(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0076389
(43) 공개일자 2023년05월31일

(51) 국제특허분류(Int. Cl.)
G06N 3/04 (2023.01) *G06F 21/55* (2013.01)
G06N 20/20 (2019.01)
(52) CPC특허분류
G06N 3/045 (2023.01)
G06F 21/554 (2013.01)
(21) 출원번호 10-2021-0163254
(22) 출원일자 2021년11월24일
심사청구일자 2021년11월24일

(71) 출원인
주식회사 윈스
경기도 성남시 분당구 판교로228번길 15, 제윈스동(삼평동, 판교세븐벤처벨리1)
(72) 발명자
최병환
경기도 성남시 분당구 미금로 63, 307동 303호(구미동, 무지개마을)
(74) 대리인
특허법인충정

전체 청구항 수 : 총 20 항

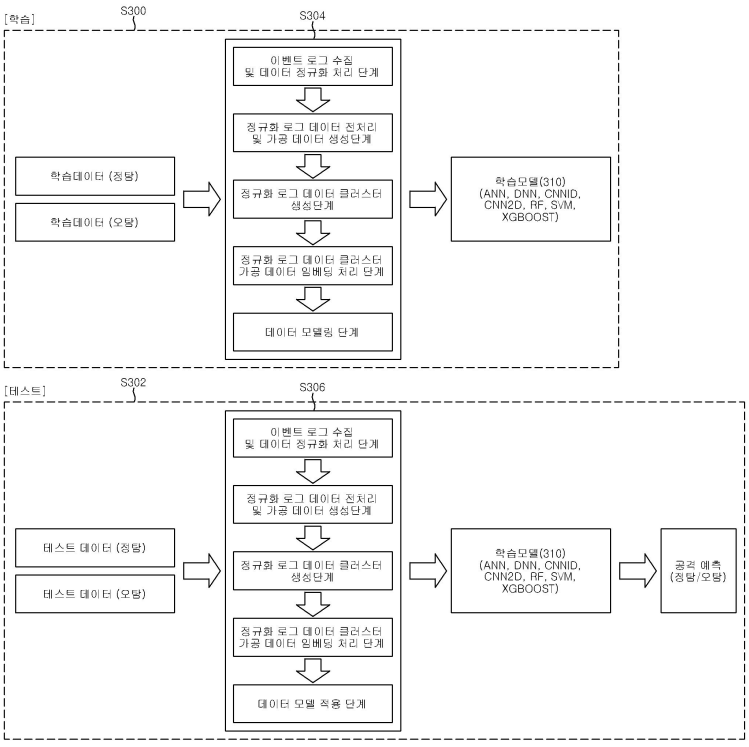
(54) 발명의 명칭 **인공지능 기반 정오탐 식별 모델 생성 방법과 장치 및 인공지능 기반 정오탐 식별 방법과 장치**

(57) 요 약

본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법은, A) 인공지능 기반 정오탐 식별 장치가, 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷에 기반하여 복수의 인공지능 기반 데이터 모델을 생성하는 단계; 및 (B) 상기 복수의 인공지능 기반 데이터 모델에 기반하여, 테스트 데이터인 정탐 또는 오탐일 수 있는 복

(뒷면에 계속)

대 표 도



수의 공격 패킷의 정탐 또는 오탐 여부를 식별하는 단계를 포함하고, 상기 단계 (A)는, (A-1) 상기 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷으로부터 각각 복수 유형의 특징들을 생성하는 단계; (A-2) 상기 복수 유형의 특징들 중 한 유형의 특징들을 클러스터링하여 복수의 클러스터를 생성하는 단계; (A-3) 상기 복수의 클러스터 각각에서 하나의 특징을 선택하는 단계; 및 (A-4) 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 공격 패킷의 정탐 또는 오탐을 식별하기 위한 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 생성하는 단계를 포함한다.

(52) CPC특허분류

G06N 20/20 (2021.08)

명 세 서

청구범위

청구항 1

(A) 인공지능 기반 정오탐 식별 모델 생성 장치가, 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷으로부터 각각 복수 유형의 특징들을 생성하는 단계;

(B) 상기 복수 유형의 특징들 중 한 유형의 특징들을 클러스터링하여 복수의 클러스터를 생성하는 단계;

(C) 상기 복수의 클러스터 각각에서 하나의 특징을 선택하는 단계; 및

(D) 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 공격 패킷의 정탐 또는 오탐을 식별하기 위한 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 생성하는 단계를 포함하는, 인공지능 기반 정오탐 식별 모델 생성 방법.

청구항 2

청구항 1에 있어서,

상기 복수 유형의 특징들은 각각, 데이터 인덱스 정보, 데이터 컨텐츠 카운트 벡터 정보 및 데이터 토큰 정보를 포함하는, 인공지능 기반 정오탐 식별 모델 생성 방법.

청구항 3

청구항 2에 있어서,

상기 단계 (B)에서, 상기 한 유형의 특징들은 상기 데이터 컨텐츠 카운트 벡터 정보를 포함하는, 인공지능 기반 정오탐 식별 모델 생성 방법.

청구항 4

청구항 2에 있어서,

상기 단계 (B)는, 상기 데이터 컨텐츠 카운트 벡터 정보들 간의 유사도에 기반하여 상기 복수의 데이터 컨텐츠 카운트 벡터 정보를 클러스터링하여 복수의 클러스터를 생성하는, 인공지능 기반 정오탐 식별 모델 생성 방법.

청구항 5

청구항 1에 있어서,

상기 단계 (D)는,

(D-1) 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 상기 공격 패킷의 정탐 또는 오탐을 식별하기 위한 각 유형의 특징에 대한 복수의 인공지능 기반 데이터 모델을 생성하는 단계; 및

(D-2) 상기 각 유형의 특징에 대한 복수의 인공지능 기반 데이터 모델 중 정확도가 가장 높은 데이터 모델을 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델로 최종 선정하는 단계를 포함하는, 인공지능 기반 정오탐 식별 모델 생성 방법.

청구항 6

청구항 5에 있어서,

상기 단계 (D-1)의 복수의 인공지능 기반 데이터 모델은,

인공신경망(ANN: Artificial Neural Network), 심층 신경망(DNN: Deep Neural Network), 1D 컨벌루션 신경망(CNN1D: 1D Convolution Neural Network), 2D 컨벌루션 신경망(CNN2D: 2D Convolution Neural Network), 랜덤 포레스트(RF: Random Forest), 서포트 벡터 머신(SVM: Support Vector Machine) 및 XGBoost(Extreme Gradient

Boosting)에 기반한 데이터 모델들을 포함하는, 인공지능 기반 정오탐 식별 모델 생성 방법.

청구항 7

청구항 2에 있어서,

상기 데이터 인덱스 정보는, 상기 공격 패킷의 패킷 데이터의 디코딩된 데이터의 헥사(hexa)값들을 아스키 코드의 정수값들로 변환하여 생성된 인덱스 데이터를 포함하고,

상기 데이터 컨텐츠 카운트 벡터 정보는, 상기 디코딩된 데이터를 청킹(Chunking) 알고리즘을 이용하여 소정 개수의 문자 당 하나의 블록을 생성하여 해시값을 생성한 후 중복된 해시값을 누적카운트하여 생성된 소정 크기의 벡터값을 포함하며,

상기 데이터 토큰 정보는, 상기 디코딩된 데이터를 URL 디코딩하여 문장으로 형성하고 형성된 문장에서 중복문자를 제거하고, 상기 형성된 문장 내의 공백 및 특수문자를 소정 바이트의 식별가능한 문자로 대체한 후 문장을 분리하여 생성된 토큰값을 포함하는, 인공지능 기반 정오탐 식별 모델 생성 방법.

청구항 8

청구항 1에 있어서,

상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들 각각을 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위로 데이터를 구분하고 해싱처리한 후 해싱 처리된 각각의 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위 데이터를 연결하여 생성된 이진화된 값을, 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 생성하기 위한 입력 벡터값으로 사용하는, 인공지능 기반 정오탐 식별 모델 생성 방법.

청구항 9

학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷으로부터 각각 복수 유형의 특징들을 생성하는 가공 데이터 생성부;

상기 복수 유형의 특징들 중 한 유형의 특징들을 클러스터링하여 복수의 클러스터를 생성하는 데이터 클러스터 생성부;

상기 복수의 클러스터 각각에서 하나의 특징을 선택하는 데이터 선택부; 및

상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 공격 패킷의 정탐 또는 오탐을 식별하기 위한 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 생성하는 데이터 모델 생성부를 포함하는, 인공지능 기반 정오탐 식별 모델 생성 장치.

청구항 10

(A) 인공지능 기반 정오탐 식별 장치가, 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷에 기반하여 복수의 인공지능 기반 데이터 모델을 생성하는 단계; 및

(B) 상기 복수의 인공지능 기반 데이터 모델에 기반하여, 테스트 데이터인 정탐 또는 오탐일 수 있는 복수의 공격 패킷의 정탐 또는 오탐 여부를 식별하는 단계를 포함하고,

상기 단계 (A)는,

(A-1) 상기 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷으로부터 각각 복수 유형의 특징들을 생성하는 단계;

(A-2) 상기 복수 유형의 특징들 중 한 유형의 특징들을 클러스터링하여 복수의 클러스터를 생성하는 단계;

(A-3) 상기 복수의 클러스터 각각에서 하나의 특징을 선택하는 단계; 및

(A-4) 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 공격 패킷의 정탐 또는 오탐을 식별하기 위한 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 생성하는 단계를 포함하는, 인공지능 기반 정오탐 식별 방법.

청구항 11

청구항 10에 있어서,

상기 단계 (B)는,

(B-1) 상기 테스트 데이터인 정탐 또는 오탐일 수 있는 복수의 공격 패킷으로부터 각각 복수 유형의 특징들을 생성하는 단계;

(B-2) 상기 복수 유형의 특징들 중 한 유형의 특징들을 클러스터링하여 복수의 클러스터를 생성하는 단계;

(B-3) 상기 복수의 클러스터 각각에서 하나의 특징을 선택하는 단계; 및

(B-4) 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 사용하여 상기 복수의 클러스터 각각에서 선택된 하나의 특징에 대응하는 정탐 또는 오탐일 수 있는 공격 패킷의 정탐 또는 오탐 여부를 최종 결정하는 단계를 포함하는, 인공지능 기반 정오탐 식별 방법.

청구항 12

청구항 10에 있어서,

상기 복수 유형의 특징들은 각각, 데이터 인덱스 정보, 데이터 컨텐츠 카운트 벡터 정보 및 데이터 토큰 정보를 포함하는, 인공지능 기반 정오탐 식별 방법.

청구항 13

청구항 12에 있어서,

상기 단계 (A-2) 및 상기 단계 (B-2)에서, 상기 한 유형의 특징들은 상기 데이터 컨텐츠 카운트 벡터 정보를 포함하는, 인공지능 기반 정오탐 식별 방법.

청구항 14

청구항 12에 있어서,

상기 단계 (A-2) 및 상기 단계 (B-2) 각각은, 상기 데이터 컨텐츠 카운트 벡터 정보들 간의 유사도에 기반하여 상기 복수의 데이터 컨텐츠 카운트 벡터 정보를 클러스터링하여 복수의 클러스터를 생성하는, 인공지능 기반 정오탐 식별 방법.

청구항 15

청구항 10에 있어서,

상기 단계 (A-4)는,

(A-4-1) 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 상기 공격 패킷의 정탐 또는 오탐을 식별하기 위한 각 유형의 특징에 대한 복수의 인공지능 기반 데이터 모델을 생성하는 단계; 및

(A-4-2) 상기 각 유형의 특징에 대한 복수의 인공지능 기반 데이터 모델 중 정확도가 가장 높은 데이터 모델을 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델로 최종 선정하는 단계를 포함하는, 인공지능 기반 정오탐 식별 방법.

청구항 16

청구항 15에 있어서,

상기 단계 (A-4-1)의 복수의 인공지능 기반 데이터 모델은,

인공신경망(ANN: Artificial Neural Network), 심층 신경망(DNN: Deep Neural Network), 1D 컨벌루션 신경망(CNN1D: 1D Convolution Neural Network), 2D 컨벌루션 신경망(CNN2D: 2D Convolution Neural Network), 랜덤 포레스트(RF: Random Forest), 서포트 벡터 머신(SVM: Support Vector Machine) 및 XGBoost(Extreme Gradient

Boosting)에 기반한 데이터 모델들을 포함하는, 인공지능 기반 정오탐 식별 방법.

청구항 17

청구항 12에 있어서,

상기 데이터 인덱스 정보는, 상기 공격 패킷의 패킷 데이터의 디코딩된 데이터의 헥사(hexa)값들을 아스키 코드의 정수값들로 변환하여 생성된 인덱스 데이터를 포함하고,

상기 데이터 컨텐츠 카운트 벡터 정보는, 상기 디코딩된 데이터를 청킹(Chunking) 알고리즘을 이용하여 소정 개수의 문자 당 하나의 블록을 생성하여 해시값을 생성한 후 중복된 해시값을 누적카운트하여 생성된 소정 크기의 벡터값을 포함하며,

상기 데이터 토큰 정보는, 상기 디코딩된 데이터를 URL 디코딩하여 문장으로 형성하고 형성된 문장에서 중복문자를 제거하고, 상기 형성된 문장 내의 공백 및 특수문자를 소정 바이트의 식별가능한 문자로 대체한 후 문장을 분리하여 생성된 토큰값을 포함하는, 인공지능 기반 정오탐 식별 방법.

청구항 18

청구항 11에 있어서,

상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들 각각을 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위로 데이터를 구분하고 해싱처리한 후 해싱 처리된 각각의 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위 데이터를 연결하여 생성된 이진화된 값을, 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델의 입력 벡터값으로 사용하는, 인공지능 기반 정오탐 식별 방법.

청구항 19

청구항 11에 있어서,

상기 단계 (B-4)는,

상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 사용하여 상기 단계 (B-3)의 상기 복수의 클러스터 각각에서 선택된 하나의 특징에 대응하는 정탐 또는 오탐일 수 있는 공격 패킷의 정탐 또는 오탐을 식별하는 단계; 및

상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델의 정오탐 식별 결과를 집계하여 다수결에 따라 상기 복수의 클러스터 각각에서 선택된 하나의 특징에 대응하는 정탐 또는 오탐일 수 있는 공격 패킷의 정탐 또는 오탐 여부를 최종 결정하는 단계를 포함하는, 인공지능 기반 정오탐 식별 방법.

청구항 20

학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷에 기반하여 복수의 인공지능 기반 데이터 모델을 생성하는 정오탐 식별 모델 생성부; 및

상기 복수의 인공지능 기반 데이터 모델에 기반하여, 테스트 데이터인 정탐 또는 오탐일 수 있는 복수의 공격 패킷의 정탐 또는 오탐 여부를 식별하는 정오탐 식별부를 포함하고,

상기 정오탐 식별 모델 생성부는,

상기 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷으로부터 각각 복수 유형의 특징들을 생성하는 가공 데이터 생성부;

상기 복수 유형의 특징들 중 한 유형의 특징들을 클러스터링하여 복수의 클러스터를 생성하는 데이터 클러스터 생성부;

상기 복수의 클러스터 각각에서 하나의 특징을 선택하는 데이터 선택부; 및

상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 공격 패킷의 정탐 또는 오탐을 식별하기 위한 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 생성하는 데이터 모델 생성부를 포함하는, 인공지능 기반 정오탐 식별 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 정탐 또는 오탐일 수 있는 공격 패킷의 정오탐 여부를 식별할 수 있는 인공지능 기반 정오탐 식별 모델 생성 및 인공지능 기반 정오탐 식별에 관한 것으로, 특히 네트워크 트래픽 환경에서 공격 패킷을 수집하여 패킷 데이터 기반으로 학습하여 인공지능(AI)/머신러닝(ML) 모델을 생성하고 이를 이용하여 공격을 탐지하기 위한, 인공지능 기반 정오탐 식별 모델 생성 방법과 장치 및 인공지능 기반 정오탐 식별 방법과 장치에 관한 것이다.

[0002] 본 발명은 하기의 국가연구개발사업 과제에 의해 지원되었다.

[0003] 부처명: 과학기술정보통신부

[0004] 과제관리 기관명: 한국인터넷진흥원(KISA)

[0005] 연구사업명: 2021년 AI기반 보안 제품 및 서비스 개발 지원사업

[0006] 연구과제명: 인공지능 지속학습과 분산마이닝 기법을 적용한 차세대 SIEM 솔루션 개발

[0007] 기여율: 1/1

[0008] 과제수행기관명: (주)윈스

[0009] 연구기간: 2021.06.09 ~ 2021.11.30 (6개월)

배 경 기 술

[0010] 네트워크 트래픽 환경에서 공격의 정오탐을 식별하기 위해 공격 패킷을 분석해야 하며, 공격 패킷을 매번 사람이 분석하기에는 한계가 있다.

[0011] 또한, 공격 패킷의 정오탐을 식별하기 위한 인공지능 학습 모델을 생성하는데 있어서, 정탐으로 판정된 복수의 공격 패킷 및 오탐으로 판정된 복수의 공격 패킷을 사용하여 학습 모델을 생성해야 하는데, 정오탐을 식별하기 위한 최적의 학습 모델을 생성하기 위해서는 정탐으로 판정된 대량의 공격 패킷 및 오탐으로 판정된 대량의 공격 패킷을 사용하여 학습을 해야 하기 때문에, 학습 모델을 생성하는 데 시간이 많이 소요되는 문제점이 있다.

[0012] 또한, 네트워크 트래픽 환경에서, 공격 패킷이 급증함으로 인하여 공격 패킷의 정오탐 여부를 식별하는데 상당한 시간이 요소되고 있으며, 또한 변종 공격 패킷이 급증하고 있다.

[0013] 따라서, 사람을 대신하여 인공지능(AI) 학습 모델을 활용하여 공격 패킷의 정오탐을 식별하고 공격을 탐지할 수 있으며, 최적의 학습 모델을 생성하는데 소요되는 시간을 단축할 수 있고, 공격 패킷의 정오탐 여부를 식별하는데 소요되는 시간을 단축할 수 있으며, 변종 공격 패킷을 탐지할 수 있는 기술이 필요하다.

선행기술문헌

특허문헌

[0014] (특허문헌 0001) KR 10-2271449 B1

발명의 내용

해결하려는 과제

[0015] 본 발명이 해결하고자 하는 과제는 네트워크 트래픽 환경에서 공격 패킷의 특성을 파악하여 사람이 아닌 인공지능(AI)/머신러닝(ML)을 통해 공격 패킷을 자동 분석하고 이를 통해 공격 패킷의 정탐 또는 오탐을 식별하고 공격을 탐지할 수 있는 인공지능 기반 정오탐 식별 모델 생성 방법과 장치 및 인공지능 기반 정오탐 식별 방법과 장치를 제공하는 것이다.

[0016] 본 발명이 해결하고자 하는 다른 과제는 네트워크 트래픽 환경에서 공격 패킷의 특성을 파악하여 변종 공격을

탐지할 수 있는 인공지능 기반 정오탐 식별 모델 생성 방법과 장치 및 인공지능 기반 정오탐 식별 방법과 장치를 제공하는 것이다.

[0017] 본 발명이 해결하고자 하는 또 다른 과제는 최적의 학습 모델을 생성하는데 소요되는 시간을 단축할 수 있는 인공지능 기반 정오탐 식별 모델 생성 방법과 장치 및 인공지능 기반 정오탐 식별 방법과 장치를 제공하는 것이다.

[0018] 본 발명이 해결하고자 하는 또 다른 과제는 대량의 공격 패킷의 정탐 또는 오탐 여부를 식별하는데 소요되는 시간을 단축할 수 있는 인공지능 기반 정오탐 식별 모델 생성 방법과 장치 및 인공지능 기반 정오탐 식별 방법과 장치를 제공하는 것이다.

과제의 해결 수단

[0019] 상기 과제를 해결하기 위한 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 모델 생성 방법은,

[0020] (A) 인공지능 기반 정오탐 식별 모델 생성 장치가, 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷으로부터 각각 복수 유형의 특징들을 생성하는 단계;

[0021] (B) 상기 복수 유형의 특징들 중 한 유형의 특징들을 클러스터링하여 복수의 클러스터를 생성하는 단계;

[0022] (C) 상기 복수의 클러스터 각각에서 하나의 특징을 선택하는 단계; 및

[0023] (D) 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 공격 패킷의 정탐 또는 오탐을 식별하기 위한 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 생성하는 단계를 포함한다.

[0024] 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 모델 생성 방법에 있어서, 상기 복수 유형의 특징들은 각각, 데이터 인덱스 정보, 데이터 컨텐츠 카운트 벡터 정보 및 데이터 토큰 정보를 포함할 수 있다.

[0025] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 모델 생성 방법에 있어서, 상기 단계 (B)에서, 상기 한 유형의 특징들은 상기 데이터 컨텐츠 카운트 벡터 정보를 포함할 수 있다.

[0026] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 모델 생성 방법에 있어서, 상기 단계 (B)는, 상기 데이터 컨텐츠 카운트 벡터 정보들 간의 유사도에 기반하여 상기 복수의 데이터 컨텐츠 카운트 벡터 정보를 클러스터링하여 복수의 클러스터를 생성할 수 있다.

[0027] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 모델 생성 방법에 있어서, 상기 단계 (D)는,

[0028] (D-1) 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 상기 공격 패킷의 정탐 또는 오탐을 식별하기 위한 각 유형의 특징에 대한 복수의 인공지능 기반 데이터 모델을 생성하는 단계; 및

[0029] (D-2) 상기 각 유형의 특징에 대한 복수의 인공지능 기반 데이터 모델 중 정확도가 가장 높은 데이터 모델을 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델로 최종 선정하는 단계를 포함할 수 있다.

[0030] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 모델 생성 방법에 있어서, 상기 단계 (D-1)의 복수의 인공지능 기반 데이터 모델은,

[0031] 인공신경망(ANN: Artificial Neural Network), 심층 신경망(DNN: Deep Neural Network), 1D 컨벌루션 신경망(CNN1D: 1D Convolution Neural Network), 2D 컨벌루션 신경망(CNN2D: 2D Convolution Neural Network), 랜덤 포레스트(RF: Random Forest), 서포트 벡터 머신(SVM: Support Vector Machine) 및 XGBoost(Extreme Gradient Boosting)에 기반한 데이터 모델들을 포함할 수 있다.

[0032] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 모델 생성 방법에 있어서, 상기 데이터 인덱스 정보는, 상기 공격 패킷의 패킷 데이터의 디코딩된 데이터의 헥사(hexa)값들을 아스키 코드의 정수값들로 변환하여 생성된 인덱스 데이터를 포함하고,

[0033] 상기 데이터 컨텐츠 카운트 벡터 정보는, 상기 디코딩된 데이터를 청킹(Chunking) 알고리즘을 이용하여 소정 개수의 문자 당 하나의 블록을 생성하여 해시값을 생성한 후 중복된 해시값을 누적카운트하여 생성된 소정 크기의 벡터값을 포함하며,

[0034] 상기 데이터 토큰 정보는, 상기 디코딩된 데이터를 URL 디코딩하여 문장으로 형성하고 형성된 문장에서 중복문자를 제거하고, 상기 형성된 문장 내의 공백 및 특수문자를 소정 바이트의 식별가능한 문자로 대체한 후 문장을 분리하여 생성된 토큰값을 포함할 수 있다.

[0035] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 모델 생성 방법에 있어서, 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들 각각을 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위로 데이터를 구분하고 해싱처리한 후 해싱 처리된 각각의 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위 데이터를 연결하여 생성된 이진화된 값을, 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 생성하기 위한 입력 벡터값으로 사용할 수 있다.

[0036] 상기 과제를 달성하기 위한 인공지능 기반 정오탐 식별 모델 생성 장치는,

[0037] 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷으로부터 각각 복수 유형의 특징들을 생성하는 가공 데이터 생성부;

[0038] 상기 복수 유형의 특징들 중 한 유형의 특징들을 클러스터링하여 복수의 클러스터를 생성하는 데이터 클러스터 생성부;

[0039] 상기 복수의 클러스터 각각에서 하나의 특징을 선택하는 데이터 선택부; 및

[0040] 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 공격 패킷의 정탐 또는 오탐을 식별하기 위한 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 생성하는 데이터 모델 생성부를 포함한다.

[0041] 상기 과제를 달성하기 위한 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법은,

[0042] (A) 인공지능 기반 정오탐 식별 장치가, 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷에 기반하여 복수의 인공지능 기반 데이터 모델을 생성하는 단계; 및

[0043] (B) 상기 복수의 인공지능 기반 데이터 모델에 기반하여, 테스트 데이터인 정탐 또는 오탐일 수 있는 복수의 공격 패킷의 정탐 또는 오탐 여부를 식별하는 단계를 포함하고,

[0044] 상기 단계 (A)는,

[0045] (A-1) 상기 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷으로부터 각각 복수 유형의 특징들을 생성하는 단계;

[0046] (A-2) 상기 복수 유형의 특징들 중 한 유형의 특징들을 클러스터링하여 복수의 클러스터를 생성하는 단계;

[0047] (A-3) 상기 복수의 클러스터 각각에서 하나의 특징을 선택하는 단계; 및

[0048] (A-4) 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 공격 패킷의 정탐 또는 오탐을 식별하기 위한 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 생성하는 단계를 포함한다.

[0049] 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법에 있어서, 상기 단계 (B)는,

[0050] (B-1) 상기 테스트 데이터인 정탐 또는 오탐일 수 있는 복수의 공격 패킷으로부터 각각 복수 유형의 특징들을 생성하는 단계;

[0051] (B-2) 상기 복수 유형의 특징들 중 한 유형의 특징들을 클러스터링하여 복수의 클러스터를 생성하는 단계;

[0052] (B-3) 상기 복수의 클러스터 각각에서 하나의 특징을 선택하는 단계; 및

[0053] (B-4) 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 사용하여 상기 복수의 클러스터 각각에서 선택된 하나의 특징에 대응하는 정탐 또는 오탐일 수 있는 공격 패킷의 정탐 또는 오탐 여부를 최종 결정하는 단계를 포함할 수 있다.

[0054] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법에 있어서, 상기 복수 유형의 특징들은 각각, 데이터 인덱스 정보, 데이터 컨텐츠 카운트 벡터 정보 및 데이터 토큰 정보를 포함할 수 있다.

[0055] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법에 있어서, 상기 단계 (A-2) 및 상기 단계

(B-2)에서, 상기 한 유형의 특징들은 상기 데이터 컨텐츠 카운트 벡터 정보를 포함할 수 있다.

[0056] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법에 있어서, 상기 단계 (A-2) 및 상기 단계 (B-2) 각각은, 상기 데이터 컨텐츠 카운트 벡터 정보들 간의 유사도에 기반하여 상기 복수의 데이터 컨텐츠 카운트 벡터 정보를 클러스터링하여 복수의 클러스터를 생성할 수 있다.

[0057] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법에 있어서, 상기 단계 (A-4)는,

[0058] (A-4-1) 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 상기 공격 패킷의 정탐 또는 오탐을 식별하기 위한 각 유형의 특징에 대한 복수의 인공지능 기반 데이터 모델을 생성하는 단계; 및

[0059] (A-4-2) 상기 각 유형의 특징에 대한 복수의 인공지능 기반 데이터 모델 중 정확도가 가장 높은 데이터 모델을 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델로 최종 선정하는 단계를 포함할 수 있다.

[0060] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법에 있어서, 상기 단계 (A-4-1)의 복수의 인공지능 기반 데이터 모델은,

[0061] 인공신경망(ANN: Artificial Neural Network), 심층 신경망(DNN: Deep Neural Network), 1D 컨벌루션 신경망(CNN1D: 1D Convolution Neural Network), 2D 컨벌루션 신경망(CNN2D: 2D Convolution Neural Network), 랜덤 포레스트(RF: Random Forest), 서포트 벡터 머신(SVM: Support Vector Machine) 및 XGBoost(Extreme Gradient Boosting)에 기반한 데이터 모델들을 포함할 수 있다.

[0062] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법에 있어서, 상기 데이터 인덱스 정보는, 상기 공격 패킷의 패킷 데이터의 디코딩된 데이터의 헥사(hexa)값들을 아스키 코드의 정수값들로 변환하여 생성된 인덱스 데이터를 포함하고,

[0063] 상기 데이터 컨텐츠 카운트 벡터 정보는, 상기 디코딩된 데이터를 청킹(Chunking) 알고리즘을 이용하여 소정 개수의 문자 당 하나의 블록을 생성하여 해시값을 생성한 후 중복된 해시값을 누적카운트하여 생성된 소정 크기의 벡터값을 포함하며,

[0064] 상기 데이터 토큰 정보는, 상기 디코딩된 데이터를 URL 디코딩하여 문장으로 형성하고 형성된 문장에서 중복문자를 제거하고, 상기 형성된 문장 내의 공백 및 특수문자를 소정 바이트의 식별가능한 문자로 대체한 후 문장을 분리하여 생성된 토큰값을 포함할 수 있다.

[0065] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법에 있어서, 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들 각각을 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위로 데이터를 구분하고 해싱처리한 후 해싱 처리된 각각의 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위 데이터를 연결하여 생성된 이진화된 값을, 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델의 입력 벡터값으로 사용할 수 있다.

[0066] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법에 있어서, 상기 단계 (B-4)는,

[0067] 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 사용하여 상기 단계 (B-3)의 상기 복수의 클러스터 각각에서 선택된 하나의 특징에 대응하는 정탐 또는 오탐일 수 있는 공격 패킷의 정탐 또는 오탐을 식별하는 단계; 및

[0068] 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델의 정오탐 식별 결과를 집계하여 다수결에 따라 상기 복수의 클러스터 각각에서 선택된 하나의 특징에 대응하는 정탐 또는 오탐일 수 있는 공격 패킷의 정탐 또는 오탐 여부를 최종 결정하는 단계를 포함할 수 있다.

[0069] 상기 과제를 달성하기 위한 인공지능 기반 정오탐 식별 장치는,

[0070] 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷에 기반하여 복수의 인공지능 기반 데이터 모델을 생성하는 정오탐 식별 모델 생성부; 및

[0071] 상기 복수의 인공지능 기반 데이터 모델에 기반하여, 테스트 데이터인 정탐 또는 오탐일 수 있는 복수의 공격 패킷의 정탐 또는 오탐 여부를 식별하는 정오탐 식별부를 포함하고,

[0072] 상기 정오탐 식별 모델 생성부는,

[0073] 상기 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷으로부터 각각 복수 유형의 특징들을 생성하는 가공 데이터 생성부;

[0074] 상기 복수 유형의 특징들 중 한 유형의 특징들을 클러스터링하여 복수의 클러스터를 생성하는 데이터 클러스터 생성부;

[0075] 상기 복수의 클러스터 각각에서 하나의 특징을 선택하는 데이터 선택부; 및

[0076] 상기 선택된 한 유형의 특징들 및 상기 선택된 한 유형의 특징들 각각에 대응하는 나머지 유형들의 특징들에 기반하여, 공격 패킷의 정탐 또는 오탐을 식별하기 위한 상기 각 유형의 특징에 대한 인공지능 기반 데이터 모델을 생성하는 데이터 모델 생성부를 포함할 수 있다.

발명의 효과

[0077] 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 모델 생성 방법과 장치 및 인공지능 기반 정오탐 식별 방법과 장치에 의하면, 네트워크 트래픽 환경에서 공격 패킷의 특성을 파악하여 사람이 아닌 인공지능(AI)/머신러닝(ML)을 통해 공격 패킷을 자동 분석하고 이를 통해 공격 패킷의 정탐 또는 오탐을 식별하고 공격을 탐지할 수 있다.

[0078] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 모델 생성 방법과 장치 및 인공지능 기반 정오탐 식별 방법과 장치에 의하면, 네트워크 트래픽 환경에서 공격 패킷의 특성을 파악하여 변종 공격을 탐지할 수 있다.

[0079] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 모델 생성 방법과 장치 및 인공지능 기반 정오탐 식별 방법과 장치에 의하면, 최적의 학습 모델을 생성하는데 소요되는 시간을 단축할 수 있다.

[0080] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 모델 생성 방법과 장치 및 인공지능 기반 정오탐 식별 방법과 장치에 의하면, 대량의 공격 패킷의 정탐 또는 오탐 여부를 식별하는데 소요되는 시간을 단축할 수 있다.

도면의 간단한 설명

[0081] 도 1은 네트워크 보안 장비로부터 발생한 이벤트 로그의 로그 항목 정규화를 위한 표준 필드 항목을 도시한 도면.

도 2는 정규화된 로그 샘플 데이터를 도시한 도면.

도 3은 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법의 전체 흐름도.

도 4는 인공지능 기반 정오탐 식별 모델을 생성하기 위한 데이터 학습 단계를 도시한 흐름도.

도 5는 인공지능 기반 정오탐 식별을 위한 데이터 모델 적용 단계를 도시한 흐름도.

도 6은 데이터 임베딩 처리의 세부 처리를 도시한 도면.

도 7은 데이터 모델을 생성하기 위한 데이터 모델링 단계의 개략도.

도 8은 정오탐 식별을 위한 데이터 모델 적용 단계의 개략도.

도 9는 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 장치를 도시한 도면.

도 10은 본 발명의 다른 실시예에 의한 인공지능 기반 정오탐 식별 장치를 도시한 도면.

발명을 실시하기 위한 구체적인 내용

[0082] 본 발명의 목적, 특정한 장점들 및 신규한 특징들은 첨부된 도면들과 연관되어지는 이하의 상세한 설명과 바람직한 실시예들로부터 더욱 명백해질 것이다.

[0083] 이에 앞서 본 명세서 및 청구범위에 사용된 용어나 단어는 통상적이고 사전적인 의미로 해석되어서는 아니되며, 발명자가 그 자신의 발명을 가장 최선의 방법으로 설명하기 위해 용어의 개념을 적절하게 정의할 수 있는 원칙에 입각하여 본 발명의 기술적 사상에 부합되는 의미와 개념으로 해석되어야 한다.

[0084] 본 명세서에서 각 도면의 구성요소들에 참조번호를 부가함에 있어서, 동일한 구성 요소들에 한해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 번호를 가지도록 하고 있음에 유의하여야 한다.

[0085] 또한, "제1", "제2", "일면", "타면" 등의 용어는, 하나의 구성요소를 다른 구성요소로부터 구별하기 위해 사용되는 것으로, 구성요소가 상기 용어들에 의해 제한되는 것은 아니다.

[0086] 이하, 본 발명을 설명함에 있어, 본 발명의 요지를 불필요하게 흐릴 수 있는 관련된 공지 기술에 대한 상세한 설명은 생략한다.

[0087] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시형태를 상세히 설명하기로 한다.

[0088] 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법 및 장치는, 정탐인 공격 패킷 및 오탐인 공격 패킷의 데이터로부터 데이터 인덱스 정보와 데이터 컨텐츠 카운트 벡터 정보 및 데이터 토큰 정보를 기준으로 특징들을 추출하고, 이를 그룹핑하고 벡터화하여 학습함으로써 데이터 학습 모델을 생성한다. 생성된 데이터 학습 모델을 통해 정탐 또는 오탐일 수 있는 공격 패킷의 정탐 또는 오탐을 분류하고, 이를 기반으로 공격을 탐지한다.

[0089] 또한, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법 및 장치는 네트워크 트래픽 환경에서 공격 패킷의 데이터로부터 데이터 인덱스 정보, 데이터 컨텐츠 카운트 벡터 정보 및 데이터 토큰 정보를 기준으로 특징을 추출하여 다중 벡터를 생성할 수 있다.

[0090] 본 발명은 네트워크 트래픽 환경에서 공격 패킷에서 추출된 다중 벡터 정보를 클러스터링하여 복수의 클러스터를 생성하고 분류하여 각 클러스터에서 하나의 특징을 선택하여 중복된 특징을 제거함으로써 각 클러스터에서 유일한 특징 벡터를 생성할 수 있다.

[0091] 본 발명은 네트워크 트래픽 환경에서 각 클러스터의 유일한 특징 벡터를 활용하여 인공지능(AI) 학습이 가능하고, 학습된 데이터 모델을 이용하여 공격 패킷의 정탐과 오탐을 식별할 수 있다.

[0092] 도 3은 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법의 전체 흐름도이다.

[0093] 도 3을 참조하면, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법은, 학습 데이터인 정탐 및 오탐으로 판정된 복수의 공격 패킷에 기반하여 복수의 인공지능 기반 데이터 모델(310)을 생성하는 단계(단계 S300) 및 상기 복수의 인공지능 기반 데이터 모델(310)에 기반하여, 테스트 데이터인 정탐 또는 오탐일 수 있는 복수의 공격 패킷의 정탐 또는 오탐 여부를 식별하는 단계(단계 S302)를 포함한다.

[0094] 참조번호 S304는 데이터 모델링 단계를 나타낸 것이고, 참조번호 S306은 데이터 모델 적용 단계를 나타낸 것이다. 이에 대해서는 추후 상세히 설명하기로 한다.

[0095] 도 9는 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 장치를 도시한 도면이고, 도 10은 본 발명의 다른 실시예에 의한 인공지능 기반 정오탐 식별 장치를 도시한 도면이다.

[0096] 우선 도 9를 참조하면, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 장치(900)는 인공지능 기반 정오탐 식별 모델 생성부(903) 및 인공지능 기반 정오탐 식별부(905)를 포함한다.

[0097] 도 10을 참조하면, 본 발명의 다른 실시예에 의한 인공지능 기반 정오탐 식별 장치(1000)는 인공지능 기반 정오탐 식별 장치(1000)의 동작을 제어하기 위한 프로세서(1002), 상기 프로세서(1002)에 메모리 제어부(1006)를 통해 연결된 메모리(1004) 및 인터페이스부(1008)를 포함한다.

[0098] 상기 프로세서(1002)는, 다양한 소프트웨어 프로그램과 메모리(1004)에 저장되어 있는 명령어 집합을 실행하여 여러 기능을 수행하고 데이터를 처리하는 기능을 수행할 수 있다.

[0099] 상기 메모리(1004)는 고속 랜덤 액세스 메모리, 하나 이상의 자기 디스크 저장 장치, 플래시 메모리 장치와 같은 불휘발성 메모리 등을 포함할 수 있다. 또한, 메모리(1004)는 프로세서(1002)로부터 떨어져 위치하는 저장장치나, 인터넷 등의 통신 네트워크를 통하여 액세스되는 네트워크 부착형 저장장치 등을 더 포함할 수 있다.

[0100] 상기 메모리(1004)는, 상기 프로세서(1002)에 의해 실행되도록 구성되는 하나 이상의 모듈을 포함하는데, 상기 하나 이상의 모듈은, 인공지능 기반 정오탐 식별 장치(1000)의 전반적인 동작을 제어하기 위한 명령어들을 포함하는 운영제체(1010), 인공지능 기반 정오탐 식별 모델 생성 모듈(1012) 및 인공지능 기반 정오탐 식별 모듈(1014)을 포함한다.

[0101] 상기 인터페이스부(1008)는 통신 네트워크(미도시)를 통해 학습 데이터 및 테스트 데이터를 수집하고 최종 정오탐 결과를 출력할 수 있으며, 입출력 주변 장치를 프로세서(1002) 또는 메모리(1004)에 연결할 수 있고, 메모리 제어부(1006)는 프로세서(1002)나 인터페이스부(1008)가 메모리(1004)에 접근하는 경우에, 메모리 액세스를 제어하는 기능을 수행할 수 있다. 실시예에 따라서는, 프로세서(1002), 메모리(1004), 메모리 제어부(1006) 및 인터페이스부(1008)를 단일 칩 상에 구현하거나, 별개의 칩으로 구현할 수 있다.

[0102] 도 3에 도시된 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법은 도 9에 도시된 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 장치(900) 또는 도 10에 도시된 본 발명의 다른 실시예에 의한 인공지능 기반 정오탐 식별 장치(1000)에 의해 수행될 수 있다.

[0103] 또한, 도 3에 도시된 데이터 모델링 단계(단계 S304)는, 인공지능 기반 정오탐 식별 모델 생성부(903)에 의해 수행될 수 있거나, 도 10에 도시된 프로세서(1002)가 메모리(1004)에 저장된 인공지능 기반 정오탐 식별 모델 생성 모듈(1012)을 실행함으로써 수행될 수 있다.

[0104] 또한, 도 3에 도시된 데이터 모델 적용 단계(단계 S306)는, 인공지능 기반 정오탐 식별부(905)에 의해 수행될 수 있거나, 도 10에 도시된 프로세서(1002)가 메모리(1004)에 저장된 인공지능 기반 정오탐 식별 모듈(1014)을 실행함으로써 수행될 수 있다.

[0105] 도 3에 도시된 데이터 모델링 단계(단계 S304)에 대해 상세히 설명하기로 한다.

[0106] 도 4는 도 3에 도시된 데이터 모델링 단계(단계 S304)의 상세 흐름도로서, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법에서 인공지능 기반 정오탐 식별 모델을 생성하기 위한 데이터 학습 단계를 도시한 흐름도이다.

[0107] 단계 S400에서, 제1 데이터 정규화부(901)는, 정탐 및 오탐 학습 데이터로서 정탐으로 판정된 공격 패킷 및 오탐으로 판정된 공격 패킷과 관련된 이벤트 로그를 수집하고, 수집된 이벤트 로그의 로그 데이터를 정규화한다.

[0108] 네트워크 보안 장비(IPS, IDS, DDX, FW, Server, System 등)로부터 발생한 이벤트 로그를 도 1에 도시된 로그 항목 정규화를 위한 표준 필드 항목에 따라 정규화하여 중복 필드를 제거하고 최적화된 특징 후보군을 선정한다. 도 2는 이벤트 로그를 정규화하여 생성된 정규화된 로그 샘플 데이터를 도시한 것이다.

[0109] 도 2에서 참조번호 200은 공격 패킷의 패킷 데이터를 나타낸 것이다.

[0110] 단계 S402에서, 제1 데이터 전처리부(902)는, 정규화 로그 데이터를 전처리한다.

[0111] 제1 데이터 전처리부(902)는 도 2에 도시된 정규화 로그 데이터에서 패킷 데이터(200)(B64)를 추출한다.

[0112] 제1 데이터 전처리부(902)는 패킷 데이터(200)의 데이터가 base64 인코딩된 데이터인지, 바이너리(Binary) 데이터인지, 스트링(String) 데이터인지 분류하고, 유형별 데이터를 디코딩하여 디코딩 데이터를 생성한다.

[0113] 단계 S404에서, 제1 가공 데이터 생성부(904)는, 디코딩 데이터에 기반하여 제1 유형 내지 제3 유형의 가공 데이터를 생성한다.

[0114] 제1 가공 데이터 생성 모듈(904_1)은 디코딩 데이터의 헥사(hexa)값들을 아스키 코드의 정수값들로 변환하여 제1 유형의 가공 데이터인 인덱스 데이터 정보를 생성한다. 인덱스 데이터 정보는 디코딩 데이터의 헥사값들을 아스키 코드의 정수값들로 변환한 것이므로, 원본 데이터의 특징을 가지고 있다.

[0115] 제2 가공 데이터 생성 모듈(904_2)은, 디코딩된 데이터를 청킹(Chunking) 알고리즘을 이용하여 윈도우 크기를 4로 설정하여 문자 4개씩 하나의 블록을 생성하여 해시값을 생성한 후 중복된 해시값을 누적카운트하여 제2 유형의 가공 데이터인 크기 512 바이트의 데이터 컨텐츠 카운트 벡터 정보를 생성한다. 데이터 컨텐츠 카운트 벡터 정보는 원본 데이터가 4 바이트씩 변형된 것으로, 변형된 공격 패킷, 즉 변종 공격을 탐지하기 위한 것이다.

[0116] 제3 가공 데이터 생성 모듈(904_3)은, 디코딩된 데이터를 URL 디코딩하여 문장으로 형성하고 생성된 문장에서 중복문자를 제거하고, 상기 생성된 문자 내의 공백 및 특수문자를 최대 5 바이트의 식별가능한 문자로 대체한 후 문장을 분리하여 제3 유형의 가공 데이터인 데이터 토큰 정보를 생성한다. 데이터 토큰 정보는 공격 패킷의 패킷 데이터를 문장으로 만들었을 때의 특징을 가지고 있다.

[0117] 단계 S406에서, 제1 데이터 클러스터 생성부(906)는, 데이터 컨텐츠 카운트 벡터 정보들 간의 유사도(예를 들어, 코사인 유사도)에 기반하여 복수의 데이터 컨텐츠 카운트 벡터 정보를 클러스터링하여 복수의 데이터 클러스터를 생성한다.

[0118] 예를 들어, 데이터 컨텐츠 카운트 벡터 정보들의 유사도가 소정값 이상인 경우, 하나의 클러스터로 구성함으로써, 복수의 데이터 컨텐츠 카운트 벡터 정보들을 을 클러스터링하여 복수의 클러스터를 생성한다.

[0119] 단계 S408에서, 제1 데이터 선택부(908)는, 복수의 클러스터 각각에서 하나의 데이터 컨텐츠 카운트 벡터 정보를 선택하고, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보에 대응하는 데이터 인덱스 정보, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보 및 데이터 토큰 정보를 출력한다.

[0120] 본 발명의 일 실시예에 의하면, 정탐으로 판정된 대량의 공격 패킷들 및 오탐으로 판정된 대량의 공격 패킷들 모두를 학습 데이터로서 사용하여 데이터 모델을 생성하는 것이 아니라, 한 유형의 특징의 유사도가 매우 높은 공격 패킷들을 하나의 클러스터로 클러스터링하고, 각 클러스터에서 하나의 공격 패킷만을 선택하여 선택된 공격 패킷의 특징들을 학습 데이터로서 사용하기 때문에, 최적의 데이터 학습 모델을 생성하는 데 소요되는 시간을 대폭 단축할 수 있다.

[0121] 단계 S410에서, 제1 가공 데이터 임베딩 처리부(910)는, 도 6에 도시된 바와 같이, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보에 대응하는 데이터 인덱스 정보, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보 및 상기 선택된 데이터 컨텐츠 카운트 벡터 정보에 대응하는 데이터 토큰 정보 각각을 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위로 데이터를 구분하고 해싱처리한 후 해싱 처리된 각각의 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위 데이터를 연결하여 생성된 이진화된 총 12288 바이트(4096×3)의 입력 벡터값을 생성한다.

[0122] 제1 가공 데이터 임베딩 처리 모듈(910_1)은, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보에 대응하는 데이터 인덱스 정보를 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위로 데이터를 구분하고 해싱처리한 후 해싱 처리된 각각의 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위 데이터를 연결하여 생성된 이진화된 총 12288 바이트(4096×3)의 입력 벡터값을 생성한다.

[0123] 제2 가공 데이터 임베딩 처리 모듈(910_2)은, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보를 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위로 데이터를 구분하고 해싱처리한 후 해싱 처리된 각각의 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위 데이터를 연결하여 생성된 이진화된 총 12288 바이트(4096×3)의 입력 벡터값을 생성한다.

[0124] 제3 가공 데이터 임베딩 처리 모듈(910_3)은, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보에 대응하는 데이터 토큰 정보를 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위로 데이터를 구분하고 해싱처리한 후 해싱 처리된 각각의 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위 데이터를 연결하여 생성된 이진화된 총 12288 바이트(4096×3)의 입력 벡터값을 생성한다.

[0125] 단계 S412에서, 데이터 모델 생성부(912)는, 도 7에 도시된 바와 같이, 데이터 인덱스 정보, 데이터 컨텐츠 카운트 벡터 정보 및 데이터 토큰 정보의 입력 벡터값이 인공지능 및 머신러닝 알고리즘의 입력 정보로 사용되어, 복수의 데이터 모델을 생성한다.

[0126] 제1 데이터 모델 생성 모듈(912_1)은 데이터 인덱스 정보의 입력 벡터값에 기반하여 복수의 인공지능 기반 데이터 모델을 생성한다.

[0127] 제2 데이터 모델 생성 모듈(912_2)은 상기 선택된 데이터 컨텐츠 카운트 벡터 정보의 입력 벡터값에 기반하여 복수의 인공지능 기반 데이터 모델을 생성한다.

[0128] 제3 데이터 모델 생성 모듈(912_3)은 데이터 토큰 정보의 입력 벡터값에 기반하여 복수의 인공지능 기반 데이터 모델을 생성한다.

[0129] 상기 복수의 인공지능 기반 데이터 모델은, 인공신경망(ANN: Artificial Neural Network), 심층 신경망(DNN: Deep Neural Network), 1D 컨벌루션 신경망(CNN1D: 1D Convolution Neural Network), 2D 컨벌루션 신경망(CNN2D: 2D Convolution Neural Network), 랜덤 포레스트(RF: Random Forest), 서포트 벡터 머신(SVM: Support Vector Machine) 및 XGBoost(Extreme Gradient Boosting)에 기반한 데이터 모델들을 포함할 수 있다.

[0130] 단계 S414에서, 데이터 모델 선정부(914)는, 데이터 인덱스 정보, 데이터 컨텐츠 카운트 벡터 정보 및 데이터 토큰 정보 각각에 대한 복수의 데이터 모델 중 가장 정확도가 높은 데이터 모델을 선정한다.

[0131] 제1 데이터 모델 선정 모듈(914_1)은, 데이터 인덱스 정보에 대한 복수의 데이터 모델 중 가장 정확도가 높은 데이터 모델을 선정한다.

[0132] 제2 데이터 모델 선정 모듈(914_2)은, 데이터 컨텐츠 카운트 벡터 정보에 대한 복수의 데이터 모델 중 가장 정

확도가 높은 데이터 모델을 선정한다.

[0133] 제3 데이터 모델 선정 모듈(914_3)은, 데이터 토큰 정보에 대한 복수의 데이터 모델 중 가장 정확도가 높은 데이터 모델을 선정한다.

[0134] 도 3에 도시된 데이터 모델 적용 단계(단계 S306)에 대해 상세히 설명하기로 한다.

[0135] 도 5는 도 3에 도시된 데이터 모델 적용 단계(단계 S306)의 상세 흐름도로서, 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 방법에서 인공지능 기반 정오탐 식별 모델에 기반하여, 테스트 데이터인 정탐 또는 오탐일 수 있는 복수의 공격 패킷의 정탐 또는 오탐 여부를 식별하기 위한 데이터 모델 적용 단계를 도시한 흐름도이다.

[0136] 도 5에 도시된 인공지능 기반 정오탐 식별을 위한 데이터 모델 적용 단계에서는, 상기 데이터 인덱스 정보에 대해 선정된 데이터 모델, 상기 데이터 컨텐츠 카운트 벡터 정보에 대해 선정된 데이터 모델 및 상기 데이터 토큰 정보에 대해 선정된 데이터 모델을 적용하여, 테스트 데이터인 정탐 또는 오탐일 수 있는 공격 패킷의 정탐 또는 오탐을 최종적으로 결정한다.

[0137] 단계 S500에서, 제2 데이터 정규화부(916)는, 정탐 및 오탐 테스트 데이터로서 정탐 또는 오탐일 수 있는 공격 패킷과 관련된 이벤트 로그를 수집하고, 수집된 이벤트 로그의 로그 데이터를 정규화한다.

[0138] 네트워크 보안 장비(IPS, IDS, DDX, FW, Server, System 등)로부터 발생한 이벤트 로그를 도 1에 도시된 로그 항목 정규화를 위한 표준 필드 항목에 따라 정규화하여 중복 필드를 제거하고 최적화된 특징 후보군을 선정한다. 도 2는 이벤트 로그를 정규화하여 생성된 정규화된 로그 샘플 데이터를 도시한 것이다.

[0139] 도 2에서 참조번호 200은 공격 패킷의 패킷 데이터를 나타낸 것이다.

[0140] 단계 S502에서, 제2 데이터 전처리부(918)는, 정규화 로그 데이터를 전처리한다.

[0141] 제2 데이터 전처리부(916)는 도 2에 도시된 정규화 로그 데이터에서 패킷 데이터(200)(B64)를 추출한다.

[0142] 제2 데이터 전처리부(916)는 패킷 데이터(200)의 데이터가 base64 인코딩된 데이터인지, 바이너리(Binary) 데이터인지, 스트링(String) 데이터인지 분류하고, 유형별 데이터를 디코딩하여 디코딩 데이터를 생성한다.

[0143] 단계 S504에서, 제2 가공 데이터 생성부(920)는, 디코딩 데이터에 기반하여 제1 유형 내지 제3 유형의 가공 데이터를 생성한다.

[0144] 제4 가공 데이터 생성 모듈(920_1)은 디코딩 데이터의 헥사(hexa)값들을 아스키 코드의 정수값들로 변환하여 제1 유형의 가공 데이터인 인덱스 데이터 정보를 생성한다. 인덱스 데이터 정보는 디코딩 데이터의 헥사값들을 아스키 코드의 정수값들로 변환한 것이므로, 원본 데이터의 특징을 가지고 있다.

[0145] 제5 가공 데이터 생성 모듈(920_2)은, 디코딩된 데이터를 청킹(Chunking) 알고리즘을 이용하여 윈도우 크기를 4로 설정하여 문자 4개씩 하나의 블록을 생성하여 해시값을 생성한 후 중복된 해시값을 누적카운트하여 제2 유형의 가공 데이터인 크기 512 바이트의 데이터 컨텐츠 카운트 벡터 정보를 생성한다. 데이터 컨텐츠 카운트 벡터 정보는 원본 데이터가 4 바이트씩 변형된 것으로, 변형된 공격 패킷, 즉 변종 공격을 탐지하기 위한 것이다.

[0146] 제6 가공 데이터 생성 모듈(920_3)은, 디코딩된 데이터를 URL 디코딩하여 문장으로 형성하고 형성된 문장에서 중복문자를 제거하고, 상기 형성된 문장 내의 공백 및 특수문자를 최대 5 바이트의 식별가능한 문자로 대체한 후 문장을 분리하여 제3 유형의 가공 데이터인 데이터 토큰 정보를 생성한다. 데이터 토큰 정보는 공격 패킷의 패킷 데이터를 문장으로 만들었을 때의 특징을 가지고 있다.

[0147] 단계 S506에서, 제2 데이터 클러스터 생성부(922)는, 데이터 컨텐츠 카운트 벡터 정보들 간의 유사도(예를 들어, 코사인 유사도)에 기반하여 복수의 데이터 컨텐츠 카운트 벡터 정보를 클러스터링하여 복수의 데이터 클러스터를 생성한다.

[0148] 예를 들어, 데이터 컨텐츠 카운트 벡터 정보들의 유사도가 소정값 이상인 경우, 하나의 클러스터로 구성함으로써, 복수의 데이터 컨텐츠 카운트 벡터 정보들을 을 클러스터링하여 복수의 클러스터를 생성한다.

[0149] 단계 S508에서, 제2 데이터 선택부(924)는, 복수의 클러스터 각각에서 하나의 데이터 컨텐츠 카운트 벡터 정보를 선택하고, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보에 대응하는 데이터 인덱스 정보, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보 및 데이터 토큰 정보를 출력한다.

[0150] 본 발명의 일 실시예에 의하면, 정탐일 수 있는 공격 패킷 및 오탐일 수 있는 대량의 공격 패킷 모두를 테스트 데이터로서 사용하여 정오탐을 식별하는 것이 아니라, 한 유형의 특징의 유사도가 매우 높은 공격 패킷들을 하나의 클러스터로 클러스터링하고, 각 클러스터에서 하나의 공격 패킷만을 선택하여 선택된 공격 패킷의 특징들을 테스트 데이터로서 사용하기 때문에, 테스트 데이터인 대량의 공격 패킷의 정탐 또는 오탐 여부를 식별하는 데 소요되는 시간을 대폭 단축할 수 있다.

[0151] 단계 S510에서, 제2 가공 데이터 임베딩 처리부(926)는, 도 6에 도시된 바와 같이, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보에 대응하는 데이터 인덱스 정보, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보 및 상기 선택된 데이터 컨텐츠 카운트 벡터 정보에 대응하는 데이터 토큰 정보들 각각을 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위로 데이터를 구분하고 해싱처리한 후 해싱 처리된 각각의 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위 데이터를 연결하여 생성된 이진화된 총 12288 바이트(4096×3)의 입력 벡터값을 생성한다.

[0152] 제4 가공 데이터 임베딩 처리 모듈(926_1)은, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보에 대응하는 데이터 인덱스 정보를 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위로 데이터를 구분하고 해싱처리한 후 해싱 처리된 각각의 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위 데이터를 연결하여 생성된 이진화된 총 12288 바이트(4096×3)의 입력 벡터값을 생성한다.

[0153] 제5 가공 데이터 임베딩 처리 모듈(926_2)은, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보를 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위로 데이터를 구분하고 해싱처리한 후 해싱 처리된 각각의 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위 데이터를 연결하여 생성된 이진화된 총 12288 바이트(4096×3)의 입력 벡터값을 생성한다.

[0154] 제6 가공 데이터 임베딩 처리 모듈(926_3)은, 상기 선택된 데이터 컨텐츠 카운트 벡터 정보에 대응하는 데이터 토큰 정보를 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위로 데이터를 구분하고 해싱처리한 후 해싱 처리된 각각의 3-그램(gram), 4-그램(gram) 및 5-그램(gram) 단위 데이터를 연결하여 생성된 이진화된 총 12288 바이트(4096×3)의 입력 벡터값을 생성한다.

[0155] 단계 S512에서, 정오탐 식별부(928)는, 도 8에 도시된 바와 같이, 데이터 인덱스 정보, 데이터 컨텐츠 카운트 벡터 정보 및 데이터 토큰 정보 각각의 데이터 모델에 기반하여, 공격 패킷의 정오탐을 식별한다(도 8에서 단계 S800, S802, S804).

[0156] 제1 정오탐 식별 모듈(928_1)은, 데이터 인덱스 정보의 입력 벡터값을 입력으로 하여, 데이터 인덱스 정보의 데이터 모델에 기반하여, 공격 패킷의 정오탐을 식별한다.

[0157] 제2 정오탐 식별 모듈(928_2)은, 데이터 컨텐츠 카운트 벡터 정보의 입력 벡터값을 입력으로 하여, 데이터 컨텐츠 카운트 벡터 정보의 데이터 모델에 기반하여, 공격 패킷의 정오탐을 식별한다.

[0158] 제3 정오탐 식별 모듈(928_3)은, 데이터 토큰 정보의 입력 벡터값을 입력으로 하여, 데이터 토큰 정보의 데이터 모델에 기반하여, 공격 패킷의 정오탐을 식별한다.

[0159] 정오탐 결정부(930)는, 제1 정오탐 식별 모듈(928_1)의 정오탐 결과, 제2 정오탐 식별 모듈(928_2)의 정오탐 결과 및 제3 정오탐 식별 모듈(928_3)의 정오탐 결과를 집계하여 다수결에 의하여 공격 패킷의 정탐 또는 오탐 여부를 식별하여 최종 정오탐 결과를 출력한다(도 8에서 단계 S806).

[0160] 이상 본 발명을 구체적인 실시예를 통하여 상세하게 설명하였으나, 이는 본 발명을 구체적으로 설명하기 위한 것으로, 본 발명은 이에 한정되지 않으며, 본 발명의 기술적 사상 내에서 당 분야의 통상의 지식을 가진 자에 의해 그 변형이나 개량이 가능함은 명백하다고 할 것이다.

[0161] 본 발명의 단순한 변형 내지 변경은 모두 본 발명의 영역에 속하는 것으로, 본 발명의 구체적인 보호 범위는 첨부된 청구범위에 의하여 명확해질 것이다.

부호의 설명

[0162] 200: 공격 패킷의 패킷 데이터

900: 본 발명의 일 실시예에 의한 인공지능 기반 정오탐 식별 장치

901: 제1 데이터 정규화부 902: 제1 데이터 전처리부

903: 인공지능 기반 정오탐 식별 모델 생성부

904: 제1 가공 데이터 생성부

904_1 내지 904_3: 제1 내지 제3 가공 데이터 생성 모듈

905: 인공지능 기반 정오탐 식별부 906: 제1 데이터 클러스터 생성부

908: 제1 데이터 선택부

910: 제1 가공 데이터 임베딩 처리부

910_1 내지 910_3: 제1 내지 제3 가공 데이터 임베딩 처리 모듈

912: 데이터 모델 생성부

912_1 내지 912_3: 제1 내지 제3 데이터 모델 생성 모듈

914: 데이터 모델 선정부

914_1 내지 914_3: 제1 내지 제3 데이터 모델 선정 모듈

916: 제2 데이터 정규화부 918: 제2 데이터 전처리부

920: 제2 가공 데이터 생성부

920_1 내지 920_3: 제4 내지 제6 가공 데이터 생성 모듈

922: 제2 데이터 클러스터 생성부

924: 제2 데이터 선택부

926: 제2 가공 데이터 임베딩 처리부

926_1 내지 926_3: 제1 내지 제3 가공 데이터 임베딩 처리 모듈

928: 정오탐 식별부

928_1 내지 928_3: 제1 내지 제3 정오탐 식별 모듈

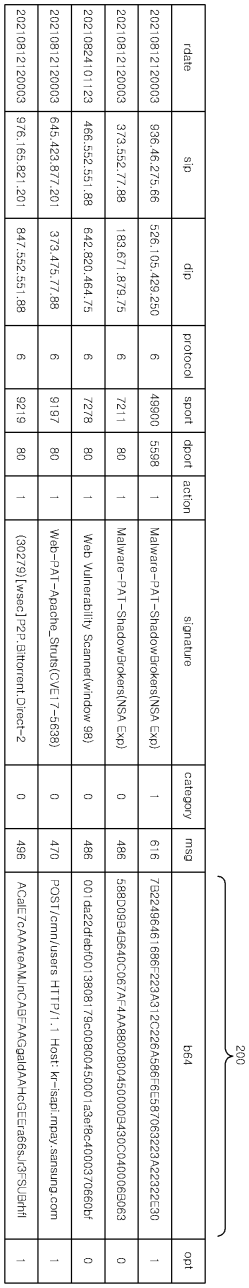
930: 정오탐 결정부

도면

도면1

필드명	필드정보	설명
Rdate	수집시간	이벤트가 발생되어 수집된 시간
Sip	출발지주소	출발지 IP 주소
Dip	목적지주소	목적지 IP 주소
Protocol	프로토콜	통신프로토콜 (TCP:6, UDP:17,ICMP:1)
Sport	출발지포트	출발지 통신 Port
Dport	목적지포트	목적지 통신 Port
action	조치방법	탐지로그 조치방법(탐지:1, 차단:2)
signature	탐지명	탐지명/공격명
category	공격유형	공격유형/공격분류
msg	공격설명	공격 설명
B64	패킷데이터	패킷정보(Base64인코딩데이터, Binary데이터, String 데이터 등)
opt	옵션	공격특이사항 및 추가 정보

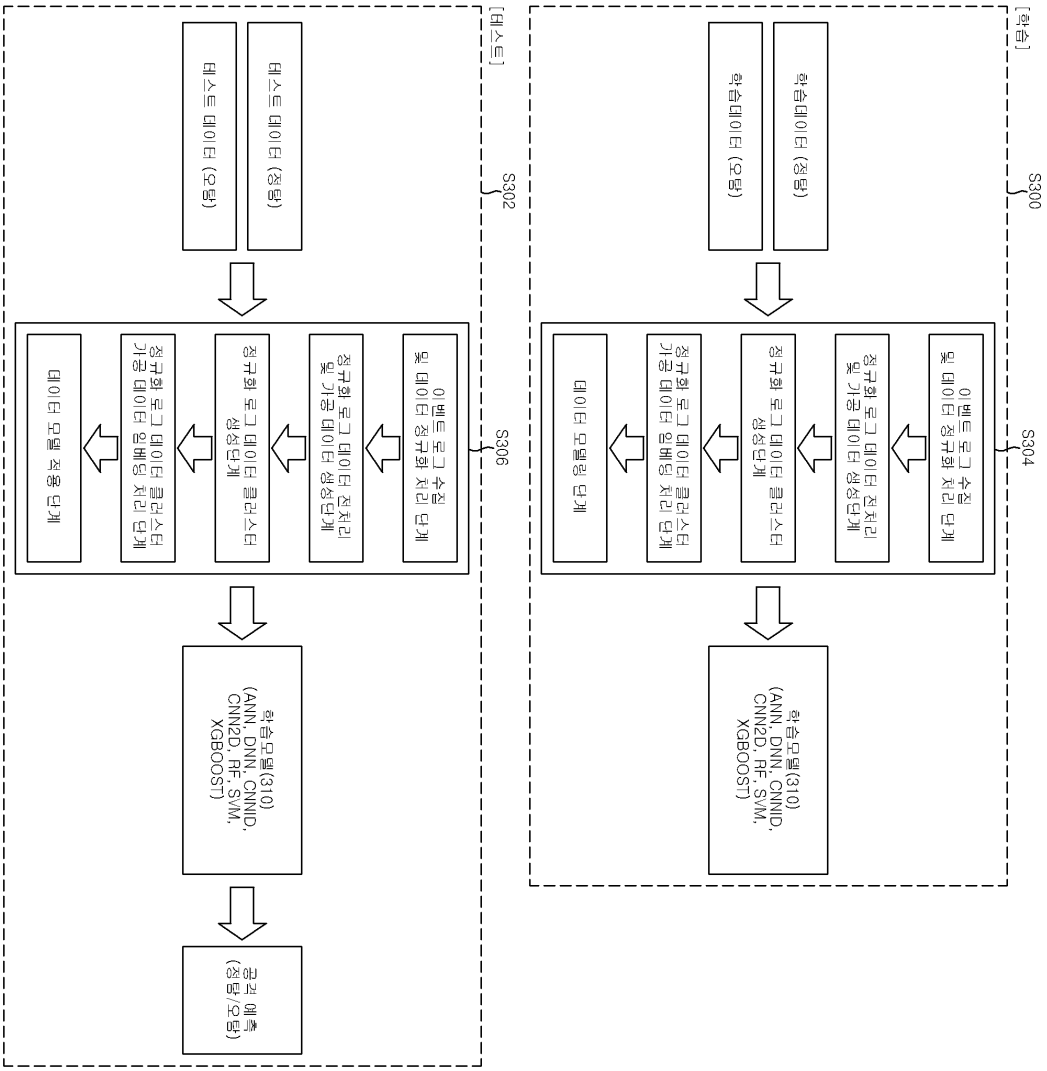
도면2



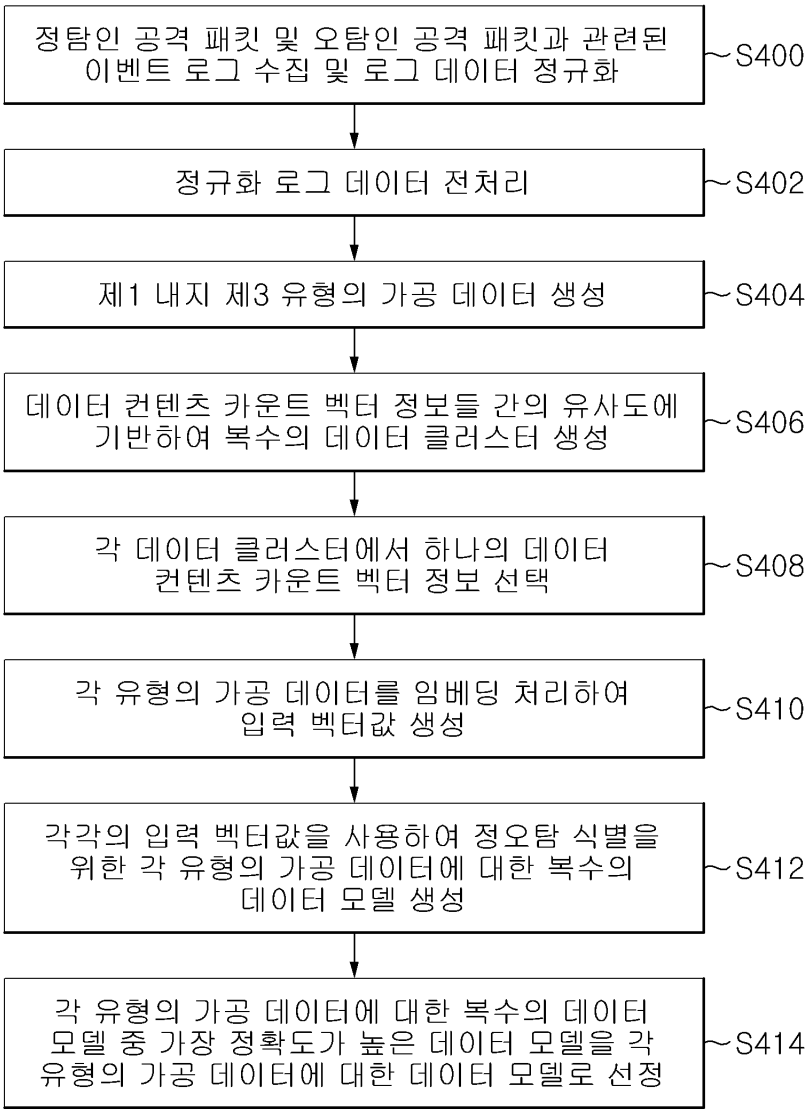
rdate	sip	dip	protocol	sport	dport	action	signature	category	msg	b64	opt
20210812120003	936.46.275.66	526.105.429.250	6	49900	5598	1	Malware-PAT-ShadowBrokers(NSA Exp)	1	616	732249646:686:223A3:2C226A586:6E587/063223A223225C30	1
20210812120003	373.552.77.88	183.67.1.879.75	6	7211	80	1	Malware-PAT-ShadowBrokers(NSA Exp)	0	486	588D0954B640C067AF4AA8800800450000934300C0400063063	0
20210824101123	466.552.551.88	642.820.464.75	6	7278	80	1	Web Vulnerability Scanner(window 98)	0	486	001da22dfcbf001380811790080045000 1a3ef8c400037068 0bf	0
20210812120003	645.423.877.201	373.475.77.88	6	9197	80	1	Web-PAT-Apache_Struts(CVE17-5638)	0	470	POST/cmm/users HTTP/1.1 Host kr-isapi.mpay.samsung.com	1
20210812120003	976.165.821.201	847.552.551.88	6	9219	80	1	(302791[wsoc] P2P.BHtorrent.Direct-2	0	496	ACalE7cAAAreAMJnCA5FAAGgaIdAAHcGEEra66sJr3FSUbfhfl	1

200

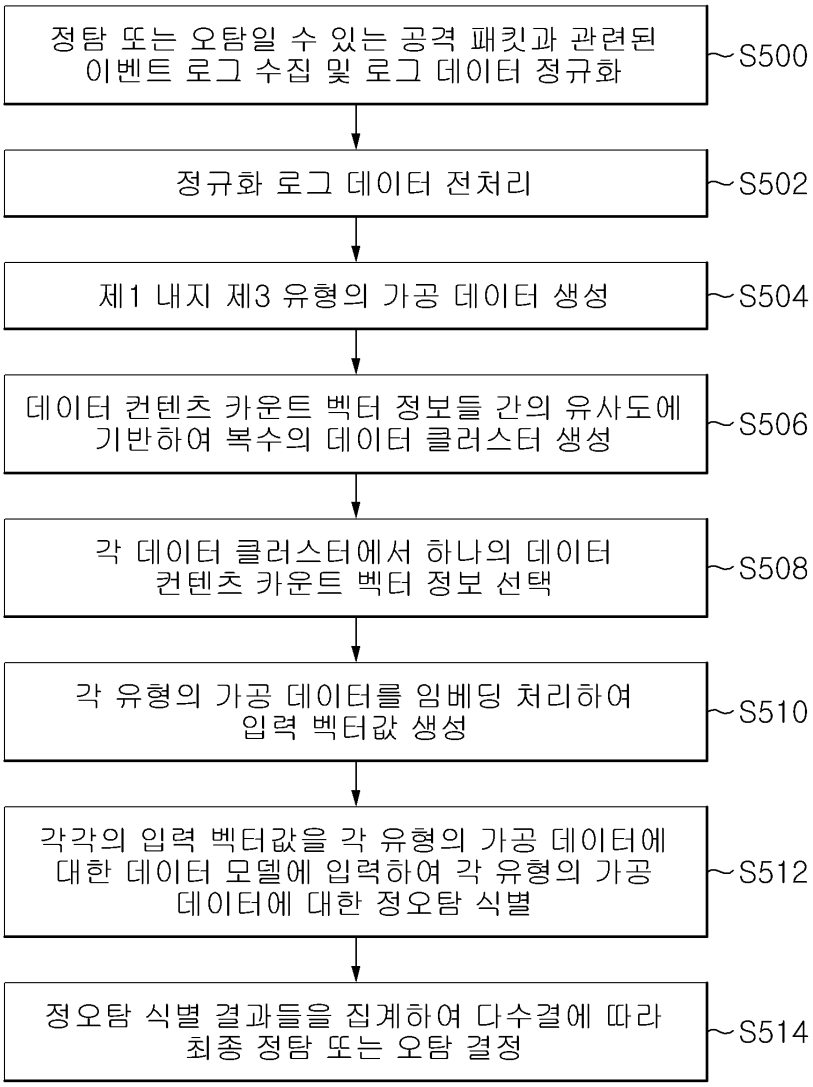
도면3


[학습]
S300
학습데이터 (정탑)
학습데이터 (오탑)
S304
이벤트 로그 수집 및 데이터 정규화 처리 단계
정규화 로그 데이터 전처리 및 기공 데이터 생성단계
정규화 로그 데이터 클러스터 생성 단계
정규화 로그 데이터 클러스터 기공 데이터 임베딩 처리 단계
데이터 모델링 단계
학습모델(310)
(ANN, DNN, CNNID, CNN2D, RF, SVM, XGBOOST)
[테스트]
S302
테스트 데이터 (정탑)
테스트 데이터 (오탑)
S306
이벤트 로그 수집 및 데이터 정규화 처리 단계
정규화 로그 데이터 전처리 및 기공 데이터 생성단계
정규화 로그 데이터 클러스터 생성 단계
정규화 로그 데이터 클러스터 기공 데이터 임베딩 처리 단계
데이터 모델 적용 단계
학습모델(310)
(ANN, DNN, CNNID, CNN2D, RF, SVM, XGBOOST)
공격 예측 (정탑/오탑)

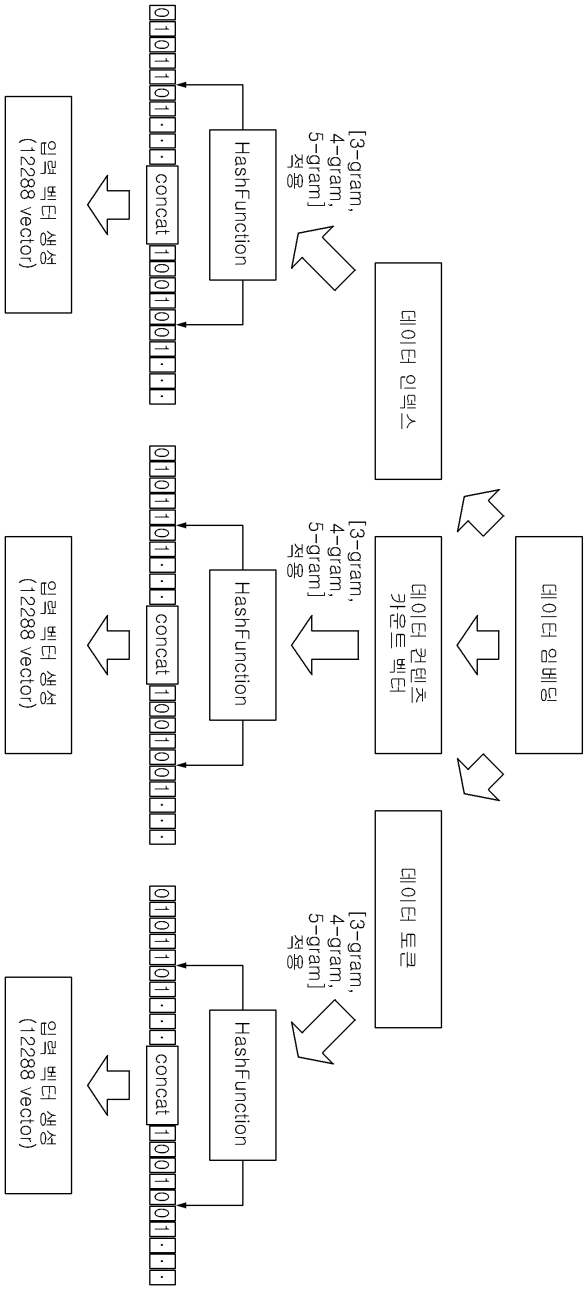
도면4


정탐인 공격 패킷 및 오탐인 공격 패킷과 관련된 이벤트 로그 수집 및 로그 데이터 정규화
S400
정규화 로그 데이터 전처리
S402
제1 내지 제3 유형의 가공 데이터 생성
S404
데이터 컨텐츠 카운트 벡터 정보들 간의 유사도에 기반하여 복수의 데이터 클러스터 생성
S406
각 데이터 클러스터에서 하나의 데이터 컨텐츠 카운트 벡터 정보 선택
S408
각 유형의 가공 데이터를 임베딩 처리하여 입력 벡터값 생성
S410
각각의 입력 벡터값을 사용하여 정오탐 식별을 위한 각 유형의 가공 데이터에 대한 복수의 데이터 모델 생성
S412
각 유형의 가공 데이터에 대한 복수의 데이터 모델 중 가장 정확도가 높은 데이터 모델을 각 유형의 가공 데이터에 대한 데이터 모델로 선정
S414

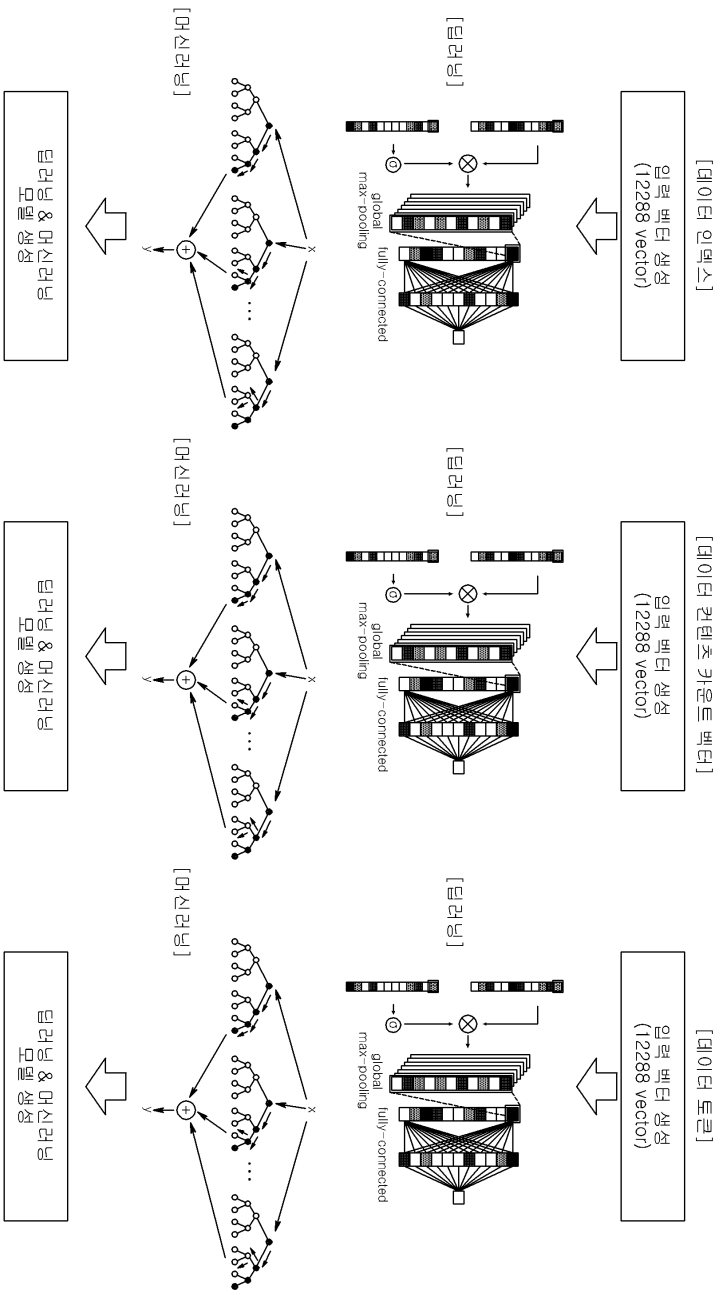
도면5


정탐 또는 오탐일 수 있는 공격 패킷과 관련된 이벤트 로그 수집 및 로그 데이터 정규화
S500
정규화 로그 데이터 전처리
S502
제1 내지 제3 유형의 가공 데이터 생성
S504
데이터 컨텐츠 카운트 벡터 정보들 간의 유사도에 기반하여 복수의 데이터 클러스터 생성
S506
각 데이터 클러스터에서 하나의 데이터 컨텐츠 카운트 벡터 정보 선택
S508
각 유형의 가공 데이터를 임베딩 처리하여 입력 벡터값 생성
S510
각각의 입력 벡터값을 각 유형의 가공 데이터에 대한 데이터 모델에 입력하여 각 유형의 가공 데이터에 대한 정오탐 식별
S512
정오탐 식별 결과들을 집계하여 다수결에 따라 최종 정탐 또는 오탐 결정
S514

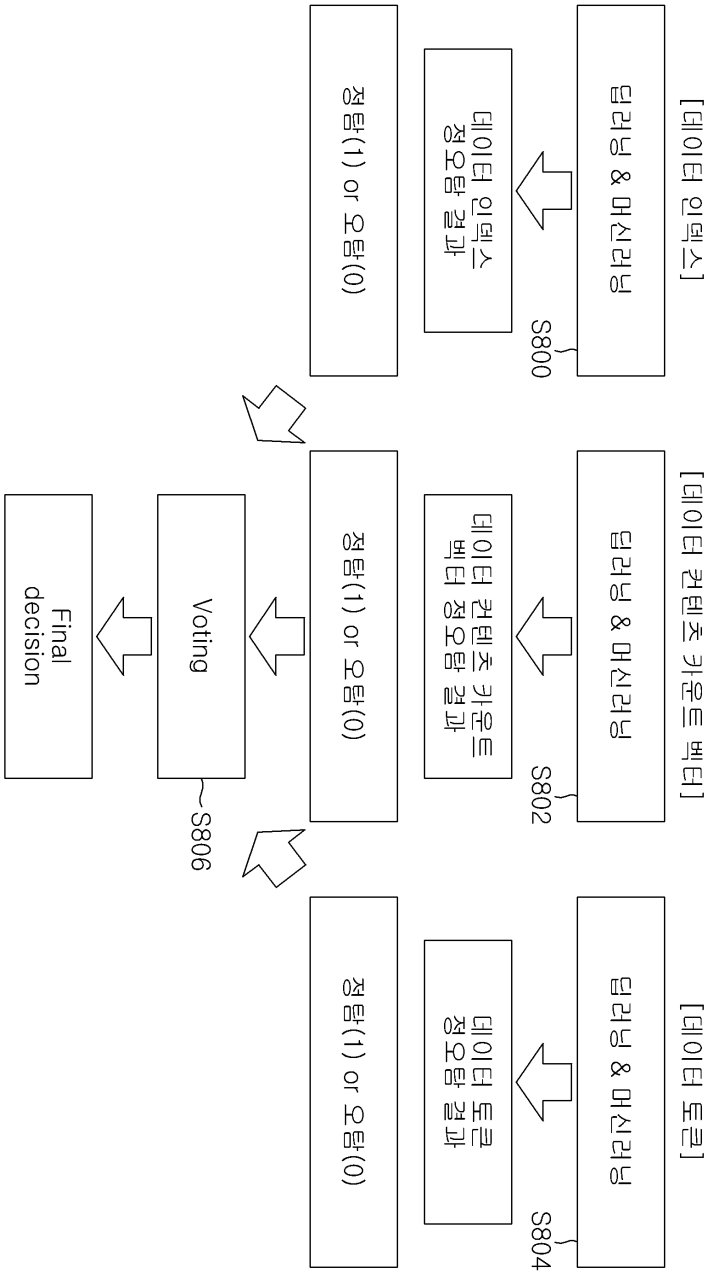
도면6


데이터 임베딩
데이터 인덱스
데이터 컨텐츠 카운트 벡터
데이터 토큰
[3-gram, 4-gram, 5-gram] 적용
[3-gram, 4-gram, 5-gram] 적용
[3-gram, 4-gram, 5-gram] 적용
HashFunction
HashFunction
HashFunction
concat
concat
concat
입력 벡터 생성 (12288 vector)
입력 벡터 생성 (12288 vector)
입력 벡터 생성 (12288 vector)

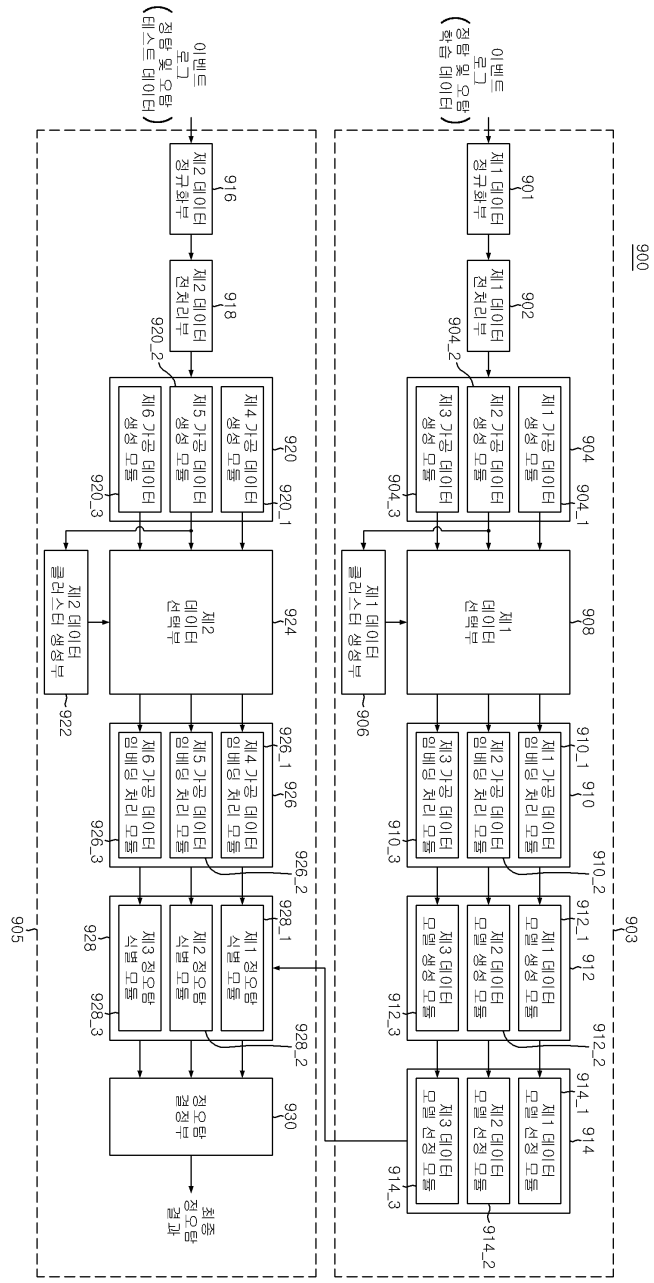
도면7


[데이터 인덱스]
[데이터 컨텐츠 카운트 벡터]
[데이터 토큰]
입력 벡터 샘플
(12288 vector)
[임베딩]
fully-connected
global
max-pooling
[머신러닝]
x
y
딥러닝 & 머신러닝
융합 출력 샘플

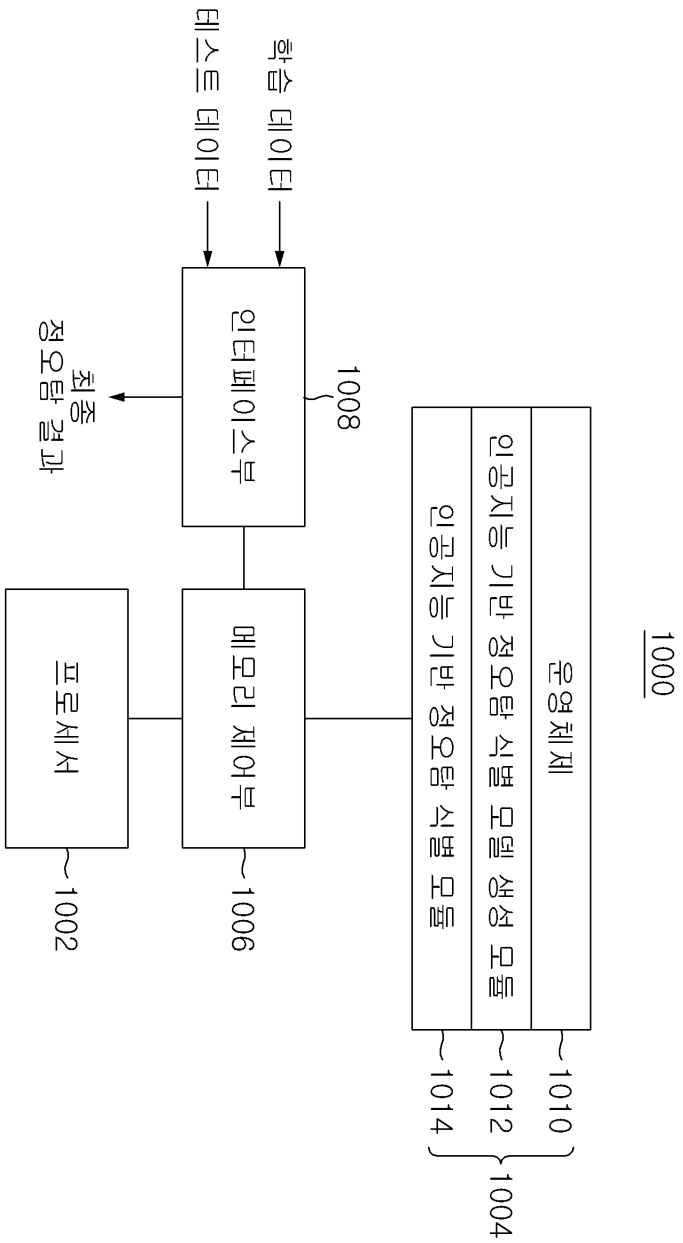
도면8


[데이터 인덱스]
딥러닝 & 머신러닝
S800
데이터 인덱스 정오탑 결과
정탑(1) or 오탑(0)
[데이터 컨텐츠 카운트 벡터]
딥러닝 & 머신러닝
S802
데이터 컨텐츠 카운트 벡터 정오탑 결과
정탑(1) or 오탑(0)
[데이터 토큰]
딥러닝 & 머신러닝
S804
데이터 토큰 정오탑 결과
정탑(1) or 오탑(0)
Voting
S806
Final decision

도면9


900
903
905
이벤트 로그 (정탐 및 오탐 혹은 데이터)
이벤트 로그 (정탐및오탐 테스트 데이터)
제1 데이터 정규화부
901
제1 데이터 전처리부
902
제1 가공 데이터 생성 모듈
904_1
904
제2 가공 데이터 생성 모듈
904_2
제3 가공 데이터 생성 모듈
904_3
제1 데이터 클러스터 생성부
906
제1 데이터 선택부
908
제1 가공 데이터 임베딩 처리 모듈
910_1
910
제2 가공 데이터 임베딩 처리 모듈
910_2
제3 가공 데이터 임베딩 처리 모듈
910_3
제1 데이터 모델 생성 모듈
912_1
912
제2 데이터 모델 생성 모듈
912_2
제3 데이터 모델 생성 모듈
912_3
제1 데이터 모델 선정 모듈
914_1
914
제2 데이터 모델 선정 모듈
914_2
제3 데이터 모델 선정 모듈
914_3
제2 데이터 정규화부
916
제2 데이터 전처리부
918
제4 가공 데이터 생성 모듈
920
920_1
제5 가공 데이터 생성 모듈
920_2
제6 가공 데이터 생성 모듈
920_3
제2 데이터 클러스터 생성부
922
제2 데이터 선택부
924
제4 가공 데이터 임베딩 처리 모듈
926_1
926
제5 가공 데이터 임베딩 처리 모듈
926_2
제6 가공 데이터 임베딩 처리 모듈
926_3
제1 정오탐 식별 모듈
928_1
928
제2 정오탐 식별 모듈
928_2
제3 정오탐 식별 모듈
928_3
정오탐 결정부
930
최종 정오탐 결과

도면10


1000
운영 체제
인공지능 기반 정오탐 식별 모델 생성 모듈
인공지능 기반 정오탐 식별 모듈
1010
1012
1014
1004
메모리 제어부
프로세서
인터페이스부
1006
1002
1008
학습 데이터
테스트 데이터
최종 정오탐 결과