Source IP 번호	Dst IP	Destination IP 번호
Src port	Source 포트 번호	Dst port	Destination 포트 번호
Protocol	프로토콜 종류	Average Packet Size	패킷의 평균 크기
Flow duration	플로우 지속 시간 (ms)	Packet Size Mean	전체 패킷 크기의 평균 값 (mean)
total Fwd Packet	순방향의 총 패킷 수	Packet Size Variance	전체 패킷 크기의 분산 값
total Bwd packets	역방향의 총 패킷 수	Packet Size Std	전체 패킷 크기의 표준편차 값
total Size of Fwd Packet	순방향 패킷 크기의 총량	FIN Flag Count	FIN 플래그 수
total Size of Bwd Packet	역방향 패킷 크기의 총량	SYN Flag Count	SYN 플래그 수
Fwd Packet size Min	순방향 패킷 크기의 최솟값	RST Flag Count	RST 플래그 수
Fwd Packet size Max	순방향 패킷 크기의 최댓값	PSH Flag Count	PUSH 플래그 수
Fwd Packet Size Mean	순방향 패킷 크기의 평균값	ACK Flag Count	ACK 플래그 수
Fwd Packet Size Std	순방향 패킷 크기의 표준편차 값	URG Flag Count	URG 플래그 수
Bwd Packet Size Min	역방향 패킷 크기의 최솟값	CWR Flag Count	CWE 플래그 수
Bwd Packet Size Max	역방향 패킷 크기의 최댓값	ECE Flag Count	ECE 플래그 수
Bwd Packet Size Mean	역방향 패킷 크기의 평균 값	Down/Up Ratio	수신/전송 비율
Bwd Packet Size Std	역방향 패킷 크기의 표준편차 값	Avg Fwd Segment Size	순방향 세그먼트 크기의 평균 값
Flow Byte/s	초당 플로우 바이트 수	AVG Bwd Segment Size	역방향 세그먼트 크기의 평균 값
Flow Packets/s	초당 플로우 패킷 수	Fwd Header Size	순방향 패킷 헤더의 크기
Flow IAT Mean	플로우에서 패킷 간 평균 시간	Fwd Avg Bytes/Bulk	순방향 Bulk 당 평균 바이트
Flow IAT Std	플로우에서 패킷 간 표준 편차 시간	Fwd AVG Packet/Bulk	순방향 Bulk 당 평균 패킷 수

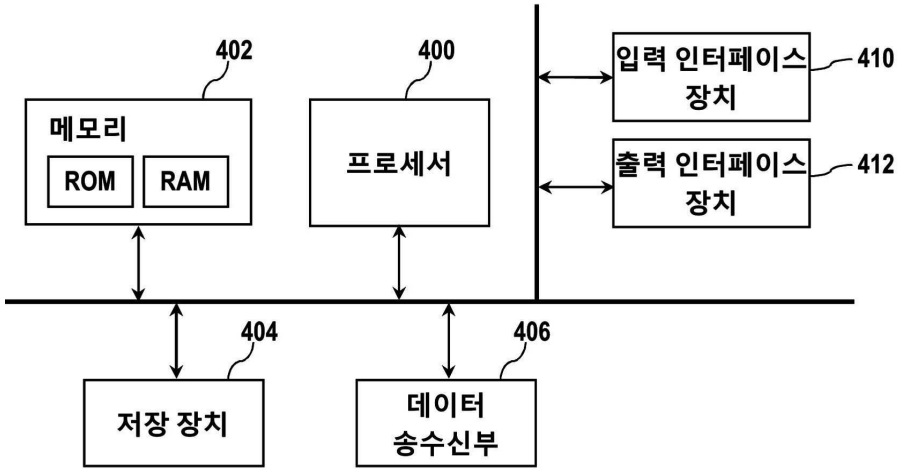
도면4b

Flow IAT Max	플로우에서 패킷 간 최대 시간	Fwd AVG Bulk Rate	순방향 Bulk 비율
Flow IAT Min	플로우에서 패킷 간 최소 시간	Bwd Avg Bytes/Bulk	역방향 Bulk 당 평균 바이트
Fwd IAT Min	순방향 패킷 간 최소 시간	Bwd AVG Packet/Bulk	역방향 Bulk 당 평균 패킷 수
Fwd IAT Max	순방향 패킷 간 최대 시간	Bwd AVG Bulk Rate	역방향 Bulk 비율
Fwd IAT Mean	순방향 패킷 간 평균 시간	Subflow Fwd Packets	순방향 서브 플로우 평균 패킷 수
Fwd IAT Std	순방향 패킷 간 표준 편차 시간	Subflow Fwd Bytes	순방향 서브 플로우 평균 바이트 수
Fwd IAT Total	순방향 패킷 간 총 시간	Subflow Bwd Packets	역방향 서브 플로우 평균 패킷 수
Bwd IAT Min	역방향 패킷 간 최소 시간	Subflow Bwd Bytes	역방향 서브 플로우 평균 바이트 수
Bwd IAT Max	역방향 패킷 간 최대 시간	Init_Win_bytes_forward	순방향 윈도우 바이트 수
Bwd IAT Mean	역방향 패킷 간 평균 시간	Init_Win_bytes_backward	역방향 윈도우 바이트 수
Bwd IAT Std	역방향 패킷 간 표준 편차 시간	Act_data_pkt_forward	페이로드가 1 바이트 이상인 패킷 수
Bwd IAT Total	역방향 패킷 간 총 시간	min_seg_size_forward	세그먼트 크기의 최솟값
Fwd PSH flag	PSH 플래그 수 (순방향 패킷)	Active Min	플로우 활성화 시간 (최솟값)
Bwd PSH Flag	PSH 플래그 수 (역방향 패킷)	Active Mean	플로우 활성화 시간 (평균 값)
Fwd URG Flag	URG 플래그 수 (순방향 패킷)	Active Max	플로우 활성화 시간 (최댓값)
Bwd URG Flag	URG 플래그 수 (역방향 패킷)	Active Std	플로우 활성화 시간 (표준편차 값)
Fwd Header Size	순방향 헤더 크기 (Byte)	Idle Min	플로우 유휴 시간 (최솟값)
Bwd Header Size	역방향 헤더 크기 (Byte)	Idle Mean	플로우 유휴 시간 (평균 값)
FWD Packets/s	초당 패킷 수 (순방향)	Idle Max	플로우 유휴 시간 (최댓값)
Bwd Packets/s	초당 패킷 수 (역방향)	Idle Std	플로우 유휴 시간 (표준편차 값)
Min Packet Size	패킷의 최소 크기	Max Packet Size	패킷의 최대
Label	레이블 값		

도면5


시작
600
네트워크 트래픽 모니터링
610
데이터 전처리
620
추론에 의한 분류
630
경보 발생
종료

도면6


402
메모리
ROM
RAM
400
프로세서
입력 인터페이스 장치
410
출력 인터페이스 장치
412
404
저장 장치
406
데이터 송수신부