

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2022-0151050
(43) 공개일자 2022년11월14일

(51) 국제특허분류(Int. Cl.)
H04L 9/40 (2022.01) H04L 47/00 (2022.01)
(52) CPC특허분류
H04L 63/1408 (2013.01)
H04L 47/2441 (2022.05)
(21) 출원번호 10-2021-0057630
(22) 출원일자 2021년05월04일
심사청구일자 2021년05월04일

(71) 출원인
엘아이지넥스원 주식회사
경기도 용인시 기흥구 마북로 207 (마북동)
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
이상근
경기도 하남시 덕풍서로 65 503동 103호
서성연
서울특별시 노원구 공릉로34길 28, 나동 106호 (공릉동, 공릉아파트)
(뒷면에 계속)
(74) 대리인
전경석

전체 청구항 수 : 총 40 항

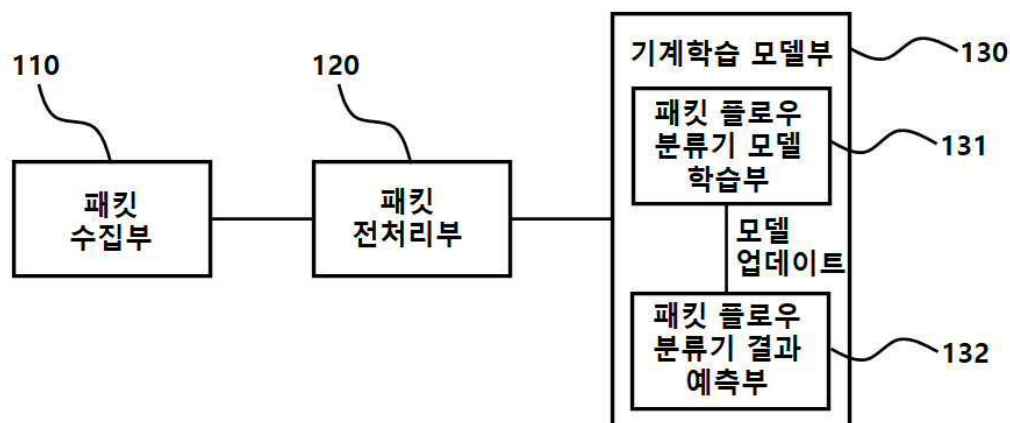
(54) 발명의 명칭 네트워크 침입 탐지 시스템 및 네트워크 침입 탐지 방법

(57) 요약

네트워크 침입 탐지 시스템 및 네트워크 침입 탐지 방법이 개시될 수 있다. 일 실시예에 따른 네트워크 침입 탐지 시스템은 네트워크 장비에 연결되어 네트워크 장비를 통해 송수신되는 패킷을 수집하는 패킷 수집부; 패킷 수집부에 연결되고, 패킷에 전처리를 수행하여 패킷 플로우를 생성하는 패킷 전처리부; 및 패킷 전처리부에 연결되고, 패킷 플로우에 임베딩을 수행하고, 임베딩된 패킷 플로우에 기초하여 네트워크 공격 여부를 탐지하는 모델을 학습하고, 학습된 모델을 사용하여 새로이 수집되는 패킷에 대한 공격 여부 또는 공격 유형을 탐지한 결과를 도출하는 기계학습 모델부를 포함할 수 있다.

대표도 - 도1

100



(52) CPC특허분류

H04L 69/22 (2022.05)

H04L 2212/002 (2013.01)

(72) 발명자

박장현

서울특별시 노원구 수락산로 174 1423동 1502호

조형민

서울특별시 은평구 갈현로16길 12, 601호

심신우

경기도 성남시 분당구 판교로 333

조병모

경기도 성남시 분당구 판교로 333

류한열

경기도 성남시 분당구 판교로 333

박성진

경기도 성남시 분당구 판교로 333

명세서

청구범위

청구항 1

네트워크 침입 탐지 시스템으로서,

네트워크 장비에 연결되어 상기 네트워크 장비를 통해 송수신되는 패킷을 수집하는 패킷 수집부;

상기 패킷 수집부에 연결되고, 상기 패킷에 전처리를 수행하여 패킷 플로우를 생성하는 패킷 전처리부; 및

상기 패킷 전처리부에 연결되고, 상기 패킷 플로우에 임베딩을 수행하고, 상기 임베딩된 패킷 플로우에 기초하여 네트워크 공격 여부를 탐지하는 모델을 학습하고, 상기 학습된 모델을 사용하여 새로이 수집되는 패킷에 대한 공격 여부 또는 공격 유형을 탐지한 결과를 도출하는 기계학습 모델부

를 포함하는 네트워크 침입 탐지 시스템.

청구항 2

제1항에 있어서, 상기 패킷 수집부는 데이터베이스를 포함하고,

상기 패킷 수집부는

상기 패킷의 맥(media access control, MAC) 주소, 아이피 주소(IP address), 및 트랜스포트 레벨 프로토콜(transport level protocol)을 확인하여 상기 패킷을 구분하고, 상기 구분된 패킷을 상기 데이터베이스에 저장하는 네트워크 침입 탐지 시스템.

청구항 3

제1항에 있어서, 상기 패킷 수집부는 상기 패킷을 순차적인 흐름으로 분석하기 위해 상기 패킷의 패킷 헤더 정보에 기초하여 상기 패킷을 분류하는 네트워크 침입 탐지 시스템.

청구항 4

제3항에 있어서, 상기 패킷 수집부는 상기 패킷 헤더 정보를 기준으로 상기 패킷의 송신자 및 수신자가 동일한 패킷의 흐름을 나타내는 플로우에 해당하는 패킷을 분류하는 네트워크 침입 탐지 시스템.

청구항 5

제4항에 있어서, 상기 패킷 수집부는 상기 패킷의 소스 아이피 주소(source IP address), 목적지 아이피 주소(destination IP address), 소스 포트 번호(source port number), 목적지 포트 번호(destination port number), 및 트랜스포트 레벨 프로토콜의 필드가 동일한 패킷을 상기 송신자와 상기 수신자가 동일한 패킷으로 간주하여 동일한 플로우에 해당하는 패킷으로 분류하는 네트워크 침입 탐지 시스템.

청구항 6

제1항에 있어서, 상기 패킷 플로우는 복수의 패킷으로 구성되고 상기 복수의 패킷 각각이 소정의 길이를 갖는 네트워크 침입 탐지 시스템.

청구항 7

제6항에 있어서, 상기 패킷 플로우는 20개의 패킷으로 구성되고, 상기 복수의 패킷 각각은 800 바이트의 길이를 갖는 네트워크 침입 탐지 시스템.

청구항 8

제7항에 있어서, 상기 패킷 플로우는 20×800의 크기의 행렬 형태로 구성되는 네트워크 침입 탐지 시스템.

청구항 9

제6항에 있어서, 상기 패킷 전처리부는 상기 패킷의 소스 아이피 주소, 목적지 아이피 주소, 소스 맥 주소 및 목적지 맥 주소를 사전 설정된 값으로 대체하는 네트워크 침입 탐지 시스템.

청구항 10

제1항에 있어서, 상기 기계학습 모델부는

상기 패킷 플로우에 기초하여 상기 패킷을 구성하는 바이트들에 대한 바이트 임베딩을 생성하는 바이트 임베딩 처리부;

상기 바이트 임베딩 처리부에 연결되고, 상기 바이트 임베딩을 종합하여 상기 패킷에 대한 패킷 임베딩을 생성하는 패킷 임베딩 처리부;

상기 패킷 임베딩 처리부에 연결되고, 상기 패킷 임베딩에 기초하여 상기 패킷 플로우에 대한 플로우 임베딩을 생성하는 플로우 임베딩 처리부; 및

상기 플로우 임베딩 처리부에 연결되고, 상기 플로우 임베딩에 기초하여 상기 패킷 플로우를 분류하는 패킷 플로우 분류 처리부

를 포함하는 네트워크 침입 탐지 시스템.

청구항 11

제10항에 있어서, 상기 바이트 임베딩 처리부는 상기 패킷을 구성하고 있는 바이트 정보를 반영한 상기 바이트 임베딩을 생성하는 네트워크 침입 탐지 시스템.

청구항 12

제11항에 있어서, 상기 바이트 임베딩 처리부는

단순 임베딩 부분 및 상기 바이트 임베딩에 위치 정보를 추가하는 위치 정보 인코딩 부분;

상기 패킷을 구성하는 상기 바이트들의 관계를 파악하고 상기 바이트들 각각의 문맥 정보를 포함하는 임베딩을 찾아내는 멀티 헤드 어텐션을 수행하는 멀티 헤드 어텐션(multi-head attention) 부분; 및

상기 멀티 헤드 어텐션을 통해 출력되는 상기 바이트 임베딩을 비선형 변환시키는 전결합 층

을 포함하는 네트워크 침입 탐지 시스템.

청구항 13

제12항에 있어서, 상기 바이트 임베딩 처리부는 상기 위치 정보 인코딩을 통해 상기 패킷의 바이트의 순서 정보를 유지시키는 네트워크 침입 탐지 시스템.

청구항 14

제12항에 있어서, 상기 바이트 임베딩 처리부는 상기 멀티 헤드 어텐션을 통해 상기 바이트 임베딩 처리부에 입력되는 입력 값 사이의 관계를 파악하는 네트워크 침입 탐지 시스템.

청구항 15

제12항에 있어서, 상기 패킷 임베딩 처리부는 애디티브 어텐션(additive attention)을 사용하여 상기 패킷 임베딩을 생성하는 네트워크 침입 탐지 시스템.

청구항 16

제15항에 있어서, 상기 애디티브 어텐션은 전결합 층 및 소프트 맥스 함수로 구성되는 네트워크 침입 탐지 시스템.

청구항 17

제10항에 있어서, 상기 플로우 임베딩 처리부는 애디티브 어텐션을 사용하여 상기 플로우 임베딩을 생성하는 네트워크 침입 탐지 시스템.

청구항 18

제10항에 있어서, 상기 플로우 임베딩 처리부는 전결합 층 및 소프트 맥스 함수를 사용하여 상기 플로우 임베딩을 생성하는 네트워크 침입 탐지 시스템.

청구항 19

제10항에 있어서, 상기 패킷 플로우 분류 처리부는 상기 플로우 임베딩을 사용하여 패킷 플로우의 공격 여부 또는 공격 유형의 탐지를 수행하는 네트워크 침입 탐지 시스템.

청구항 20

제19항에 있어서, 상기 패킷 플로우 분류 처리부는 전결합 층 및 소프트 맥스 함수에 상기 플로우 임베딩을 통과시켜 공격 여부 또는 공격 유형의 탐지를 수행하는 네트워크 침입 탐지 시스템.

청구항 21

네트워크 침입 탐지 시스템에서의 네트워크 침입 탐지 방법으로서,

상기 네트워크 침입 탐지 시스템의 패킷 수집부에서, 네트워크 장비를 통해 송수신되는 패킷을 수집하는 단계;

상기 네트워크 침입 탐지 시스템의 패킷 전처리부에서, 상기 패킷에 전처리를 수행하여 패킷 플로우를 생성하는 단계;

상기 네트워크 침입 탐지 시스템의 기계학습 모델부에서, 상기 패킷 플로우에 임베딩을 수행하여 네트워크 공격 여부를 탐지하는 모델을 학습하는 단계; 및

상기 기계학습 모델부에서, 상기 학습된 모델을 사용하여 새로이 수집되는 패킷에 대한 공격 여부 또는 공격 유형을 탐지한 결과를 도출하는 단계

를 포함하는 네트워크 침입 탐지 방법.

청구항 22

제21항에 있어서, 상기 패킷 수집부는 데이터베이스를 포함하고,

상기 패킷을 수집하는 단계는

상기 패킷의 맥 주소, 아이피 주소, 및 트랜스포트 레벨 프로토콜을 확인하여 상기 패킷을 구분하는 단계; 및

상기 구분된 패킷을 상기 데이터베이스에 저장하는 단계

를 포함하는 네트워크 침입 탐지 방법.

청구항 23

제21항에 있어서, 상기 패킷을 수집하는 단계는

상기 패킷을 순차적인 흐름으로 분석하기 위해 상기 패킷의 패킷 헤더 정보에 기초하여 상기 패킷을 분류하는 단계

를 포함하는 네트워크 침입 탐지 방법.

청구항 24

제23항에 있어서, 상기 패킷을 분류하는 단계는

상기 패킷 헤더 정보를 기준으로 상기 패킷의 송신자 및 수신자가 동일한 패킷의 흐름을 나타내는 플로우에 해당하는 패킷을 분류하는 단계

를 포함하는 네트워크 침입 탐지 방법.

청구항 25

제24항에 있어서, 상기 패킷을 분류하는 단계는

상기 패킷의 소스 아이피 주소, 목적지 아이피 주소, 소스 포트 번호, 목적지 포트 번호, 및 트랜스포트 레벨 프로토콜의 필드가 동일한 패킷을 상기 송신자와 상기 수신자가 동일한 패킷으로 간주하여 동일한 플로우에 해당하는 패킷으로 분류하는 단계

를 포함하는 네트워크 침입 탐지 방법.

청구항 26

제21항에 있어서, 상기 패킷 플로우는 복수의 패킷으로 구성되고 상기 복수의 패킷 각각이 소정의 길이를 갖는 네트워크 침입 탐지 방법.

청구항 27

제26항에 있어서, 상기 패킷 플로우는 20개의 패킷으로 구성되고, 상기 복수의 패킷 각각은 800 바이트의 길이를 갖는 네트워크 침입 탐지 방법.

청구항 28

제27항에 있어서, 상기 패킷 플로우는 20×800 의 크기의 행렬 형태로 구성되는 네트워크 침입 탐지 방법.

청구항 29

제26항에 있어서, 상기 패킷에 전처리부를 수행하여 패킷 플로우를 생성하는 단계는

상기 패킷의 소스 아이피 주소, 목적지 아이피 주소, 소스 맥 주소 및 목적지 맥 주소를 사전 설정된 값으로 대체하는 단계

를 포함하는 네트워크 침입 탐지 방법.

청구항 30

제21항에 있어서, 상기 기계학습 모델부는 바이트 임베딩 처리부, 패킷 임베딩 처리부, 플로우 임베딩 처리부 및 패킷 플로우 분류 처리부를 포함하고,

상기 패킷 플로우를 임베딩으로 변환하여 네트워크 공격 여부를 탐지하는 모델을 학습하는 단계는

상기 바이트 임베딩 처리부에서, 상기 패킷 플로우에 기초하여 상기 패킷을 구성하는 바이트들에 대한 바이트 임베딩을 생성하는 단계;

상기 패킷 임베딩 처리부에서, 상기 바이트 임베딩을 종합하여 상기 패킷에 대한 패킷 임베딩을 생성하는 단계;

상기 플로우 임베딩 처리부에서, 상기 패킷 임베딩에 기초하여 상기 패킷 플로우에 대한 플로우 임베딩을 생성하는 단계; 및

상기 패킷 플로우 분류 처리부에서, 상기 플로우 임베딩에 기초하여 상기 패킷 플로우를 분류하는 단계

를 포함하는 네트워크 침입 탐지 방법.

청구항 31

제30항에 있어서, 상기 바이트 임베딩을 생성하는 단계는

상기 패킷을 구성하고 있는 바이트 정보를 반영한 상기 바이트 임베딩을 생성하는 단계

를 포함하는 네트워크 침입 탐지 방법.

청구항 32

제31항에 있어서, 상기 바이트 임베딩 처리부는

단순 임베딩 부분 및 상기 바이트 임베딩에 위치 정보를 추가하는 위치 정보 인코딩 부분;

상기 패킷을 구성하는 상기 바이트들의 관계를 파악하고 상기 바이트들 각각의 문맥 정보를 포함하는 임베딩을

찾아내는 멀티 헤드 어텐션을 수행하는 멀티 헤드 어텐션 부분; 및
상기 멀티 헤드 어텐션을 통해 출력되는 상기 바이트 임베딩을 비선형 변환시키는 전결합 층
을 포함하는 네트워크 침입 탐지 방법.

청구항 33

제32항에 있어서, 상기 바이트 임베딩을 생성하는 단계는
상기 위치 정보 인코딩을 통해 상기 패킷의 바이트의 순서 정보를 유지시키는 단계
를 포함하는 네트워크 침입 탐지 방법.

청구항 34

제32항에 있어서, 상기 바이트 임베딩을 생성하는 단계는
상기 멀티 헤드 어텐션을 통해 상기 바이트 임베딩 처리부에 입력되는 입력 값 사이의 관계를 파악하는 단계
를 포함하는 네트워크 침입 탐지 방법.

청구항 35

제32항에 있어서, 상기 패킷 임베딩을 생성하는 단계는
애디티브 어텐션을 사용하여 상기 패킷 임베딩을 생성하는 단계
를 포함하는 네트워크 침입 탐지 방법.

청구항 36

제35항에 있어서, 상기 패킷 임베딩을 생성하는 단계는
전결합 층 및 소프트 맥스 함수를 사용하여 상기 패킷 임베딩을 생성하는 단계
를 포함하는 네트워크 침입 탐지 방법.

청구항 37

제30항에 있어서, 상기 플로우 임베딩을 생성하는 단계는
애디티브 어텐션을 사용하여 상기 플로우 임베딩을 생성하는 단계
를 포함하는 네트워크 침입 탐지 방법.

청구항 38

제30항에 있어서, 상기 플로우 임베딩을 생성하는 단계는
전결합 층 및 소프트 맥스 함수를 사용하여 상기 플로우 임베딩을 생성하는 단계
를 포함하는 네트워크 침입 탐지 방법.

청구항 39

제30항에 있어서, 상기 패킷 플로우를 분류하는 단계는
상기 플로우 임베딩을 사용하여 패킷 플로우의 공격 여부 또는 공격 유형의 탐지를 수행하는 단계
를 포함하는 네트워크 침입 탐지 방법.

청구항 40

제39항에 있어서, 상기 패킷 플로우를 분류하는 단계는
전결합 층 및 소프트 맥스 함수에 상기 플로우 임베딩을 통과시켜 공격 여부 또는 공격 유형의 탐지를 수행하는

단계

를 포함하는 네트워크 침입 탐지 방법.

발명의 설명

기술 분야

[0001] 본 발명은 네트워크 침입 탐지 시스템 및 네트워크 침입 탐지 방법에 관한 것이다.

배경 기술

[0002] 네트워크 침입 탐지 시스템(intrusion detection system, IDS)은 라우터를 통해 수입한 패킷들을 분석하여 네트워크 공격 여부를 판단한다. 최근, 기계학습이 많은 분야에서 각광을 받음에 따라 기계학습이 네트워크 침입 탐지에 적용되고 있다.

[0003] 네트워크 침입 탐지에서는 패킷을 통해 공격의 패턴을 파악하는 것이 중요하며, 기계학습 기반 네트워크 침입 탐지에서 가장 중요한 부분은 기계학습 모델이 패킷을 잘 분류할 수 있는 표현형을 찾는 것이다. 일반적으로, 기계학습에서 패킷을 잘 분류할 수 있는 표현형을 찾는 것을 임베딩이라 한다. 종래에는 순환 신경망, 단어 유사도 기반 임베딩 모델(word2vec), 차원 축소 기법 등을 통해 임베딩을 찾았다.

[0004] 예를 들면, 종래(특허문헌 1)에는 라우터와 클라이언트 컴퓨터 간에 송수신되는 패킷을 수집하고, 수집된 패킷을 세션 단위로 분리한 후 통계적 특징(예를 들어, 단위시간 당 패킷 수, 단위시간 당 패킷의 크기 변화량 등)을 추출하였다. 또한, 종래(특허문헌 1)에는 추출된 통계적 특징을 기계학습 모델(예를 들어, 격리 포레스트, 원클래스 서포트 벡터 머신)의 입력으로 하여 네트워크 침입 탐지를 수행하였다.

[0005] 그러나, 종래에는 순환 신경망과 word2vec의 경우, 재귀적 구조, 학습 단어의 수(예를 들어, 네트워크 침입 탐지에서는 패킷 헤더와 페이로드가 가질 수 있는 값의 종류) 등과 같은 구조의 한계점으로 인해, 학습 시간과 예측 시간이 오래 소요되어 실시간 탐지에 부적합하다는 문제점이 있다. 또한, 패킷의 흐름에 대해 고려하고 있지 않고 패킷 자체에 대한 통계적 특성만을 고려하기 때문에 패킷의 흐름을 고려한 임베딩에 비해 정보의 손실이 발생하는 문제점이 있다. 이로 인해, 기계학습 모델이 소위 제로 데이 공격이라 불리는 처음 보는 공격에 취약하게 되는 문제점이 있다. 또한, 공격 특성 탐지에 중요한 패킷의 특정 부분을 집중해서 보는 것이 아니라 전체적으로 보고 탐지하기 때문에 중요한 정보가 고정된 크기의 임베딩에 전부 포함되지 못하는 문제점이 있다.

[0006] 한편, 종래(특허문헌 2)에는 정상/악성 플로우 데이터를 수집하고 딥 러닝을 사용하여 네트워크 공격을 탐지하고 있다. 그러나, 종래에는 패킷의 흐름을 고려할 수 있는 딥 러닝 모델 구조를 사용하지 않기 때문에 임베딩에 패킷의 흐름에 대한 정보는 포함되어 있지 않다. 또한, 종래에는 패킷의 헤더만 보고 침입 여부를 탐지하므로 페이로드까지 보는 방법에 비해 정보량이 부족하여 성능이 떨어지는 문제점이 있다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 한국공개특허 제10-2019-0048264호

(특허문헌 0002) 한국등록특허 제10-2032249호

발명의 내용

해결하려는 과제

[0008] 본 발명은 패킷 흐름에 대한 임베딩을 병렬적으로 계산하고 네트워크 침입 탐지에 중요한 특징 정보만을 담고 있는 패킷 플로우 임베딩을 구하여 네트워크 침입 탐지를 수행하는 네트워크 침입 탐지 시스템 및 네트워크 침입 탐지 방법을 제공할 수 있다.

과제의 해결 수단

- [0009] 본 발명의 일 실시예에 따르면, 네트워크 침입 탐지 시스템이 개시될 수 있다. 일 실시예에 따른 네트워크 침입 탐지 시스템은 네트워크 장비에 연결되어 상기 네트워크 장비를 통해 송수신되는 패킷을 수집하는 패킷 수집부; 상기 패킷 수집부에 연결되고, 상기 패킷에 전처리를 수행하여 패킷 플로우를 생성하는 패킷 전처리부; 및 상기 패킷 전처리부에 연결되고, 상기 패킷 플로우에 임베딩을 수행하고, 상기 임베딩된 패킷 플로우에 기초하여 네트워크 공격 여부를 탐지하는 모델을 학습하고, 상기 학습된 모델을 사용하여 새로이 수집되는 패킷에 대한 공격 여부 또는 공격 유형을 탐지한 결과를 도출하는 기계학습 모델부를 포함할 수 있다.
- [0010] 일 실시예에 있어서, 상기 패킷 수집부는 데이터베이스를 포함하고, 상기 패킷 수집부는 상기 패킷의 맥(media access control, MAC) 주소, 아이피 주소(IP address), 및 트랜스포트 레벨 프로토콜(transport level protocol)을 확인하여 상기 패킷을 구분하고, 상기 구분된 패킷을 상기 데이터베이스에 저장할 수 있다.
- [0011] 일 실시예에 있어서, 상기 패킷 수집부는 상기 패킷을 순차적인 흐름으로 분석하기 위해 상기 패킷의 패킷 헤더 정보에 기초하여 상기 패킷을 분류할 수 있다.
- [0012] 일 실시예에 있어서, 상기 패킷 수집부는 상기 패킷 헤더 정보를 기준으로 상기 패킷의 송신자 및 수신자가 동일한 패킷의 흐름을 나타내는 플로우에 해당하는 패킷을 분류할 수 있다.
- [0013] 일 실시예에 있어서, 상기 패킷 수집부는 상기 패킷의 소스 아이피 주소(source IP address), 목적지 아이피 주소(destination IP address), 소스 포트 번호(source port number), 목적지 포트 번호(destination port number), 및 트랜스포트 레벨 프로토콜의 필드가 동일한 패킷을 상기 송신자와 상기 수신자가 동일한 패킷으로 간주하여 동일한 플로우에 해당하는 패킷으로 분류할 수 있다.
- [0014] 일 실시예에 있어서, 상기 패킷 플로우는 복수의 패킷으로 구성되고 상기 복수의 패킷 각각이 소정의 길이를 가질 수 있다.
- [0015] 일 실시예에 있어서, 상기 패킷 플로우는 20개의 패킷으로 구성되고, 상기 복수의 패킷 각각은 800 바이트의 길이를 가질 수 있다.
- [0016] 일 실시예에 있어서, 상기 패킷 플로우는 20×800 의 크기의 행렬 형태로 구성될 수 있다.
- [0017] 일 실시예에 있어서, 상기 패킷 전처리부는 상기 패킷의 소스 아이피 주소, 목적지 아이피 주소, 소스 맥 주소 및 목적지 맥 주소를 사전 설정된 값으로 대체할 수 있다.
- [0018] 일 실시예에 있어서, 상기 기계학습 모델부는 상기 패킷 플로우에 기초하여 상기 패킷을 구성하는 바이트들에 대한 바이트 임베딩을 생성하는 바이트 임베딩 처리부; 상기 바이트 임베딩 처리부에 연결되고, 상기 바이트 임베딩을 종합하여 상기 패킷에 대한 패킷 임베딩을 생성하는 패킷 임베딩 처리부; 상기 패킷 임베딩 처리부에 연결되고, 상기 패킷 임베딩에 기초하여 상기 패킷 플로우에 대한 플로우 임베딩을 생성하는 플로우 임베딩 처리부; 및 상기 플로우 임베딩 처리부에 연결되고, 상기 플로우 임베딩에 기초하여 상기 패킷 플로우를 분류하는 패킷 플로우 분류 처리부를 포함할 수 있다.
- [0019] 일 실시예에 있어서, 상기 바이트 임베딩 처리부는 상기 패킷을 구성하고 있는 바이트 정보를 반영한 상기 바이트 임베딩을 생성할 수 있다.
- [0020] 일 실시예에 있어서, 상기 바이트 임베딩 처리부는 단순 임베딩 부분 및 상기 바이트 임베딩에 위치 정보를 추가하는 위치 정보 인코딩 부분; 상기 패킷을 구성하는 상기 바이트들의 관계를 파악하고 상기 바이트들 각각의 문맥 정보를 포함하는 임베딩을 찾아내는 멀티 헤드 어텐션을 수행하는 멀티 헤드 어텐션(multi-head attention) 부분; 및 상기 멀티 헤드 어텐션을 통해 출력되는 상기 바이트 임베딩을 비선형 변환시키는 전결합층을 포함할 수 있다.
- [0021] 일 실시예에 있어서, 상기 바이트 임베딩 처리부는 상기 위치 정보 인코딩을 통해 상기 패킷의 바이트의 순서 정보를 유지시킬 수 있다.
- [0022] 일 실시예에 있어서, 상기 바이트 임베딩 처리부는 상기 멀티 헤드 어텐션을 통해 상기 바이트 임베딩 처리부에 입력되는 입력 값 사이의 관계를 파악할 수 있다.
- [0023] 일 실시예에 있어서, 상기 패킷 임베딩 처리부는 애디티브 어텐션(additive attention)을 사용하여 상기 패킷 임베딩을 생성할 수 있다.
- [0024] 일 실시예에 있어서, 상기 애디티브 어텐션은 전결합 층 및 소프트 맥스 함수로 구성되어 있다.

- [0025] 일 실시예에 있어서, 상기 플로우 임베딩 처리부는 애디티브 어텐션을 사용하여 상기 플로우 임베딩을 생성할 수 있다.
- [0026] 일 실시예에 있어서, 상기 플로우 임베딩 처리부는 전결합 층 및 소프트 맥스 함수를 사용하여 상기 플로우 임베딩을 생성할 수 있다.
- [0027] 일 실시예에 있어서, 상기 패킷 플로우 분류 처리부는 상기 플로우 임베딩을 사용하여 패킷 플로우의 공격 여부 또는 공격 유형의 탐지를 수행할 수 있다.
- [0028] 일 실시예에 있어서, 상기 패킷 플로우 분류 처리부는 전결합 층 및 소프트 맥스 함수에 상기 플로우 임베딩을 통과시켜 공격 여부 또는 공격 유형의 탐지를 수행할 수 있다.
- [0029] 본 발명의 일 실시예에 따르면, 네트워크 침입 탐지 시스템에서의 네트워크 침입 탐지 방법이 개시될 수 있다. 일 실시예에 따른 네트워크 침입 탐지 방법은 상기 네트워크 침입 탐지 시스템의 패킷 수집부에서, 네트워크 장비를 통해 송수신되는 패킷을 수집하는 단계; 상기 네트워크 침입 탐지 시스템의 패킷 전처리부에서, 상기 패킷에 전처리를 수행하여 패킷 플로우를 생성하는 단계; 상기 네트워크 침입 탐지 시스템의 기계학습 모델부에서, 상기 패킷 플로우에 임베딩을 수행하여 네트워크 공격 여부를 탐지하는 모델을 학습하는 단계; 및 상기 기계학습 모델부에서, 상기 학습된 모델을 사용하여 새로이 수집되는 패킷에 대한 공격 여부 또는 공격 유형을 탐지한 결과를 도출하는 단계를 포함할 수 있다.
- [0030] 일 실시예에 있어서, 상기 패킷 수집부는 데이터베이스를 포함하고, 상기 패킷을 수집하는 단계는 상기 패킷의 맥 주소, 아이피 주소, 및 트랜스포트 레벨 프로토콜을 확인하여 상기 패킷을 구분하는 단계; 및 상기 구분된 패킷을 상기 데이터베이스에 저장하는 단계를 포함할 수 있다.
- [0031] 일 실시예에 있어서, 상기 패킷을 수집하는 단계는 상기 패킷을 순차적인 흐름으로 분석하기 위해 상기 패킷의 패킷 헤더 정보에 기초하여 상기 패킷을 분류하는 단계를 포함할 수 있다.
- [0032] 일 실시예에 있어서, 상기 패킷을 분류하는 단계는 상기 패킷 헤더 정보를 기준으로 상기 패킷의 송신자 및 수신자가 동일한 패킷의 흐름을 나타내는 플로우에 해당하는 패킷을 분류하는 단계를 포함할 수 있다.
- [0033] 일 실시예에 있어서, 상기 패킷을 분류하는 단계는 상기 패킷의 소스 아이피 주소, 목적지 아이피 주소, 소스 포트 번호, 목적지 포트 번호, 및 트랜스포트 레벨 프로토콜의 필드가 동일한 패킷을 상기 송신자와 상기 수신자가 동일한 패킷으로 간주하여 동일한 플로우에 해당하는 패킷으로 분류하는 단계를 포함할 수 있다.
- [0034] 일 실시예에 있어서, 상기 패킷 플로우는 복수의 패킷으로 구성되고 상기 복수의 패킷 각각이 소정의 길이를 가질 수 있다.
- [0035] 일 실시예에 있어서, 상기 패킷 플로우는 20개의 패킷으로 구성되고, 상기 복수의 패킷 각각은 800 바이트의 길이를 가질 수 있다.
- [0036] 일 실시예에 있어서, 상기 패킷 플로우는 20×800의 크기의 행렬 형태로 구성될 수 있다.
- [0037] 일 실시예에 있어서, 상기 패킷에 전처리부를 수행하여 패킷 플로우를 생성하는 단계는 상기 패킷의 소스 아이피 주소, 목적지 아이피 주소, 소스 맥 주소 및 목적지 맥 주소를 사전 설정된 값으로 대체하는 단계를 포함할 수 있다.
- [0038] 일 실시예에 있어서, 상기 기계학습 모델부는 바이트 임베딩 처리부, 패킷 임베딩 처리부, 플로우 임베딩 처리부 및 패킷 플로우 분류 처리부를 포함하고, 상기 패킷 플로우를 임베딩으로 변환하여 네트워크 공격 여부를 탐지하는 모델을 학습하는 단계는 상기 바이트 임베딩 처리부에서, 상기 패킷 플로우에 기초하여 상기 패킷을 구성하는 바이트들에 대한 바이트 임베딩을 생성하는 단계; 상기 패킷 임베딩 처리부에서, 상기 바이트 임베딩을 종합하여 상기 패킷에 대한 패킷 임베딩을 생성하는 단계; 상기 플로우 임베딩 처리부에서, 상기 패킷 임베딩에 기초하여 상기 패킷 플로우에 대한 플로우 임베딩을 생성하는 단계; 및 상기 패킷 플로우 분류 처리부에서, 상기 플로우 임베딩에 기초하여 상기 패킷 플로우를 분류하는 단계를 포함할 수 있다.
- [0039] 일 실시예에 있어서, 상기 바이트 임베딩을 생성하는 단계는 상기 패킷을 구성하고 있는 바이트 정보를 반영한 상기 바이트 임베딩을 생성하는 단계를 포함할 수 있다.
- [0040] 일 실시예에 있어서, 상기 바이트 임베딩 처리부는 단순 임베딩 부분 및 상기 바이트 임베딩에 위치 정보를 추가하는 위치 정보 인코딩 부분; 상기 패킷을 구성하는 상기 바이트들의 관계를 파악하고 상기 바이트들 각각의

문맥 정보를 포함하는 임베딩을 찾아내는 멀티 헤드 어텐션을 수행하는 멀티 헤드 어텐션 부분; 및 상기 멀티 헤드 어텐션을 통해 출력되는 상기 바이트 임베딩을 비선형 변환시키는 전결합 층을 포함할 수 있다.

- [0041] 일 실시예에 있어서, 상기 바이트 임베딩을 생성하는 단계는 상기 위치 정보 인코딩을 통해 상기 패킷의 바이트의 순서 정보를 유지시키는 단계를 포함할 수 있다.
- [0042] 일 실시예에 있어서, 상기 바이트 임베딩을 생성하는 단계는 상기 멀티 헤드 어텐션을 통해 상기 바이트 임베딩 처리부에 입력되는 입력 값 사이의 관계를 파악하는 단계를 포함할 수 있다.
- [0043] 일 실시예에 있어서, 상기 패킷 임베딩을 생성하는 단계는 애디티브 어텐션을 사용하여 상기 패킷 임베딩을 생성하는 단계를 포함할 수 있다.
- [0044] 일 실시예에 있어서, 상기 패킷 임베딩을 생성하는 단계는 전결합 층 및 소프트 맥스 함수를 사용하여 상기 패킷 임베딩을 생성하는 단계를 포함할 수 있다.
- [0045] 일 실시예에 있어서, 상기 플로우 임베딩을 생성하는 단계는 애디티브 어텐션을 사용하여 상기 플로우 임베딩을 생성하는 단계를 포함할 수 있다.
- [0046] 일 실시예에 있어서, 상기 플로우 임베딩을 생성하는 단계는 전결합 층 및 소프트 맥스 함수를 사용하여 상기 플로우 임베딩을 생성하는 단계를 포함할 수 있다.
- [0047] 일 실시예에 있어서, 상기 패킷 플로우를 분류하는 단계는 상기 플로우 임베딩을 사용하여 패킷 플로우의 공격 여부 또는 공격 유형의 탐지를 수행하는 단계를 포함할 수 있다.
- [0048] 일 실시예에 있어서, 상기 패킷 플로우를 분류하는 단계는 전결합 층 및 소프트 맥스 함수에 상기 플로우 임베딩을 통과시켜 공격 여부 또는 공격 유형의 탐지를 수행하는 단계를 포함할 수 있다.

발명의 효과

- [0049] 본 발명의 다양한 실시예에 따르면, 자연어 처리 분야에서 제안된 트랜스포머 모델 구조를 채용하여 패킷 흐름에 대한 임베딩을 병렬적으로 계산할 수 있다. 따라서, 개별 패킷에 대한 임베딩만을 구할 수 있는 종래 기술의 한계점을 극복하고 네트워크 침입 탐지에 중요한 특징 정보만을 담고 있는 패킷 플로우 임베딩을 구하여 네트워크 침입 탐지를 수행할 수 있다.
- [0050] 또한, 본 발명의 일 실시예에 따른 어텐션(attention) 패킷 플로우 임베딩 및 분류기는 개별 패킷 및 패킷 플로우에 대한 임베딩 학습뿐만 아니라, 어텐션 기법을 통해 종래 기술보다 유의미한 공격 여부(예를 들어, 정상 또는 악성)를 판단할 수 있다. 즉, 어텐션 기법이 적용되지 않는 종래 기술의 경우, 공격 여부의 판단과 관계없는 패킷 또는 패킷 플로우의 정보를 포함하기 때문에 부정확한 임베딩을 생성한다. 이에 반해, 본 발명의 다양한 실시예에 따르면 어텐션 기법을 통해 중요한 정보에 가중치를 두어 보다 정확한 임베딩을 생성하여 종래 기술보다 유의미한 공격 여부(예를 들어, 정상 또는 악성)를 판단할 수 있다.
- [0051] 더욱이, 본 발명의 다양한 실시예에 따르면, 기계학습 기반 네트워크 침입 탐지 시스템에 활용할 수 있으며, 패킷 기반 분석에도 사용할 수 있다.

도면의 간단한 설명

- [0052] 도 1은 본 발명의 일 실시예에 따른 네트워크 침입 탐지 시스템의 구성을 개략적으로 나타낸 블록도이다.
- 도 2는 본 발명의 일 실시예에 따른 패킷 플로우 분류기 모델 학습부의 구성을 개략적으로 나타낸 블록도이다.
- 도 3은 본 발명의 일 실시예에 따른 바이트 임베딩 처리부, 패킷 임베딩 처리부, 플로우 임베딩 처리부 및 패킷 플로우 분류 처리부를 개략적으로 나타낸 도면이다.
- 도 4는 본 발명의 일 실시예에 따라 네트워크 침입 탐지 방법을 나타낸 흐름도이다.
- 도 5는 본 발명의 일 실시예에 따라 패킷 플로우 임베딩을 수행하여 네트워크 공격 여부를 탐지하는 모델을 학습하는 방법을 나타낸 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0053] 본 발명의 실시예들은 본 발명의 기술적 사상을 설명하기 위한 목적으로 예시된 것이다. 본 발명에 따른 권리

범위가 이하에 제시되는 실시예들이나 이들 실시예들에 대한 구체적 설명으로 한정되는 것은 아니다.

- [0054] 본 발명에 사용되는 모든 기술적 용어들 및 과학적 용어들은, 달리 정의되지 않는 한, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 일반적으로 이해되는 의미를 갖는다. 본 발명에 사용되는 모든 용어들은 본 발명을 더욱 명확히 설명하기 위한 목적으로 선택된 것이며 본 발명에 따른 권리범위를 제한하기 위해 선택된 것이 아니다.
- [0055] 본 발명에서 사용되는 "포함하는", "구비하는", "갖는" 등과 같은 표현은, 해당 표현이 포함되는 어구 또는 문장에서 달리 언급되지 않는 한, 다른 실시예를 포함할 가능성을 내포하는 개방형 용어(open-ended terms)로 이해되어야 한다.
- [0056] 본 발명에서 기술된 단수형의 표현은 달리 언급하지 않는 한 복수형의 의미를 포함할 수 있으며, 이는 청구범위에 기재된 단수형의 표현에도 마찬가지로 적용된다.
- [0057] 본 발명에서 사용되는 "제1", "제2" 등의 표현들은 복수의 구성요소들을 상호 구분하기 위해 사용되며, 해당 구성요소들의 순서 또는 중요도를 한정하는 것은 아니다.
- [0058] 본 발명에서 사용되는 용어 "부"는, 소프트웨어, 또는 FPGA(field-programmable gate array), ASIC(application specific integrated circuit)과 같은 하드웨어 구성요소를 의미한다. 그러나, "부"는 하드웨어 및 소프트웨어에 한정되는 것은 아니다. "부"는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고, 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서, "부"는 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스, 함수, 속성, 프로시저, 서브루틴, 프로그램 코드의 세그먼트, 드라이버, 펌웨어, 마이크로코드, 회로, 데이터, 데이터베이스, 데이터 구조, 테이블, 어레이 및 변수를 포함한다. 구성요소와 "부" 내에서 제공되는 기능은 더 작은 수의 구성요소 및 "부"로 결합되거나 추가적인 구성요소와 "부"로 분리될 수 있다.
- [0059] 본 발명에서 사용되는 "~에 기초하여"라는 표현은, 해당 표현이 포함되는 어구 또는 문장에서 기술되는, 결정 판단의 행위 또는 동작에 영향을 주는 하나 이상의 인자를 기술하는데 사용되며, 이 표현은 결정, 판단의 행위 또는 동작에 영향을 주는 추가적인 인자를 배제하지 않는다.
- [0060] 본 발명에서, 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 경우, 어떤 구성요소가 다른 구성요소에 직접적으로 연결될 수 있거나 접속될 수 있는 것으로, 또는 새로운 다른 구성요소들 매개로 하여 연결될 수 있거나 접속될 수 있는 것으로 이해되어야 한다.
- [0061] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예들을 설명한다. 첨부된 도면에서, 동일하거나 대응하는 구성요소에는 동일한 참조부호가 부여되어 있다. 또한, 이하의 실시예들의 설명에 있어서, 동일하거나 대응하는 구성요소를 중복하여 기술하는 것이 생략될 수 있다. 그러나, 구성요소에 관한 기술이 생략되어도, 그러한 구성요소가 어떤 실시예에 포함되지 않는 것으로 의도되지는 않는다.
- [0062] 도 1은 본 발명의 일 실시예에 따른 네트워크 침입 탐지 시스템의 구성을 개략적으로 나타낸 블록도이다. 도 1을 참조하면, 네트워크 침입 탐지 시스템(100)은 패킷 수집부(110), 패킷 전처리부(120) 및 기계학습 모델부(130)를 포함할 수 있다.
- [0063] 패킷 수집부(110)는 네트워크 장비(도시하지 않음)를 통해 패킷들을 수집할 수 있다. 일 실시예에 있어서, 네트워크 장비는 라우터를 포함할 수 있다. 예를 들면, 패킷 수집부(110)는 라우터에 연결되고, 라우터를 통해 송수신되는 패킷들을 수집할 수 있다.
- [0064] 일 실시예에 있어서, 패킷 수집부(110)는 데이터베이스(도시하지 않음)를 포함할 수 있다. 예를 들면, 패킷 수집부(110)는 패킷 수집 시 패킷의 맥(media access control, MAC) 주소, 아이피 주소(IP address), 및 전송 포트 레벨 프로토콜(transport level protocol)을 확인하여 패킷을 구분하고, 구분된 패킷을 데이터베이스(도시하지 않음)에 저장할 수 있다. 여기서, 전송 포트 레벨 프로토콜은 네트워크 통신 계층 중 전송 계층의 규약 정보를 나타낸다. 예를 들면, 전송 포트 레벨 프로토콜은 TCP(transmission control protocol), UDP(user datagram protocol) 등을 포함할 수 있다.
- [0065] 일 실시예에 있어서, 패킷 수집부(110)는 패킷을 순차적인 흐름으로 분석하기 위해 패킷 헤더 정보에 기초하여 패킷을 분류할 수 있다. 예를 들면, 패킷 수집부(110)는 패킷 헤더 정보를 기준으로 동일한 플로우에 해당하는 패킷을 분류할 수 있다. 여기서, 플로우는 송신자와 수신자가 동일한 패킷의 흐름을 나타낸다. 예를 들면, 플로우는 소스 아이피 주소(source IP address), 목적지 아이피 주소(destination IP address), 소스 포트 번호

(source port number), 목적지 포트 번호(destination port number), 및 트랜스포트 레벨 프로토콜의 필드가 동일한 패킷을 송신자와 수신자가 동일한 패킷으로 간주하고, 동일한 플로우에 해당하는 패킷으로 분류할 수 있다.

- [0066] 패킷 전처리부(120)는 패킷 수집부(110)에 연결될 수 있다. 패킷 전처리부(120)는 패킷 수집부(110)에 의해 분류된 패킷들에 대해 패킷 플로우를 구성하기 위한 전처리를 수행할 수 있다.
- [0067] 일 실시예에 있어서, 패킷 전처리부(120)는 복수의 패킷으로 구성되고 복수의 패킷 각각이 소정의 길이를 갖는 복수의 패킷 플로우를 생성할 수 있다. 예를 들면, 복수의 패킷 플로우 각각은 20개의 패킷으로 구성될 수 있다. 또한, 복수의 패킷 각각은 800 바이트(bytes)의 길이를 가질 수 있다. 여기서, 패킷 플로우의 크기를 동일하게 하는 것은 기계학습 모델이 고정된 크기의 값을 입력으로 하기 때문이다. 일 실시예에 있어서, 패킷 플로우는 20×800 의 크기의 행렬 형태로 구성될 수 있다.
- [0068] 일 실시예에 있어서, 패킷 전처리부(120)는 패킷을 기계학습 모델부(130)의 임베딩 모델에 입력하기 전에, 패킷의 헤더 정보에 대해 임의의 값으로 전처리를 수행할 수 있다. 여기서, 임베딩은 범주형 또는 수치형 데이터를 잘 표현할 수 있는 저차원의 연속형 벡터(특징)을 나타낸다. 패킷 헤더 정보, 예를 들어 맥 주소, 소스 아이피 주소, 목적지 아이피 주소, 소스 포트 번호 및 목적지 포트 번호의 필드 값들은 기계학습 모델부(130)의 기계학습 모델의 과적합을 유도하여 예측 정확도가 떨어질 수 있다. 따라서, 기계학습 모델의 예측 정확도에 영향을 미치지 않도록 해당 필드 값을 임의의 값으로 대체할 필요가 있다. 예를 들면, 패킷 전처리부(120)는 각 패킷의 소스 아이피 주소, 목적지 아이피 주소, 소스 맥 주소 및 목적지 맥 주소를 임의의 값으로 대체할 수 있다. 이것은 패킷을 악성 또는 양성으로 분류할 때 기계학습 모델이 특정 주소 정보만을 통해 분류하는 것을 방지하기 위함이며 이를 통해 기계학습 모델은 정상 패킷 또는 악성 패킷의 의미와, 패킷 플로우의 의미를 학습하여 이해하게 될 수 있다.
- [0069] 기계학습 모델부(130)는 패킷 전처리부(120)에 연결될 수 있다. 기계학습 모델부(130)는 패킷 전처리부(120)에 의해 생성된 패킷 플로우에 기초하여 네트워크 공격 여부를 탐지하는 모델을 학습하고, 학습된 모델에 기초하여 새로운 패킷 플로우(예를 들어, 미지의 패킷 플로우)에 대한 공격 여부 결과를 도출할 수 있다. 일 실시예에 있어서, 기계학습 모델부(130)는 패킷 플로우에 임베딩을 수행하고, 임베딩된 패킷 플로우에 기초하여 네트워크 공격 여부를 탐지하는 모델을 학습하고, 학습된 모델을 사용하여 새로이 수집되는 패킷에 대한 공격 여부 또는 공격 유형을 탐지한 결과를 도출할 수 있다.
- [0070] 일 실시예에 있어서, 기계학습 모델부(130)는 도 1에 도시된 바와 같이, 패킷 플로우 분류기 모델 학습부(131) 및 패킷 플로우 분류기 결과 예측부(132)를 포함할 수 있다.
- [0071] 패킷 플로우 분류기 모델 학습부(131)는 패킷 전처리부(120)에 의해 생성된 패킷 플로우를 수신하고, 수신된 패킷 플로우를 임베딩으로 변환할 수 있다. 또한, 패킷 플로우 분류기 모델 학습부(131)는 변환된 임베딩에 기초하여 네트워크 공격 여부를 탐지하는 모델을 학습할 수 있다. 여기서, 학습은 패킷 플로우가 입력되었을 때 기계학습 모델부(130)가 공격 여부 또는 공격 타입을 예측하도록 하는 과정을 나타낸다.
- [0072] 일 실시예에 있어서, 패킷 플로우 분류기 모델 학습부(131)는 패킷 플로우와 패킷 플로우에 해당하는 공격 여부(예를 들어, 정상 또는 악성 여부), 또는 공격 타입의 라벨 쌍을 사용하여 모델을 학습할 수 있다. 예를 들면, 해당하는 형식의 공개 데이터셋은 CIC-IDS2017 및 NSL-KDD를 포함할 수 있다. 그러나, 공개 데이터셋은 반드시 이에 한정되는 것은 아니다.
- [0073] 패킷 플로우 분류기 결과 예측부(132)는 패킷 플로우 분류기 모델 학습부(131)에 의해 학습된 모델을 업데이트하고, 업데이트된 모델을 사용하여 미지의 패킷 플로우, 예를 들어 새로이 입력되는 패킷 플로우에 대한 공격(예를 들어, 정상 또는 악성) 여부를 분류할 수 있다. 또한, 패킷 플로우 분류기 결과 예측부(132)는 업데이트된 모델을 사용하여 미지의 패킷 플로우에 대한 공격 타입을 분류할 수 있다.
- [0074] 도 2는 본 발명의 일 실시예에 따른 패킷 플로우 모델 학습부(131)의 구성을 개략적으로 나타낸 블록도이다. 도 2를 참조하면, 패킷 플로우 모델 학습부(131)는 바이트 임베딩 처리부(210), 패킷 임베딩 처리부(220), 플로우 임베딩 처리부(230) 및 패킷 플로우 분류 처리부(240)를 포함할 수 있다.
- [0075] 바이트 임베딩 처리부(210)는 패킷 플로우에 기초하여 패킷을 구성하는 바이트에 대한 바이트 임베딩을 생성할 수 있다. 일 실시예에 있어서, 바이트 임베딩 처리부(210)는 패킷을 구성하고 있는 바이트 정보(예를 들어, 중요한 바이트 정보)를 반영한 바이트 임베딩을 생성할 수 있다. 일 실시예에 있어서, 바이트 임베딩 처리부

(210)는 임베딩을 생성하기 위해, 도 3에 도시된 바와 같이, 크게 세 가지 부분을 포함할 수 있다.

- [0076] 첫 번째는 전결합 층으로 구성된 단순 임베딩 부분(311) 및 생성된 임베딩에 위치 정보를 추가하는 위치 정보 인코딩(positional encoding) 부분(312)이다. 해당 방법은 순서 정보를 포착할 수 있지만 재귀적인 구조로 인해 학습 및 예측 시간이 오래 걸리는 순환 신경망 구조를 사용하지 않는 대신 위치 정보 인코딩을 통해 바이트의 순서 정보를 유지시킬 수 있다.
- [0077] 두 번째는 패킷을 구성하는 바이트들의 관계를 파악하고 각 바이트들의 문맥정보를 포함하는 임베딩을 찾아내는 멀티 헤드 어텐션(multi-head attention) 부분(313)이다. 여기서, 어텐션(attention)은 자연어 처리(natural language process, NLP)에 자주 사용되는 방법이다. 어텐션은 입력 값으로부터 예측을 수행할 때, 입력 값에서 예측에 중요한 정보를 주는 부분을 찾아내는 역할을 수행한다. 예를 들면, 입력 값은 번역 서비스에서 번역하고 싶은 입력 문장일 수 있다. 멀티 헤드 어텐션(313)은 어텐션을 한 번만 수행하는 것이 아니라 복수 회 반복 수행한다. 본 실시예에서는 멀티 헤드 어텐션을 입력 값과 출력 값 사이의 관계를 파악하기 위해 사용하는 대신 입력 값 사이의 관계를 파악하기 위해 사용한다.
- [0078] 세 번째는 전결합 층(314)이다. 전결합 층(314)은 멀티 헤드 어텐션(313)을 통해 출력되는 바이트 임베딩을 비선형 변환시키는 역할을 수행할 수 있다.
- [0079] 패킷 임베딩 처리부(220)는 바이트 임베딩 처리부(210)의 처리 결과를 종합하여 패킷에 대한 임베딩을 수행할 수 있다. 일 실시예에 있어서, 패킷 임베딩 처리부(220)는 바이트 임베딩 처리부(210)에 의해 생성된 복수의 바이트 임베딩을 종합하여 패킷에 대한 패킷 임베딩을 생성할 수 있다. 일 실시예에 있어서, 패킷 임베딩 처리부(220)는 중요한 바이트 정보를 강조하고 의미 없는 바이트 정보를 제외시키기 위해 애디티브 어텐션(additive attention)을 사용할 수 있다.
- [0080] 예를 들면, 상기 애디티브 어텐션은 전결합 층 및 소프트 맥스 함수로 구성되어 있다. 이는, 어텐션을 주기 위해 애디티브 어텐션을 사용하며, 해당 구조가 전 결합 층 및 소프트 맥스로 구성된다. 애디티브 어텐션을 통해 패킷에 어텐션 가중치를 부여한다. 주용한 정보를 포함하는 패킷일수록 더 큰 어텐션 가중치 값을 갖는다. 네트워크 플로우 벡터는 패킷 벡터와 어텐션 가중치의 가중합으로 나타난다.
- [0081] 플로우 임베딩 처리부(230)는 패킷 임베딩 처리부(220)의 처리 결과에 기초하여 패킷 플로우에 대한 플로우 임베딩을 생성할 수 있다. 일 실시예에 있어서, 플로우 임베딩 처리부(230)는 패킷 임베딩 처리부(220)에 의해 생성된 복수의 패킷 임베딩에 기초하여 패킷 플로우에 대한 플로우 임베딩을 생성할 수 있다. 일 실시예에 있어서, 플로우 임베딩 처리부(230)는 중요한 바이트 정보를 강조하고 의미 없는 바이트 정보를 제외시키기 위해 애디티브 어텐션을 사용할 수 있다. 예를 들면, 플로우 임베딩 처리부(230)는 도 3에 도시된 바와 같이, 전결합 층(321) 및 소프트 맥스 함수(322)를 사용하여 패킷 임베딩을 생성할 수 있다.
- [0082] 일 실시예에 있어서, 플로우 임베딩 처리부(230)는 패킷 임베딩 처리부(220)와 작동 방식이 동일하다. 다만, 패킷 임베딩 처리부(220)는 복수의 바이트 임베딩을 이용하여 패킷 임베딩을 생성하는 반면, 플로우 임베딩 처리부(230)는 패킷 임베딩 처리부(220)에 의해 생성된 복수의 패킷 임베딩으로부터 패킷 플로우에 대한 플로우 임베딩을 생성할 수 있다.
- [0083] 패킷 플로우 분류 처리부(240)는 플로우 임베딩 처리부(230)에 의해 생성된 플로우 임베딩에 기초하여 패킷 플로우를 분류할 수 있다. 일 실시예에 있어서, 패킷 플로우 분류 처리부(240)는 플로우 임베딩을 사용하여 패킷 플로우의 공격 여부(예를 들어, 정상 또는 악성) 또는 공격 유형의 탐지를 수행할 수 있다. 일 실시예에 있어서, 패킷 플로우 분류 처리부(240)는 도 3에 도시된 바와 같이, 전결합 층(341) 및 소프트 맥스 함수(342)에 플로우 임베딩을 통과시켜 공격 여부 또는 공격 유형의 탐지를 수행할 수 있다.
- [0084] 본 발명에 도시된 흐름도에서 프로세스 단계들, 방법 단계들, 알고리즘들 등이 순차적인 순서로 설명되었지만, 그러한 프로세스들, 방법들 및 알고리즘들은 임의의 적합한 순서로 작동되도록 구성될 수 있다. 다시 말하면, 본 발명의 다양한 실시예들에서 설명되는 프로세스들, 방법들 및 알고리즘들의 단계들이 본 발명에서 기술된 순서로 수행될 필요는 없다. 또한, 일부 단계들이 비동시적으로 수행되는 것으로서 설명되더라도, 다른 실시예에서는 이러한 일부 단계들이 동시에 수행될 수 있다. 또한, 도면에서의 묘사에 의한 프로세스의 예시는 예시된 프로세스가 그에 대한 다른 변화들 및 수정들을 제외하는 것을 의미하지 않으며, 예시된 프로세스 또는 그의 단계들 중 임의의 것이 본 발명의 다양한 실시예들 중 하나 이상에 필수적임을 의미하지 않으며, 예시된 프로세스가 바람직하다는 것을 의미하지 않는다.
- [0085] 도 4는 본 발명의 일 실시예에 따라 네트워크 공격을 탐지하는 방법을 나타낸 흐름도이다. 도 4를 참조하면,

단계 S402에서, 패킷 수집부(110)는 네트워크 장비를 통해 송수신되는 패킷을 수집할 수 있다. 일 실시예에 있어서, 패킷 수집부(110)는 라우터를 통해 송수신되는 복수의 패킷을 수집할 수 있다. 일 실시예에 있어서, 패킷 수집부(110)는 패킷을 순차적인 흐름으로 분석하기 위해 패킷 헤더 정보에 기초하여 패킷을 분류할 수 있다. 예를 들면, 패킷 수집부(110)는 패킷 헤더 정보를 기준으로 동일한 플로우에 있는 패킷을 분류할 수 있다.

[0086] 단계 S404에서, 패킷 전처리부(120)는 패킷 수집부(110)로부터 제공되는 패킷에 전처리를 수행할 수 있다. 일 실시예에 있어서, 패킷 전처리부(120)는 패킷 수집부(110)에 의해 분류된 패킷들에 패킷 플로우를 구성하기 위한 전처리를 수행할 수 있다.

[0087] 예를 들면, 패킷 전처리부(120)는 복수의 패킷으로 구성되고 복수의 패킷 각각이 소정의 길이를 갖는 복수의 패킷 플로우를 생성할 수 있다. 또한, 패킷 전처리부(120)는 패킷을 기계학습 모델부(130)의 임베딩 모델에 입력하기 전에, 패킷의 헤더 정보에 대해 임의의 값으로 전처리를 수행할 수 있다.

[0088] 단계 S406에서, 기계학습 모델부(130)는 패킷 전처리부(120)에 의해 생성된 패킷 플로우에 임베딩을 수행하여 네트워크 공격 여부를 탐지하는 모델을 학습할 수 있다. 일 실시예에 있어서, 기계학습 모델부(130)의 패킷 플로우 분류기 학습 모델부(131)는 패킷 전처리부(120)에 의해 생성된 패킷 플로우를 입력받고, 입력된 패킷 플로우에 임베딩을 수행하여 네트워크 공격 여부를 탐지하는 모델을 학습할 수 있다.

[0089] 단계 S408에서, 기계학습 모델부(130)는 학습된 모델을 사용하여 새로이 수집되는 패킷에 대한 공격 여부 또는 공격 유형을 탐지한 결과를 도출할 수 있다. 일 실시예에 있어서, 기계학습 모델부(130)의 패킷 플로우 분류기 결과 예측부(132)는 패킷 플로우 분류기 학습 모델부(131)에 의해 학습된 모델을 업데이트하고, 업데이트된 모델을 사용하여 미지의 패킷 플로우(예를 들어, 새로이 입력되는 패킷 플로우)에 대한 공격 여부 또는 공격 유형의 탐지 결과를 도출할 수 있다.

[0090] 도 5는 본 발명의 일 실시예에 따라 패킷 플로우에 임베딩을 수행하여 네트워크 공격 여부를 탐지하는 모델을 학습하는 방법을 나타낸 흐름도이다. 도 5를 참조하면, 단계 S502에서, 기계학습 모델부(130)는 패킷 전처리부(120)에 의해 생성된 패킷 플로우에 기초하여 패킷을 구성하는 바이트들에 대한 바이트 임베딩을 생성할 수 있다. 일 실시예에 있어서, 기계학습 모델부(130)의 바이트 임베딩 처리부(210)는 패킷을 구성하는 바이트들에 바이트 인코딩을 수행하여 바이트 임베딩을 생성할 수 있다.

[0091] 단계 S504에서, 기계학습 모델부(130)는 바이트 임베딩을 종합하여 패킷에 대한 패킷 임베딩을 생성할 수 있다. 일 실시예에 있어서, 기계학습 모델부(130)의 패킷 임베딩 처리부(220)는 바이트 임베딩 처리부(210)에 의해 생성된 복수의 바이트 임베딩에 패킷 인코딩을 수행하여 패킷에 대한 패킷 임베딩을 생성할 수 있다.

[0092] 단계 S506에서, 기계학습 모델부(130)는 패킷 임베딩에 기초하여 패킷 플로우에 대한 플로우 임베딩을 생성할 수 있다. 일 실시예에 있어서, 기계학습 모델부(130)의 플로우 임베딩 처리부(230)는 패킷 임베딩 처리부(220)에 의해 생성된 복수의 패킷 임베딩에 플로우 인코딩을 수행하여 패킷 플로우에 대한 플로우 임베딩을 생성할 수 있다.

[0093] 단계 S508에서, 기계학습 모델부(130)는 플로우 임베딩에 기초하여 패킷 플로우를 분류할 수 있다. 일 실시예에 있어서, 기계학습 모델부(130)의 패킷 플로우 분류 처리부(240)는 플로우 임베딩 처리부(230)에 의해 생성된 플로우 임베딩을 사용하여 패킷 플로우의 공격 여부(예를 들면, 정상 또는 악성) 또는 공격 유형의 탐지를 수행할 수 있다.

[0094] 위 방법은 특정 실시예들을 통하여 설명되었지만, 위 방법은 또한 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의해 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등이 있다. 또한, 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다. 그리고, 위 실시예들을 구현하기 위한 기능적인(functional) 프로그램, 코드 및 코드 세그먼트들은 본 발명이 속하는 기술분야의 프로그래머들에 의해 용이하게 추론될 수 있다.

[0095] 이상 일부 실시예들과 첨부된 도면에 도시된 예에 의해 본 발명의 기술적 사상이 설명되었지만, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 이해할 수 있는 본 발명의 기술적 사상 및 범위를 벗어나지 않는 범위에서 다양한 치환, 변형 및 변경이 이루어질 수 있다는 점을 알아야 할 것이다. 또한, 그러한 치환, 변형 및 변경은 첨부된 청구범위 내에 속하는 것으로 생각되어야 한다.

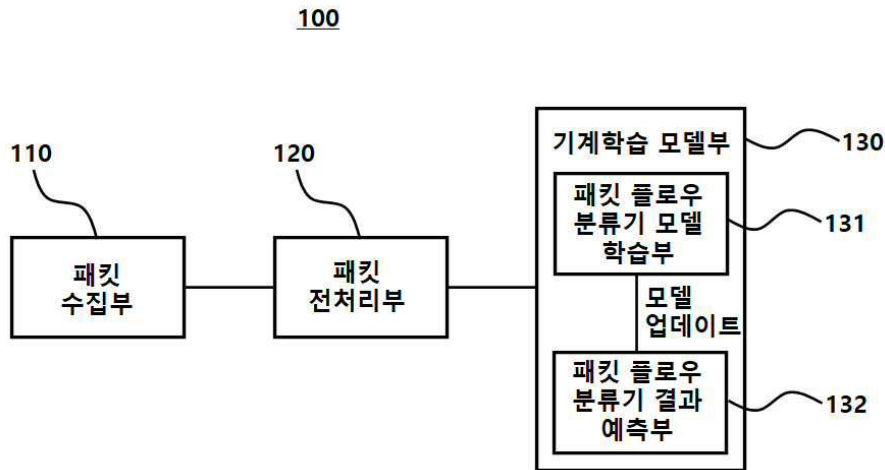
부호의 설명

[0096]

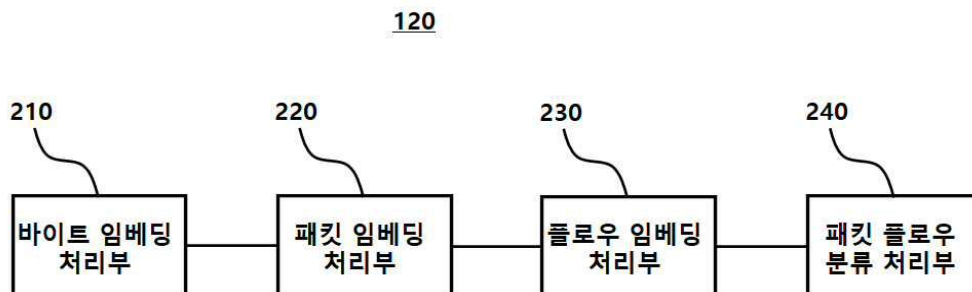
100: 네트워크 침입 탐지 시스템, 110: 패킷 수집부, 120: 패킷 전처리부, 130: 기계학습 모델부, 131: 패킷 플로우 분류기 학습 모델부, 132: 패킷 플로우 분류기 결과 예측부, 210: 바이트 임베딩 처리부, 220: 패킷 임베딩 처리부, 230: 플로우 임베딩 처리부, 240: 패킷 플로우 분류 처리부

도면

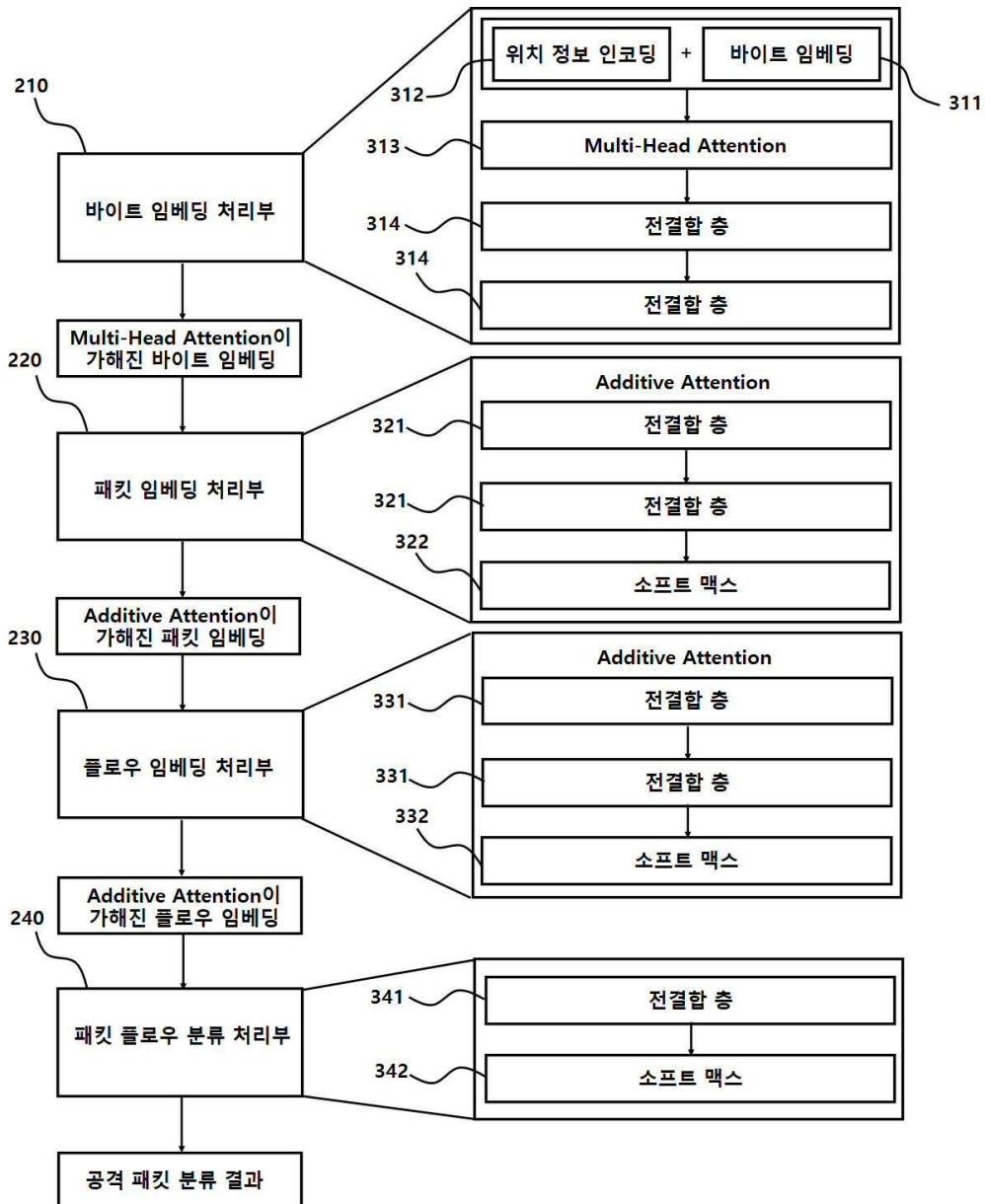
도면1



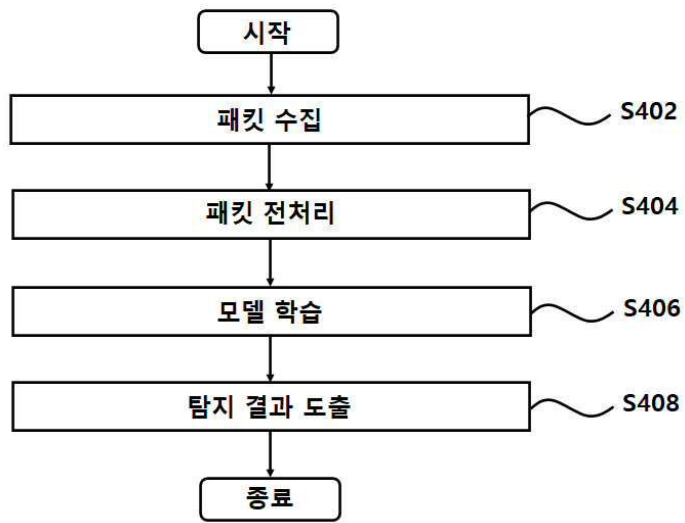
도면2



도면3



도면4



도면5

