(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년12월21일
(11) 등록번호 10-2480204
(24) 등록일자 2022년12월19일

(51) 국제특허분류(Int. Cl.)
G06F 21/55 (2013.01) *G06N 7/00* (2022.01)
H04L 9/40 (2022.01)
(52) CPC특허분류
G06F 21/55 (2013.01)
G06N 7/005 (2013.01)
(21) 출원번호 10-2019-7022466
(22) 출원일자(국제) 2018년01월22일
심사청구일자 2021년01월05일
(85) 번역문제출일자 2019년07월30일
(65) 공개번호 10-2019-0109427
(43) 공개일자 2019년09월25일
(86) 국제출원번호 PCT/US2018/014606
(87) 국제공개번호 WO 2018/140335
국제공개일자 2018년08월02일
(30) 우선권주장
15/419,933 2017년01월30일 미국(US)
(56) 선행기술조사문헌
US09306962 B1
US20100064369 A1
US20150067857 A1

(73) 특허권자
마이크로소프트 테크놀로지 라이센싱, 엘엘씨
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원 마이크로소프트 웨이
(72) 발명자
루오 펭칭
미국 워싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 마이크로소프트 테크놀로지 라이센싱, 엘엘씨
브리그스 리베스 호프
미국 워싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 마이크로소프트 테크놀로지 라이센싱, 엘엘씨
아마드 나비드
미국 워싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 마이크로소프트 테크놀로지 라이센싱, 엘엘씨
(74) 대리인
제일특허법인(유)

전체 청구항 수 : 총 20 항 심사관 : 정성훈

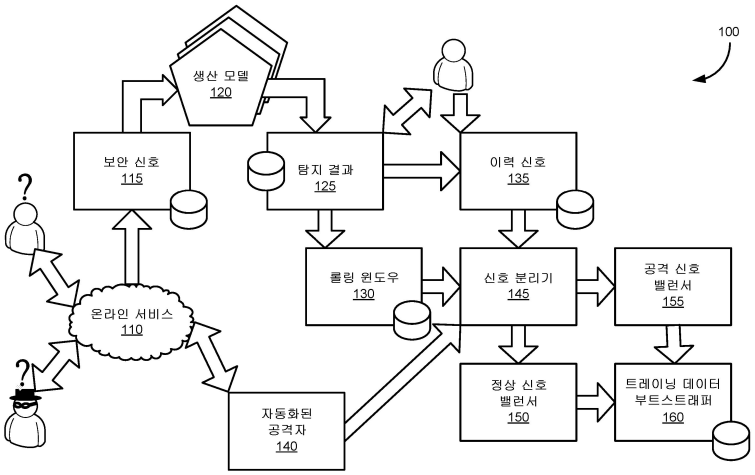
(54) 발명의 명칭 **침입 탐지를 위한 지속적인 학습**

(57) 요 약

네트워크 침입 탐지 모델을 트레이닝하는 데 사용되는 관찰 신호를 밸런싱함으로써, 악의적인 제3자로부터 네트워크를 보호하도록 컴퓨팅 리소스를 더 정확하게 할당하는 것을 가능하게 한다. 이 모델은 롤링 윈도우 내에서 정의되는 라이브 데이터 및 이력 데이터에 대해서 트레이닝되어서, 이 데이터 내에서 사용자 정의 피쳐를 탐지한

(뒷면에 계속)

대 표 도



다. 자동화된 공격을 통해서 롤링 트레이닝 윈도우 내에 다양한 종류의 공격이 항상 존재하는 것을 보장한다. 모델의 세트가 항상 트레이닝되어서, 어느 모델을 생산할 것인지 결정하고, 침입 분석가에게 경고하며, 및/또는 자동으로 대책을 배포한다. 이 모델은 피쳐가 재정의되고 롤링 윈도우의 데이터가 변경됨에 따라서 지속적으로 업데이트되고, 롤링 윈도우의 컨텐츠는 모델을 트레이닝하는데 사용할 각각의 관찰되는 타입의 충분한 데이터를 제공하도록 밸런싱된다. 데이터세트를 밸런싱할 때, 저비율 신호가 고비율 신호에 덮어쓰기되어서 이들의 상대적인 수를 배런싱한다.

(52) CPC특허분류
H04L 63/1425 (2013.01)

명 세 서

청구범위

청구항 1

프로세서에 의해 수행되어, 연속 학습 모델을 통해, 네트워크로 제공되는 온라인 서비스를 보안하는 방법으로서,

상기 온라인 서비스로부터 하나 이상의 네트워크 세션과 관련된 보안 신호의 세트를 수집하는 단계 - 상기 보안 신호의 세트는 시간의 롤링 윈도우(rolling window of time) 동안 수집됨 - 와,

상기 보안 신호의 세트의 각각의 보안 신호가 악의적인(malicious) 것인지 또는 정상적인(benign) 것인지 식별하는 단계와,

각각의 악의적 신호에 대해서 식별된 공격 타입에 기초해서 상기 보안 신호의 세트의 악의적 신호를 밸런싱하고, 각각의 정상 신호를 수신받은 장치 타입에 기초해서 상기 보안 신호의 세트의 정상 신호를 밸런싱하며, 상기 악의적 신호를 상기 정상 신호와 크로스 조인(cross-join)함으로써 상기 악의적 신호를 상기 정상 신호와 밸런싱함으로써, 상기 시간의 롤링 윈도우 동안의 밸런싱된 트레이닝 데이터세트를 생성하는 단계와,

상기 밸런싱된 트레이닝 데이터세트에 기초해서 예측 모델을 생성하는 단계 - 상기 온라인 서비스로부터 새로운 네트워크 세션과 관련된 추가 보안 신호를 수신하는 것에 응답해서, 상기 예측 모델이 적용되어서 상기 추가 보안 신호가 악의적인 것인지 또는 정상적인 것인지 판정함 -

를 포함하는 방법.

청구항 2

제 1 항에 있어서,

상기 보안 신호의 세트의 각각의 보안 신호가 악의적인 것인지 또는 정상적인 것인지를 식별하는 단계는,

생산 모델(production model)에서 각각의 보안 신호를 시험하는 단계 - 상기 생산 모델은 상기 밸런싱된 트레이닝 데이터세트에 따라서 모델 트레이너에 의해 생성되며, 특정 보안 신호가 악의적인 것인지 또는 정상적인 것인지의 탐지 결과를 생성하도록 구성됨 - 와,

상기 탐지 결과를 분석자 사용자에게 전송하는 단계와,

상기 탐지 결과와 관련해서 상기 분석자 사용자로부터의 액션을 수신하는 것에 응답해서, 상기 특정 보안 신호가 악의적인 것인지 또는 정상적인 것인지를 나타내도록 상기 탐지 결과를 업데이트하는 단계

를 포함하는

방법.

청구항 3

제 2 항에 있어서,

상기 온라인 서비스에 대해서 자동화된 공격자가 공격을 시뮬레이션하고,

상기 보안 신호의 세트의 각각의 보안 신호가 악의적인 것인지 또는 정상적인 것인지를 식별하는 단계는,

상기 자동화된 공격자로부터, 상기 공격에 응답하여 생성된 보안 신호를 식별하는 통지를 수신하는 단계와,

상기 탐지 결과에 관계없이, 상기 공격에 응답하여 생성된 상기 보안 신호를 악의적인 것으로 간주하는 단계

를 포함하는

방법.

청구항 4

제 2 항에 있어서,

상기 보안 신호의 세트의 각각의 보안 신호가 악의적인 것인지 또는 정상적인 것인지를 식별하는 단계는,

상기 특정 보안 신호로부터 피쳐를 추출하는 단계와,

상기 특정 보안 신호로부터 추출된 상기 피쳐가, 관리 사용자에 의해 공격 타입을 정의하는 것으로 지정된 피쳐 세트를 만족시키는지 여부를 결정하는 단계와,

상기 추출된 피쳐가 상기 피쳐 세트를 만족시킨다고 결정되는 것에 응답해서, 상기 특정 보안 신호를 악의적인 것으로 지정하는 단계와,

상기 추출된 피쳐가 상기 피쳐 세트를 만족시키지 않는다고 결정되는 것에 응답해서, 상기 특정 보안 신호를 정상적인 것으로 지정하는 단계

를 포함하는

방법.

청구항 5

제 4 항에 있어서,

상기 각각의 악의적 신호에 대해서 상기 식별된 공격 타입에 기초해서 상기 보안 신호의 세트의 악의적 신호를 밸런싱하는 단계는,

상기 악의적 신호에 대해 관찰된 공격 타입의 세트에 대한 공격 타입의 상대적인 수를,

상기 공격 타입의 세트에서 과소 표시된(underrepresented) 공격 타입의 상대적인 양을 증가시키는 것과,

상기 공격 타입의 세트에서 과대 표시된(overrepresented) 공격 타입의 상대적인 양을 감소시키는 것

중 적어도 하나에 의해서 밸런싱하는 단계

를 포함하는

방법.

청구항 6

제 4 항에 있어서,

상기 피쳐 세트는 상기 관리 사용자에 의해 제출되는 구조화된 문서로 식별되고,

상기 구조화된 문서는 상기 보안 신호의 세트에서 관찰할 피쳐 타입 및 데이터 필드를 식별하며,

상기 피쳐 세트의 피쳐는 코드를 수정할 필요없이 상기 구조화된 문서에 기초해서 상기 보안 신호의 세트로부터 유동적으로 추출되는

방법.

청구항 7

제 1 항에 있어서,

상기 보안 신호 세트에는 이력 데이터 신호가 포함되는

방법.

청구항 8

제 1 항에 있어서,

상기 각각의 정상 신호를 수신받은 장치 타입에 기초해서 상기 보안 신호의 세트의 정상 신호를 밸런싱하는 단계는,

상기 롤링 윈도우 동안 적어도 하나의 악의적 신호가 수집되는, 상기 온라인 서비스 내의 장치를 식별하는 단계와,

상기 보안 신호의 세트로부터 상기 장치와 관련된 정상 신호를 제거하는 단계

를 포함하는

방법.

청구항 9

제 8 항에 있어서,

상기 악의적 신호를 상기 정상 신호와 크로스 조인함으로써 상기 악의적 신호와 상기 정상 신호를 밸런싱하는 것은, 상기 악의적 신호를 상기 정상 신호와 크로스 조인해서 상기 밸런싱된 트레이닝 데이터세트에 대한 복수의 공격 시나리오를 생성하는 것을 포함하고,

상기 복수의 공격 시나리오는 상기 악의적 신호 및 상기 정상 신호를 포함하는

방법.

청구항 10

제 1 항에 있어서,

상기 보안 신호의 세트가 상기 시간의 롤링 윈도우 동안 수집되는 것은,

현재 시간으로부터 상기 롤링 윈도우를 정의하는 시간 프레임을 식별하는 것과,

상기 현재 시간에, 상기 온라인 서비스로부터 새로운 보안 신호를 수신하는 것과,

상기 현재 시간에 수신한 상기 새로운 보안 신호를 상기 보안 신호의 세트에 추가하는 것과,

상기 현재 시간부터 상기 시간 프레임 밖에서 상기 보안 신호의 세트에 추가된 기존 보안 신호를 식별하는 것과,

상기 현재 시간부터 상기 시간 프레임 밖에서 상기 보안 신호의 세트에 추가된 것으로 식별된 상기 기존 보안 신호를, 상기 보안 신호의 세트로부터 제거하는 것

을 포함하는

방법.

청구항 11

연속 학습 모델을 통해, 네트워크로 제공되는 온라인 서비스를 보안하는 시스템으로서,

프로세서와,

명령어가 포함된 메모리 저장 장치

를 포함하고,

상기 명령어는, 상기 프로세서에 의해 실행될 때,

상기 온라인 서비스 내의 장치로부터 하나 이상의 네트워크 세션과 관련된 보안 신호를 수신하고,

상기 보안 신호 각각으로부터 피쳐 벡터를 추출하며 - 특정 피쳐 벡터가, 특정 보안 신호가 수신되는 특정 장치의 상태를 나타내는 숫자 값을 제공함 - ,

관련된 예측 모델을 통해서 상기 피쳐 벡터 각각에 대한 탐지 결과를 생성하고 - 특정 탐지 결과는, 특정 네트워크 세션과 관련된 상기 특정 보안 신호가 상기 특정 장치에 대해 악의적 활동인지 정상 활동인지 여부를 식별함 - ,

복수의 보안 신호 및 관련 탐지 결과를 포함하는 롤링 윈도우를 정의하며 - 상기 관련 탐지 결과는 현재 시간으로부터 타임프레임 내에 수신된 것임 - ,

상기 롤링 윈도우 동안 밸런싱된 트레이닝 데이터세트를 생성하도록

동작 가능하고,

상기 밸런싱된 트레이닝 데이터세트를 생성하기 위해서, 상기 시스템은

악의적 활동을 나타내는 것으로 식별되는 상기 롤링 윈도우 내의 상기 보안 신호 각각의 공격 타입을 식별하고,

과대 표시된 공격 타입으로 식별되는 보안 신호에 비한, 상기 롤링 윈도우 내에서 과소 표시된 공격 타입으로 식별되는 보안 신호의 양을 증가시키며,

악의적 활동을 나타내는 것으로 식별된 상기 보안 신호를 정상 활동을 나타내는 것으로 식별된 보안 신호와 크로스 조인해서, 상기 롤링 윈도우에 대한 공격 시나리오를 생성하고,

머신 학습 알고리즘에 따라서, 상기 밸런싱된 트레이닝 데이터세트에 기초해서 상기 관련 예측 모델을 업데이트하도록

구성되는

시스템.

청구항 12

제 11 항에 있어서,

상기 관련 예측 모델을 업데이트하는 것은, 상기 머신 학습 알고리즘이, 개발 모델이 상기 밸런싱된 트레이닝 데이터세트에 따라서 상기 보안 신호가 상기 장치에 대해 악의적 활동인지 혹은 정상 활동인지 여부를 더 정확하게 식별한다고 나타내는 것에 응답해서, 상기 탐지 결과를 생산하는데 사용되는 생산 모델을 상기 머신 학습 알고리즘에 따라서 상기 밸런싱된 트레이닝 데이터세트로부터 개발되는 상기 개발 모델로 대체하는 것을 포함하는

시스템.

청구항 13

제 11 항에 있어서,

상기 롤링 윈도우에는 이력 신호가 포함되고, 상기 이력 신호는 상기 타임프레임 밖에서 수집되는 보안 신호를 포함하는

시스템.

청구항 14

제 11 항에 있어서,

상기 온라인 서비스 내의 상기 장치로부터 수신된 상기 보안 신호는, 자동화된 공격자가 상기 온라인 서비스에 대해 알려진 악의적 활동을 수행하는 것에 응답해서 생성된 보안 신호를 포함하고,

상기 자동화된 공격자가 상기 악의적 활동을 수행하는 것에 응답해서 생성된 상기 보안 신호에 대해 생산된 상기 탐지 결과는, 상기 자동화된 공격자로부터의 통지에 기초해서 상기 특정 보안 신호가 악의적 활동을 나타내는 것으로 표시하도록 설정되는

시스템.

청구항 15

제 11 항에 있어서,

상기 밸런싱된 트레이닝 데이터세트를 생산하기 위해서, 상기 시스템은, 정상 활동을 나타내는 것으로 식별된 상기 보안 신호가 수신되는 특정한 장치가 상기 롤링 윈도우 내에서 악의적 활동을 나타내는 것으로 식별된 하나 이상의 보안 신호와 관련되는 것을 식별하는 것에 응답해서, 상기 롤링 윈도우로부터 정상 활동을 나타내는 것으로 식별된 보안 신호를 제거하도록 더 구성되는

시스템.

청구항 16

제 11 항에 있어서,

상기 밸런싱된 트레이닝 데이터세트를 생산하기 위해서, 상기 시스템은,

정상 활동을 나타내는 것으로 식별된 상기 롤링 윈도우 내의 상기 보안 신호 각각이 수신되는 장치 타입을 식별하고,

과대 표시된 장치 타입으로 식별된 보안 신호에 비한 상기 롤링 윈도우 내의 과소 표시된 장치 타입으로 식별된 보안 신호의 양을 증가시키도록

더 구성되는

시스템.

청구항 17

연속 학습 모델을 통해, 네트워크로 제공되는 온라인 서비스를 보안하는 프로세서 실행 가능 명령어가 포함된 컴퓨터 판독 가능 저장 장치로서,

상기 명령어는

상기 온라인 서비스로부터 하나 이상의 네트워크 세션과 관련된 보안 신호의 세트를 수집하는 것 - 상기 보안 신호의 세트는 시간의 롤링 윈도우 동안 수집됨 - 과,

상기 보안 신호의 세트의 각각의 보안 신호를 예측 모델을 통해 시험해서 특정 네트워크 세션과 관련된 각각의 보안 신호가 악의적인 것인지 또는 정상적인 것인지 여부를 식별하는 것 - 상기 예측 모델은, 관리 사용자에 의

해 정의되는 피쳐 벡터에 기초하여 특정 보안 신호가 악의적인 것인지 또는 정상적인 것인지 여부의 탐지 결과를 생산하도록 구성됨 - 과,

상기 보안 신호의 세트를 상기 탐지 결과와 연관시켜서 상기 보안 신호의 세트를 악의적 신호 또는 정상 신호로서 식별하는 것과,

상기 악의적 신호를 상기 정상 신호와 밸런싱해서, 밸런싱된 트레이닝 데이터세트를 생산하는 것

을 포함하고,

상기 밸런싱된 트레이닝 데이터세트를 생산하는 것은,

상기 악의적 신호 각각의 공격 타입을 식별하는 것과,

상기 정상 신호 각각이 수집되는 장치 타입을 식별하는 것과,

식별되는 공격 타입에 기초해서 상기 롤링 윈도우 내의 악의적 신호의 상대적인 수를 균등화(equalizing)해서 공격 예시(attack examples)의 세트를 생산하는 것과,

식별되는 장치 타입에 기초해서 상기 롤링 윈도우 내의 정상 신호의 상대적인 수를 균등화해서 식별되는 정상 예시의 세트를 생산하는 것과,

상기 공격 예시의 세트를 상기 정상 예시의 세트의 적어도 일부와 크로스 조인해서, 상기 정상 예시의 세트 내의 정상 예시의 수에 비한 상기 공격 예시의 세트 내의 공격 예시의 수를 밸런싱하는 것과,

상기 밸런싱된 트레이닝 데이터세트 및 머신 학습 알고리즘에 기초해서 상기 예측 모델을 개선하는 것

을 포함하는

컴퓨터 판독 가능 저장 장치.

청구항 18

제 17 항에 있어서,

상기 악의적 신호를 상기 정상 신호와 밸런싱하는 것은,

상기 롤링 윈도우 동안 적어도 하나의 악의적 신호가 수집되는, 상기 온라인 서비스 내의 장치를 식별하는 것과,

상기 보안 신호의 세트로부터 상기 장치와 관련된 정상 신호를 제거하는 것

을 포함하는

컴퓨터 판독 가능 저장 장치.

청구항 19

제 17 항에 있어서,

상기 각각의 보안 신호를 상기 예측 모델을 통해 시험하는 것은,

상기 탐지 결과를 분석자 사용자에게 전송하는 것과,

상기 탐지 결과와 관련해서 상기 분석자 사용자로부터의 액션을 수신하는 것에 응답해서, 상기 특정 보안 신호가 상기 분석자 사용자로부터의 상기 액션의 견지에서 악의적인 것인지 또는 정상적인 것인지를 나타내도록 상기 탐지 결과를 업데이트하는 것

을 포함하는

컴퓨터 판독 가능 저장 장치.

청구항 20

제 17 항에 있어서,

상기 피처 벡터는, 하나 이상의 공격 타입을 식별하도록 구성된 예측 모델을 정의하도록 상기 관리 사용자에 의해 제출되는 구조화된 문서로 식별되는 상기 장치의 상태의 피처의 세트를 포함하고,

상기 구조화된 문서는 상기 보안 신호의 세트에서 관찰할 피처 타입 및 데이터 필드를 명시하며,

상기 피처의 세트의 피처는 코드를 수정할 필요없이 상기 구조화된 문서에 기초해서 상기 보안 신호의 세트로부터 유동적으로 추출되는

컴퓨터 판독 가능 저장 장치.

발명의 설명

기 술 분 야

배 경 기 술

[0001] 컴퓨터 네트워크에는, 악의적인 제3자가 호스트되는 시스템에 대해 미승인 액세스를 시도하는 위험이 항상 존재한다. 악의적인 제3자가 네트워크를 공격하는데 사용하는 전술과 네트워크 관리자가 공격을 방어하는데 사용하는 전술은 서로를 고려해서 계속해서 진화되고 있으며, 악의적인 제3자의 무기고에는 새로운 악용법(exploit)이 추가되고 비효율적인 악용법은 삭제된다. 그러나, 대책을 강구하는 것은 종종 반작용적인 측면이 있으며, 즉 네트워크 관리자는 대응책을 배포하기 전에 최신 악용법을 식별해야 하고, 대응하는 악용법이 더 이상 사용되지 않으면 언제 대응책 배포를 중단해야 할지 결정해야 한다. 네트워크 관리자에게 있어서 최신 악용법을 정확하게 식별해서 차단하는 것은 종종 어려운 일이며, 특히 악용법이 아직 널리 확산되지 않았거나 혹은 네트워크에서 제공되는 서비스의 작은 부분만을 공격하는 경우에 그러하다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0002] 본 개요는 이하 상세한 설명에서 더 상세하게 설명될 개념 중 일부를 선택해서 간략하게 소개하고자 하는 것이다. 본 개요는 청구되는 청구 대상의 모든 중요한 혹은 필수적인 특징을 나타내고자 하는 것이 아니며, 청구되는 청구 대상의 범주를 결정하는데 도움을 주고자 하는 것도 아니다.

[0003] 침입 탐지를 위해 지속적인 학습을 제공하는 시스템, 방법 및 명령어를 포함하는 컴퓨터 저장 장치가 제공된다. 다수의 머신 학습 모델은, 공격 거동 및 정상(benign) 거동을 나타내는, 네트워크 내의 머신 및 디바이스로부터 수집된 신호에 기초해서, 네트워크 신호에 대해 지속적으로 재트레이닝된다(retrain). 이 모델이 최신 데이터를 사용해서 공격을 식별할 수 있도록, 롤링 윈도우를 사용해서 신호를 수집하는 데, 이 모델은, 자신이 갖고 있는 성능이 가장 최근 데이터를 구성하는 것에 응답해서 성공적으로 혹은 그렇지 않게(wax and wane) 공격을 정확하게 탐지해 내는지에 따라서, 지속적으로 프로모트되고 디모트되면서 네트워크를 보호한다. 활성 생산(active production) 네트워크로 배포되는 모델은 보안 분석가에게 거의 실시간 탐지를 제공하며, 보안 분석가는 이 모델의 정확성에 대한 피드백을 제공해서(예를 들어, 놓친 침입/오류 부정(false negatives), 오류 긍정(false positives), 오인된 침입), 이 모델의 트레이닝 방법을 개선한다.

[0004] 탐지 모델을 지속적으로 재트레이닝하고 개선하는데 사용되는 트레이닝 데이터세트의 신뢰도를 향상시켜서 모델을 향상시키기 위해서, 정상 신호(benign signal)와 비교한 공격 신호의 희소성을 해결하도록 공격 신호는 특정

한 공격 타입을 고려해서 밸런싱된다. 정상 신호를 다른 머신으로부터의 다양한 타입의 공격 신호로 덮어쓰기해서, 이 모델을 트레이닝하고 개선하기 위한 밸런싱된 트레이닝 세트를 제공한다. 밸런싱된 트레이닝 세트 내의 공격 신호 중에서 다양한 공격 타입의 신호가, 이 모델이 모든 공격 타입에 대해서 확실히 동일하게 트레이닝되도록, 밸런싱된다. 이 신호의 피쳐가 텍스트 기반 구성을 통해 유동적으로 추출되며, 따라서 네트워크에 대한 공격을 나타내는 다양한 피쳐 세트에 응답하도록 모델의 유연성을 향상시킨다.

[0005] 다양한 측면에서, 네트워크의 준비 상태를 증가시키고 추가적인 공격 신호를 생성하기 위해서, 공격은 알려진 내부 공격자에 의해 시뮬레이션된다. 유사하게, 일부 측면에서는, 이력적으로 중요한 공격 신호가 사용되면, 특정한 타입의 공격 신호가 롤링 윈도우에서 관찰되지 않은 경우라도 그 신호는 모델에 제공된다.

[0006] 네트워크에 대한 연속 학습 침입 탐지 모델을 제공함으로써, 네트워크 내의 장치 및 소프트웨어의 기능이 향상된다. 새로운 형태의 공격이 보다 빠르고 안정적으로 식별되므로, 네트워크의 보안을 향상시키는 방법에 있어서의 컴퓨터 중심의 문제를 해결한다. 나아가, 사용되지 않는 공격 형태를 탐지하려고 시도하면서 컴퓨팅 리소스가 낭비되는 일이 없으므로, 악의적인 제3자로부터 네트워크를 보호하는 데 사용되는 처리 리소스가 감소된다.

[0007] 예시들은 컴퓨터 프로세스, 컴퓨팅 시스템으로서 또는 장치, 컴퓨터 프로그램 제품 또는 컴퓨터 판독가능 매체와 같은 제조 물품으로서 구현된다. 일 측면에 따르면, 컴퓨터 프로그램 제품은 컴퓨터 시스템에 의해 판독 가능하고 컴퓨터 프로세스를 실행하기 위한 명령어를 포함하는 컴퓨터 프로그램을 인코딩하는 컴퓨터 저장 매체이다.

[0008] 하나 이상의 측면의 세부 사항이 첨부된 도면들 및 이하의 설명에서 설명된다. 다음의 상세한 설명을 읽고 관련 도면들을 검토함으로써 다른 특징 및 이점이 명백해질 것이다. 이하의 상세한 설명은 단지 설명을 위한 것으로 청구 범위를 제한하는 것은 아니라는 것을 이해해야 한다.

도면의 간단한 설명

[0009] 첨부 도면은 본 개시에 포함되어 그 일부를 구성하며, 다양한 측면을 도시한다.

도 1a는 본 개시가 실시될 수 있는 예시적인 보안 시스템이다.

도 1b는 본 개시가 실시될 수 있는 도 1a의 예시적인 보안 시스템과 함께 사용하기 위한 예시적인 모델 트레이닝 및 선택 시스템이다.

도 2는 온라인 서비스를 보안하는 데 사용하기 위한 예측 모델을 트레이닝시키는데 사용되는 트레이닝 데이터세트를 개발하는 예시적인 방법에 포함된, 전반적인 단계를 나타내는 흐름도이다.

도 3은 온라인 서비스를 보안하는 데 사용하기 위한 예측 모델을 트레이닝시키고 선택하는 예시적인 방법에 포함된 전반적인 단계를 나타내는 흐름도이다.

도 4는 컴퓨팅 장치의 예시적인 물리적 구성 요소를 나타내는 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0010] 이하의 상세한 설명은 첨부된 도면을 참조한다. 도면 및 이하 상세한 설명에서 동일한 혹은 유사한 요소를 가리키는데 가능한 한 균등화한 참조 번호가 사용된다. 예시가 기술될 수 있지만, 수정, 개조 및 다른 구현이 가능하다. 예를 들어, 도면에 예시된 요소들에 대한 대체, 추가 또는 수정이 이루어질 수 있으며, 본 명세서에 기재된 방법은 개시된 방법의 단계들을 대체, 재정렬 또는 추가하는 방식으로 수정될 수 있다. 따라서, 이하 상세한 설명은 제한적인 것이 아니며, 적절한 범주는 첨부된 청구 범위에 의해 정의된다. 예시는 하드웨어 구현이나, 완전한 소프트웨어 구현이나 또는 소프트웨어와 하드웨어 측면을 결합한 구현의 형태를 취할 수 있다. 따라서, 이하 상세한 설명은 한정의 의미로 해석되어서는 안된다.

[0011] 본 명세서에는, 연속 학습 침입 탐지 모델을 통해서 향상된 네트워크 보안을 제공하는 시스템, 방법 및 명령어를 포함하는 컴퓨터 판독가능 저장 장치가 개시된다. 네트워크에 대한 연속 학습 침입 탐지 모델을 제공함으로써, 네트워크의 장치 및 소프트웨어의 기능이 향상된다. 새로운 형태의 공격이 보다 빠르고 안정되게 식별되므로, 네트워크의 보안을 향상시키는 방법에 있어서의 컴퓨터 중심의 문제를 해결한다. 나아가, 사용되지 않는 공격 형태를 탐지하려고 시도하면서 컴퓨팅 리소스가 낭비되는 일이 없으므로, 악의적인 제3자로부터 네트워크를 보호하는 데 사용될 처리 리소스가 감소된다.

[0012] 도 1a는 본 개시가 실시될 수 있는 예시적인 보안 시스템(100)이다. 도 1a에 도시된 바와 같이, 온라인 서비스

(110)는 다양한 사용자 - 정상 사용자일 수도 있고 악의적인 사용자일 수도 있음 - 및 보안 시스템(100)에 접속된다. 온라인 서비스(110)는 다양한 사용자에게 '클라우드' 서비스를 제공하는 클라우드 데이터 센터와 같은 네트워크화된 컴퓨팅 장치의 세트를 나타내며, '클라우드' 서비스는 비한정의 예로서, 사용자가 온라인 서비스(110)의 장치에서 실행되는 운영 체제 및 소프트웨어를 제공하는 서비스형 인프라(Infrastructure as a Service;IaaS), 또는 사용자가 소프트웨어를 제공하고 온라인 서비스(110)가 운영 체제 및 장치를 제공하는 서비스로서의 플랫폼(Platform as a Service;PaaS) 또는 온라인 서비스(110)가 사용자에게 장치에서 실행되는 운영 체제 및 소프트웨어를 모두 제공하는 서비스로서의 소프트웨어(Software as a Service;SaaS)를 포함한다. 온라인 서비스(110)에 대한 엑세스를 시도하는 사용자는 합법적인 사용자일 수도 있고 악의적인 제3자일 수도 있으며, 보안 취약점을 악용해서 온라인 서비스(110)에 침입해서 온라인 서비스(110)로부터의 데이터를 합법적인 인증없이 처리 및/또는 취득한다.

[0013] 사용자가 정상인지 혹은 악의적인지 여부를 판정하기 위해서 또는 장치가 안전한지(악의적 신호(malicious signals)를 송신하지 않음) 또는 손상되었는지(악의적 신호를 송신함)를 판정하기 위해, 온라인 서비스(110)로부터의 다양한 보안 신호(115)가 수집되어 생산 모델(120)에 보내져서, 해당 세션이 악의적인지 또는 정상적인지 여부를 나타내는 탐지 결과(125)를 생산한다. 보안 신호(115)는 이벤트 로그, 네트워크 트레이스, 시스템 명령 등을 포함하며, 이는 생산 모델(120)에 의해 분석되어서, 생산 모델(120)의 트레이닝을 통해 결정된 특성 및 이들의 피쳐(feature) 값이 악의적 거동 혹은 정상 거동을 나타내게 된다. 본 개시의 목적에서, 특정한 보안 신호(115)는 이 특정한 보안 신호(115)를 생성하는 것과 관련된 온라인 서비스(110) 내의 액션에 기초해서 '악의적'이라고 칭해지거나 혹은 '정상'이라고 칭해진다. 또한, 본 명세서에서 사용되는 용어 '피쳐'는, 네트워크에서 관찰되는 특성 또는 거동과 관련되어 있으며 '모델'에 의해 수용되는 하나 이상의 입력 신호로부터 유도된 수치적인 속성으로, 이는 피쳐의 값을 이 피쳐가 악의적 활동을 나타내는지 혹은 정상 활동을 나타내는지 여부에 대한 예측 스코어 또는 신뢰도로 변환하도록 분석가에 의해서 정의되는 피쳐 세트(모델 피쳐라고도 함)를 수용하는 알고리즘이다.

[0014] 보안 신호(115)는 생산 모델(120)에 제공되어서 보안 신호(115)로부터 다양한 피쳐를 추출하며, 이에 대해서 생산 모델(120)이 온라인 서비스(110) 상의 악의적 활동을 식별하도록 트레이닝된다. 보안 신호(115)는 온라인 서비스(110) 내의 장치에서 일어나는 하나 이상의 관련 이벤트의 집합이며, 이는 다수의 피쳐(예를 들어, 사용되는 포트, 접속된 IP 주소, 신호가 수신되는 장치의 신원/타입, 사용자, 취해진 액션)를 포함할 수 있는데, 그 서브세트가 특정 생산 모델(120)에 의한 시험을 위해서 추출되어서 보안 신호(115)가 정상적인지 악의적인지 여부를 결정한다. 하나 이상의 보안 신호(115)로부터의 피쳐는 분석을 위해서 피쳐 벡터로 결합되고, 이 피쳐는 다양한 측면에서 스코어링되어서 생산 모델(120)에 입력하기 위한 피쳐의 수치적인 분석을 제공한다.

[0015] 예를 들어, 특정 IP(Internet Protocol) 주소가 그 사용 빈도에 기초해서 스코어링될 수 있으며, 여기서 롤링 윈도우(130) 동안에 특정 IP 주소를 빈번하게 사용하면 사용 빈도가 적은 것에 비해서 생산 모델(120)에 제시되는 값은 달라질 것이다. 반대로, 중요한 파일이 액세스되거나, 금지된 동작이 취해지거나, 블랙리스트에 있는 IP 주소가 통신되는 등의 경우에는, 위험 상태가 발생했다는 것을 나타내는 2진 스코어가 보안 신호(115) 내에서 생산 모델(120)에 제공될 수 있다. 생산 모델(120)은 화이트리스트 혹은 블랙리스트에 의존하지 않으며, 이들의 트레이닝은 도 1b, 2 및 3과 관련해서 이하 상세하게 설명되는 보안 신호(115)에서 관찰되는 피쳐에 관련되며, 이는 온라인 서비스(110)에 대한 침임을 나타내는 화이트리스 혹은 블랙리스트에 관계없이 피쳐를 일정 시간 동안 학습할 수 있다.

[0016] 특정한 보안 신호(115)에 대해서, 생산 모델(120)에 의한 결정은 문제의 보안 신호(115)가 정상적인 것인지 혹은 악의적인 것인지 여부를 나타낸다. 이들 탐지 결과(125)는 보안 신호(115)와 관련되어서 이들이 정상적인 것인지 혹은 악의적인 것인지 여부를 식별한다. 일부 측면들에서, 이들 탐지 결과(125) 중 적어도 일부(예를 들어, 악의적인 탐지 결과(125))는 분석가 사용자에게 제공되고, 이 분석가 사용자는 탐지 결과(125)에 동작을 취해서 악의적인 사용자 또는 공격에 대한 대책을 배치할 수도 있고 또는 탐지 결과(125)가 생산 모델(120)이 나타내는 것과는 상이한 평가를 보증하도록 결정할 수도 있다. 예를 들어, 악의적 신호의 오류 부정이 생산 모델(120)에 의해 표시되었다면, 분석가는 이 신호가 실제로는 악의적인 것이라고 평가해서, 취해져야 할 액션을 나타낼 수 있다. 또 다른 예에서, 정상 신호에 대한 오류 긍정이 표시되었다면, 분석가는 신호가 실제로는 정상적인 것이라고 평가해서, 어떠한 액션도 취해질 필요가 없다고 나타낼 수 있다. 또 다른 예에서, 악의적인 행동에 대한 오류 긍정이 표시되었다면, 분석가는 보안 시스템(100)에 의해 권고된 것 이외의 어떠한 행동이나 다른 행동이 취해지지 않는다는 것을 나타낼 수 있다. 모델을 추가적으로 트레이닝하고 향상시키는데 분석가로부터의 수정이 사용된다.

[0017] 다양한 측면에서 탐지 결과(125)는 또한 과거 d일 동안(여기서 d는 예를 들어 2일, 10일, 15일 등과 같이 분석가 사용자 혹은 다른 네트워크 관리자에 의해 구성될 수 있다)에 관찰되는 보안 신호(115)의 롤링 윈도우(130)가 저장된 데이터베이스나 혹은 과거 d일 동안 관찰되었는지 여부에 관계없이 트레이닝을 위해서 사용되어야 하는 보안 신호(115)의 이력 신호(135)가 저장된 데이터베이스에 전송된다. 이력 신호(135)는 알려진 외부 공격과 관련된 보안 신호(115)를 포함하도록 분석가 사용자에 의해서 큐레이트된다. 추가적인 측면에서, 분석가 사용자는 의심스럽게 보일 수 있는 혹은 네트워크 침입을 오류 긍정으로 리턴할 수 있는 정상 신호를 포함하도록 이력 신호(135)를 큐레이트해서, 이력적으로 적절하게 식별하기 어려운 것으로 입증된 신호에 적절하게 응답하도록 예측 모델이 트레이닝되는 것을 보증한다.

[0018] 자동화된 공격자(140)는 알려진 공격 패턴을 사용해서 온라인 서비스(110)의 보안을 테스트하며, 생산 모델(120)이 생산하는 탐지 결과와 함께 사용할 알려진 결과를 제공한다. 보안 신호(115)에 대한 탐지 결과(125)가 자동화된 공격자(140)로부터의 공격의 결과가 이 공격이 악의적이었다는 것을 명시하지 않는 것이면, 자동화된 공격자(140)가 보안 신호(115)를 악의적인 것으로 명시함에 따라서 보안 신호(115)는 악의적인 것으로 취급될 것이다. 다양한 측면에서, 자동화된 공격자(140)는 보안 시스템(100) 또는 온라인 서비스(110)의 옵션인 구성요소이다.

[0019] 보안 신호(115)(이용 가능하다면 롤링 윈도우(130) 및 이력 신호(135) 내의 것을 포함하는)는, 보안 신호(115)가 정상적인 것으로 판정되었는지 혹은 악의적인 것으로 판정되었는지를 나타내는 생산 모델(120)로부터의 탐지 결과(125)(및 분석가 사용자로부터의 교정)와 함께 신호 분리기(145)로 전송된다. 유사하게, 자동화된 공격자(140)가 이용되는 양상에서, 온라인 서비스(110)에 대한 액션으로부터 생성된 보안 신호(115)의 정상/악의 식별 결과가 신호 분리기(145)에 제공된다. 신호 분리기(145)는 보안 신호(115)를 정상 신호 밸런서(150)에 제공되는 정상 신호 및 공격 신호 밸런서(155)로 제공되는 악의적 신호로 분할하도록 구성된다.

[0020] 정상 신호 밸런서(150) 및 공격 신호 밸런서(155)는 트레이닝 데이터 부트스트래퍼(bootstrapper)(160)에 의해 사용되는 데이터세트를 채우는데 사용되는 보안 신호(115)의 세트를 개발해서, 모델을 트레이닝해서 온라인 서비스(110)의 최신 악용을 탐지하도록 밸런싱된 정상 신호 및 악의적 신호를 제공한다. 트레이닝 데이터 부트스트래퍼(160)는 온라인 서비스(110) 내의 손상된 장치로부터 수신된 정상 신호를 제거해서, 손상된 장치로부터의 악의적 신호만을 남긴다. 손상되지 않은 장치로부터의 정상 신호는 손상된 장치로부터의 악의적 신호와 크로스 조인(cross join)되어서 BxM 공격 예시를 생성하며, 여기서 B는 정상 예시의 수를 나타내고 M은 악의적인 예시의 수를 나타낸다. 이는, 손상되지 않은 장치에서 공격이 발생된 것처럼 공격 예시를 정상 예시에 중첩시킨 확장된 데이터세트를 생성한다.

[0021] 손상되지 않은 장치는 정상 신호의 상이한 변형예를 가지며, 손상된 장치는 공격 신호의 상이한 변형예를 가지므로, 2개의 데이터세트를 크로스 조인함으로써 많은 변형예를 가진 많은 수의 시나리오를 생성한다. 그러나, 시나리오가 자동화된 공격자(140) 등에 의해서 랜덤하게 선택되는 경우에는, 다른 수의 각각의 공격 타입이 트레이닝 세트에 제시될 수 있으며, 이는 모델의 트레이닝을 왜곡할 수 있다(그 결과 일부 공격이 다른 것보다 더 양호하게 예측될 수 있다). 따라서, 공격 예시는 트레이닝 세트의 각각의 예시 공격과 실질적으로 같은 수(예컨대, ± 5%)가 존재하도록 공격 시나리오에 대해 밸런싱될 수 있다. 다양한 측면에서, 과소 표시된(underrepresented) 공격 타입(즉, 밸런싱된 수 이하의 수량의 공격 타입)은 그 상대적인 수를 증가시키도록 복제된 기존의 악의적인 시그널을 갖고 있고, 및/또는 과대 표시된(overrepresented) 공격 타입(즉, 밸런싱된 수 이상의 수량의 공격 타입)은 밸런싱된 공격 예시 세트에 도달하도록 과소 표시된 공격 타입의 예시에 의해 삭제 혹은 대체/중복된 기존 악의적인 시그널을 갖고 있다.

[0022] 악의적 신호와 유사하게, 정상 신호가, 신호를 수신받은 장치 타입 혹은 역할을 고려해서 서로에 대해서 밸런싱되고, 이로써 특정 장치 타입 혹은 역할은 트레이닝 데이터세트에 과대 표시되지 않는다(그 결과 특정 장치 타입/역할에 대해서는 일부 공격은 다른 것에 비해서 더 양호하게 예상되게 된다). 따라서, 정상 예시는, 정상 예시를 제공하는 각각의 장치 타입과 실질적으로 같은 수(예를 들어, ± 5%)가 존재하도록, 이용 가능한 장치 타입에 대해서 밸런싱된다. 다양한 측면에서, 과소 표시된 장치 타입(즉, 밸런싱된 수 이하의 수량의 장치 타입)은 그 상대적인 수를 증가시키도록 복제된 기존의 정상적인 시그널을 갖고 있고, 및/또는 과대 표시된 공격 타입(즉, 밸런싱된 수 이상의 수량의 공격 타입)은 밸런싱된 정상 예시 세트에 도달하도록 과소 표시된 장치 타입으로부터의 정상 예시에 의해 삭제 혹은 대체/중복된 기존 정상적인 시그널을 갖고 있다.

[0023] 도 1b는 본 개시가 실시될 수 있는 도 1a의 예시적인 보안 시스템(100)과 함께 사용하기 위한 예시적인 모델 트레이닝 및 선택 시스템(105)이다. 트레이닝 데이터 부트스트래퍼(160)로부터의 정상적인 및 악성적인 신호의 밸

런싱된 데이터세트가 트레이닝/테스트 분리기(165)에 제공되어서, 온라인 서비스(110)를 보안하는 데 사용하기 위한 다양한 모델을 트레이닝 및 평가한다. 데이터세트는 k개의 서브 세트들로 분할되고, 여기서 이용 가능한 서브 세트(예컨대, 2/3)의 k-1이 모델을 트레이닝하는 데 사용되며, 데이터세트의 한 서브 세트(예컨대, 1/3)는 모델을 평가하기 위해 보존된다. 다양한 측면에서, 데이터세트를 트레이닝 서브 세트와 평가 서브 세트로 분리하기 위해 다양한 비율이 상정되고, 트레이닝 서브 세트와 평가 서브 세트는 모델 트레이너(170) 및 모델 평가자(175)에 각각 제공된다.

[0024] 모델 트레이너(170)는 밸런싱 데이터의 트레이닝 서브 세트를 통해 하나 이상의 머신 학습 기술을 통해 복수의 개발 모델(180)을 트레이닝하도록 구성된다. 머신 학습 기술은 모델에 공급되는 데이터에 대한 예측을 정확하게 수행하도록 모델을 트레이닝한다(예를 들어, 보안 신호(115)가 정상적인 것인 악의적인 것인지 여부; 명사가 사람인지 장소인지 물건인지; 또는 내일 날씨가 어떨지). 학습 기간 동안에는, 특정 입력에 대한 출력을 정확하게 예측하도록 모델을 최적화하기 위해, 알려진 입력(예를 들어, 샘플 A, 샘플 B, 샘플 C)의 트레이닝 데이터세트에 대해 이 모델이 개발된다. 일반적으로 학습 기간은 지도(supervised)일 수도 있고, 반지도(semi-supervised)일 수도 있으며, 비지도(unsupervised)일 수도 있는데, 이는 트레이닝 입력에 따라 "정확한" 출력이 제공되는 감소 레벨을 나타낸다. 지도 학습 기간에는, 모든 출력이 모델에 제공되고 모델은 입력을 출력에 매핑하는 일반적인 규칙 혹은 알고리즘을 개발하도록 맞춰진다(directed to). 반대로 비지도 학습 기간에는, 입력에 대해서 원하는 출력이 제공되지 않으며, 따라서 모델은 트레이닝 데이터세트 내에서 관계를 발견하도록 자체 역할을 개발할 수 있다. 반지도 학습 기간에는, 불완전하게 라벨링된 트레이닝 세트에 알려진 출력의 일부 및 트레이닝 데이터세트의 알려지지 않은 일부가 제공된다.

[0025] 이 모델은 여러 에포크(epoch) 동안 트레이닝 데이터세트에 대해서 실행될 수 있으며, 여기서 트레이닝 데이터세트가 반복해서 모델에 제공됨으로써 그 결과를 개선한다. 예를 들어, 지도 학습 기간에는, 모델이 특정 입력 세트에 대한 출력을 예측하도록 개발되고, 여러 에포크 동안 평가되어서, 트레이닝 데이터세트의 가장 많은 수의 입력에 대한 특정 입력에 대응하는 것으로 명시되는 출력을 더 신뢰성 있게 제공한다. 다른 예에서, 비지도 학습 기간 동안, 모델은 데이터세트를 n개의 그룹으로 클러스터하도록 개발되고, 특정 입력을 특정 그룹에 어떻게 정확하게 배치할지에 대해 그리고 n개의 원하는 클러스터를 각각의 에포크에 어떻게 신뢰성 있게 생산할지에 대해 여러 에포크 동안 평가된다.

[0026] 다양한 측면에서, 각 트레이닝 기간 뿐만 아니라 교차 검증이 적용되며, 여기서 트레이닝 데이터세트의 일부가 평가 데이터세트로서 사용된다. 예를 들어, 트레이닝 데이터세트는 k개의 세그먼트로 분할될 수 있으며, 여기에서 (k-1)개의 세그먼트가 트레이닝 에포크들에서 사용되고, 나머지 세그먼트는 트레이닝된 모델이 얼마나 잘 수행되었는지를 결정하는 데 사용된다. 이러한 방식으로, 각 모델은 입력 파라미터의 모든 사용 가능한 조합에 대해 트레이닝되어서, 각각의 모델이 k회 트레이닝되고, 에포크 동안의 자신의 평균 성능에 기초해서 최상의 모델 파라미터가 선택된다.

[0027] 일단 에포크가 실행되면, 모델이 평가되고 이들의 변수의 값이 조정되어서 모델을 더 개선하는 시도를 한다. 다양한 측면에서, 평가는 오류 부정에 대해 바이어싱되거나, 오류 긍정에 대해 바이어싱되거나, 혹은 심지어 그 모델의 전체 정확성에 대해서 바이어싱된다. 이 값은, 사용되는 머신 학습 기술에 따라서 여러 가지 방법으로 조정될 수 있다. 예를 들어, 유전 알고리즘이나 진화 알고리즘에서, 원하는 출력을 예측하는 데 가장 성공적인 모델의 값은, 후속하는 에포크 동안 사용할 모델의 값을 개발하는 데 사용되며, 이는 랜덤한 변이/돌연변이를 포함해서 추가적인 데이터 포인트를 제공할 수 있다. 당업자는 본 개시와 함께 적용될 수 있는, 선형 회귀, 랜덤 포레스트, 의사 결정 트리 학습, 신경망 등을 포함한 다양한 다른 머신 학습 알고리즘에 익숙할 것이다.

[0028] 이 모델은 입력에 영향을 미치는 하나 이상의 변수의 값을 원하는 결과에 더 가깝게 매핑하도록 변경함으로써 여러 에포크에 걸쳐서 규칙 또는 알고리즘을 개발하지만, 트레이닝 데이터세트는 변경될 수 있어서 매우 큰 것이 바람직하기 때문에, 완전한 정확성 및 정밀성은 달성될 수 없다. 따라서 학습 기간을 이루는 다수의 에포크는 특정 횟수의 시험 혹은 고정된 시간/컴퓨팅 예산으로서 설정될 수도 있고, 혹은 특정 모델의 정확도가 충분히 높거나 혹은 낮거나, 혹은 횟수/예산에 도달되기 전에 확성 안정 상태(accuracy plateau)가 달성되었다면 숫자/예산에 도달하기 전에 종료될 수도 있다. 예를 들어, 학습 기간은 n개의 에포크를 실행하고 적어도 95%의 정확도를 가진 모델을 생성하도록 설계되고, 이 모델이 n번째 에포크 이전에 생성된다면, 학습 기간은 조기에 종료될 수 있으며 최종 목표 정확도 임계값을 만족하는 생성된 모델을 사용할 수 있다. 유사하게, 주어진 모델이 랜덤 기회 임계값(예를 들어, 이 모델의 정확도는 특정 입력에 대한 참/거짓 출력을 결정할 때 단지 55%)을 만족시키기에는 부정확하다면, 학습 기간의 다른 모델은 트레이닝을 계속할 수 있지만 해당 모델의 학습 단계는 조기에 종료될 수 있다. 유사하게 특정 모델이 유사한 정확성을 계속 제공하거나 혹은 다수의 에포크에 걸쳐서

그 결과를 자주 변경하는 경우 - 성능 안정 상태에 이미 도달함 - , 에포크 횟수/컴퓨팅 예산이 도달되기 전에 특정 모델의 학습 기간을 종료될 수 있다.

[0029] 학습 기간이 완료되면, 모델이 완성된다. 완성된 모델은 테스팅 기준에 따라 평가된다. 제 1 예시에서는, 입력에 대한 알려진 출력을 포함하는 테스팅 데이터세트가 완료된 모델에 제공되어서, 트레이닝되지 않은 데이터를 핸들링할 때 이 모델의 정확성을 결정한다. 제 2 예시에서, 완성 이후 모델을 평가하는데 오류 긍정 비율, 오류 부정 비율이 사용될 수 있다. 제 3 예시에서, 데이터의 클러스터에 대한 가장 명확한 경계를 생성하는 모델을 선택하는 데 클러스터링 사이의 묘사(delineation)가 사용된다. 다른 예시에서, 정확도 및 리콜 곡선(precision and recall curves) 하의 영역과 같은, 모델의 추가 메트릭이 평가된다.

[0030] 개발 모델(180)(및 따라서 생산 모델(120))은 관리 사용자에 의해 만들어진 선택에 기초해서 모델 피쳐 구성자(185)에 의해 초기에 개발된 예측 모델이다. 관리 사용자는 온라인 서비스(110)의 장치에 대해 청취되는 보안 신호(115)의 하나 이상의 피쳐를 선택하고, 이들 피쳐가 특정 보안 신호(115)가 악의적인지 또는 정상인지를 나타내기 위해서 어떻게 분석되어야 하는지를 선택한다. 다양한 측면에서, 이 피쳐는, 새로운 모델(180)에 대한 피쳐 세트를 정의하기 위해서 관리 사용자가 선택할 수 있는 구조화된 텍스트 파일(예를 들어, XML(Extensible Markup Language) 또는 JSON(JavaScript Object Notation) 태그)로 제공된다. 피쳐 구성에 기초해서, 피쳐는 장치에 대한 특정 보안 신호의 세트로부터 피쳐 벡터로서 유동적으로 추출된다. 다양한 모델에 대해서 이들의 각각의 피쳐 구성에 따라 다양한 피쳐가 추출될 수 있다. 따라서 구조화된 텍스트 파일을 통해서 관리 사용자는 피쳐를 추가 혹은 수정할 수 있고, 코드베이스에 코드를 추가하거나 수정하지 않고도 피쳐를 검사하고 모델에 대해서 검사될 수 있으며, 구조화된 텍스트 파일은 관리 사용자가 선택하는 새로운 피쳐 타입을 전달하기 위해서 개발자에 의해 확장 혹은 수정될 수 있는 코드베이스로부터 코드 세그먼트를 호출한다. 예를 들어, 관리 사용자는 보안 신호(115)로부터 특정 파라미터 또는 데이터 필드와 함께 사용될 피쳐 검사 타입으로서 데이터세트의 개별 값의 수(Count), 데이터세트의 최대 값(Max), 리스트 내에서 가장 자주 나타나는 값의 수(MaxCount), 리스트 내에서 한계를 넘어가지 않는 값의 최대 값(MaxSum) 등을 선택할 수 있다. 보안 신호에서 관찰할 데이터 필드/파라미터의 예는, 신호 타입(예를 들어, 데이터 추출, 로그온 시도, 특정 파일에 대한 액세스 요청), 사용되는 포트, 프로세스/통신에 사용되는 바이트, 특정 인터넷 프로토콜(IP) 주소 및 포트 투플(port tuple)과 서로 전송하는 바이트, 사용자 식별자, 특정 IP 주소나 액션이 블랙리스트 또는 화이트리스트에 있는지 여부을 포함하지만, 이것으로 한정되는 것은 아니다.

[0031] 모델 평가기(175)는 개발 모델(180)을 평가해서 보안 시스템(100)에서 어떤 모델이 생산 모델(120)로서 사용될 것인지를 결정하도록 구성된다. 다양한 측면에서, 생산 모델(120)이 평가를 위해서 개발 모델(180)에 다시 포함되거나 혹은 생산 모델(120)의 정확도 임계값이 특정 생산 모델(120)을 개발 모델(180)로 대체할지 여부를 결정하는데 사용된다. 다른 측면에서, 개발 모델(180)은 예를 들어, 정확성, 정확도 및 리콜 곡선(precision and recall curves) 하의 영역 등과 같은 다른 메트릭과 관련해서 생산 모델(120)과 비교되며, 여기서 최상의 모델이 생산 모델(120)로서 사용할 프로모트 모델(190)로서 선택된다. 모델 평가기(175)가 악의적 신호를 악의적인 것으로 정상 신호를 정상적인 것으로 적절하게 식별하는데 있어서의 자신의 유효성을 결정함에 따라서, 모델은 개발 모델(180)로부터 생산 모델(120)로 연속해서 프로모트될 수 있다(그리고 생산 모델(120)로부터 개발 모델(180)로 디모트될 수 있다). 다양한 측면에서, 상위 n개 최대 정확 개발 모델(180) 또는 정확성 임계값을 초과하는 모든 개발 모델(180)이 프로모트 모델(190)로서 생산 모델(120)에 프로모트된다. 다른 측면에서, 예를 들어, 다른 모델이 보안 신호(115)의 특정한 피쳐를 모니터하지 않을 때와 같이, 관리 사용자는 개발 모델(180)을 생산 모델(120)에 수동으로 프로모트할 수 있다.

[0032] 보안 시스템(100), 모델 트레이닝 및 선택 시스템(105) 및 이들의 각각의 구성 요소는, 비한정의 예로서 데스크톱 컴퓨터 시스템, 유선 및 무선 컴퓨팅 시스템, 모바일 컴퓨팅 시스템(예를 들어, 모바일 전화기, 넷북, 태블릿 혹은 슬레이트 형 컴퓨터, 노트북 컴퓨터 및 랩톱 컴퓨터), 핸드-헬드 장치, 멀티프로세서 시스템, 마이크로프로세서 기반 또는 프로그래밍 가능한 소비자 전자 제품, 미니 컴퓨터, 프린터 및 메인프레임 컴퓨터를 포함하는 다수의 컴퓨팅 시스템을 나타낸다. 이러한 컴퓨팅 시스템의 하드웨어는 도 4와 관련하여 보다 상세하게 설명될 것이다.

[0033] 설명을 위해서, 보안 시스템(100)과 모델 트레이닝 및 선택 시스템(105)의 구성 요소가 서로 별개로 도시되어 있지만, 이들 장치 중 하나 이상의 다른 예시된 장치에 로컬하게 호스팅되는 다수의 구성이 가능하며, 각각의 예시된 장치는 이 장치의 다수의 실례를 나타낼 수 있다. 도 1a 및 1b에 도시된 구성 요소 사이에는 이들 시스템 사이의 통신을 라우팅하기 위해서 당업자에게 익숙한 다양한 서버 및 중간 장치가 놓일 수 있지만, 본 개시

의 신규한 측면으로부터 벗어나지 않기 위해서 이들의 도시는 생략했다.

[0034] 도 2는 온라인 서비스(110)를 보안하는 데 사용하기 위한 예측 모델을 트레이닝하는 데 사용하기 위한 트레이닝 데이터세트를 개발하기 위한 예시적인 방법(200)에 포함된, 일반적인 단계를 도시하는 흐름도이다. 방법(200)은 동작 210에서 시작되며, 여기서 보안 신호(115)가 수집된다. 다양한 측면에서, 보안 신호(115)는 실시간으로(또는 처리 및 전송 지연을 고려하면 거의 실시간으로) 수신될 수도 있고 혹은 예를 들어 일괄 처리에서 m분마다 보안 이벤트를 리뷰하는 등의 주기적인 리뷰를 위해서 수신되어서 캐시될 수도 있다. 보안 신호(115)는 온라인 서비스(110) 내의 머신에서 발생하는 다양한 액션의 청취 이벤트 및 파라미터를 포함한다.

[0035] 청취 이벤트 및 파라미터는 동작 220에서 특정 보안 신호(115)가, 악성인 혹은 정상인 액션에 대응하는지 여부를 식별하는 데 사용된다. 다양한 측면에서, 수집된 보안 신호(115)는 라이브 온라인 서비스(110)(즉, 생산 모델(120))와 함께 사용하도록 지정된 예측 모델에 공급되어 각 보안 신호(115)가 악의적인 것인지 정상적인 것지 결정한다. 이러한 결정은 분석가 사용자에게 제시되며, 이 사용자는 이 결정에 대해 액션을 취해서 악의적인 제3자에 대해 온라인 서비스(110)를 보안할 수 있지만, 이 판정이 오류 긍정 혹은 오류 부정을 포함한다는 것을 나타내는 예측 모델에 의해 행해지는 판정을 철회할 수도 있다. 유사하게, 자동화된 공격자(140)가 온라인 서비스(110)에 대한 공격을 시뮬 레이션하는 데 사용되는 측면에서, 자동화된 공격자(140)는 공격이 악의적인 것에 응답해서 생성되는 보안 신호(115)를 식별하는 통지를 제공함으로써 이들 보안 신호(115)는 예측 모델로부터의 탐지 결과에 관계없이 악의적인 것으로 취급된다.

[0036] 동작 230에서, 롤링 윈도우(130)는 온라인 서비스(110) 상에서 실행되는 가장 최근의 악용 및 공격에 관련된 보안 신호(115)를 분석하는 현재 시간으로부터 시간 프레임을 정의하도록 설정된다. 롤링 윈도우(130)는 현재 시간으로부터 지정된 시간 동안 내에 속하는 보안 신호(115)의 세트를 정의하고, 보안 신호(115)는 지난 d일 동안 수집된 것이다. 다수 일자의 윈도우는 종래의 보안 시스템에 의한 탐지를 방지하기 위해서 며칠 동안 수행되는 슬로우 공격을 트레이닝하고 예측하는데 사용된다. 보안 신호(115)가 수집됨에 따라서 최근 보안 신호가 롤링 윈도우(130) 동안 보안 신호(115)의 세트에 추가되며 롤링 윈도우(130)의 지정 기간 이전에 수집된 보안 신호(115)는 보안 신호(115)의 세트로부터 연속해서 제거된다.

[0037] 일부 측면에서 단계 240에서 이력 신호(135)가 옵션으로 수신된다. 이력 신호(135)는 이전에 관찰된 보안 신호(115)로부터 분석가 사용자에 의해서 큐레이트되어서, 롤링 윈도우(130) 기간 내에 유사한 공격 혹은 사용 경우가 발견되었는지 여부에 관계없이 트레이닝의 목적으로 지정된 특정한 공격 타입 혹은 정상 사용 경우를 나타내는 이력적으로 중요한 보안 신호(115)를 포함하게 한다. 일례로, 이력적으로 위험한 악용은 이 악용에 대해 지속적으로 경계를 유지하기 위한 이력 신호(135)에 추가된 탐지와 관련된 보안 신호(115)를 가질 수 있다. 또 다른 예시로, 개발자는 제로 데이 약용을 발견했는데 이것을 악의적인 제3자가 아직 사용하고 있는지 여부는 알지 못할 수 있으며, 제로 데이 악용의 액션을 모방하는 예시적인 보안 신호(115)를, 악용이 보이지 않는 경우라도, 이 악용을 사전에 경계하기 위한 이력 신호(135)로서 사용하기 위해서 제공한다. 또 다른 예에서, 오류 긍정을 빈번하게 야기하는 보안 신호(115)가 이력 신호(135)에 추가되어서, 예측 모델이 특정 보안 신호(115)에 대해 트레이닝되는 것을 보장한다. 이력 신호(135)는 이용 가능하다면, 롤링 윈도우(130) 내에서 수집되는 보안 신호(115)의 세트에 추가된다.

[0038] 동작 250으로 넘어가서, 방법(200)은 롤링 윈도우(130) 동안 수집되는 악의적 신호와 정상 신호 및 옵션인 동작 240에서 세트에 추가된 임의의 이력 신호(135)를 밸런싱한다. 악의적 신호를 밸런싱할 때, 각각의 공격 타입을 나타내는 신호의 상대적인 양이 평형화(equilibrium)를 거치도록(즉 균등화됨(equalized)), 각각의 신호의 공격 타입이 결정되고, 그 결과 어떠한 특정한 공격 타입도 악의적 신호의 비율에서 과대 표시되거나 과소 표시되지 않는다. 정상 신호를 밸런싱할 때, 롤링 윈도우(130) 동안 악의적 신호를 생산한 장치로부터 수신된 정상 신호는 폐기되고, 온라인 서비스(110)의 각각의 타입의 장치로부터 수신한 정상 신호의 상대적인 양은 평형화를 거치며, 그 결과 어떠한 특정한 공격 타입도 정상 신호의 비율에서 과대 표시되거나 과소 표시되지 않는다.

[0039] 나아가, 악의적 신호의 세트의 수가 정상 신호의 세트보다 더 적을 것으로 예상됨으로써, 악의적 신호와 크로스 조인되어서 새로운 더 큰 악의적 신호의 세트를 생성하도록, 정상 신호의 세트의 일부가 동작 260에서 선택될 수 있으며, 그 결과 2개의 세트는 악의적 신호에 대한 정상 신호의 원하는 비율을 포함할 것이다. 다양한 측면에서 일단 악의적 신호의 세트와 정상 신호의 세트는 소정의 비율이 되면(예컨대, 평행화), 2개의 세트는 트레이닝 세트로서 함께 사용된다.

[0040] 동작 270에서, 롤링 윈도우(130)(및 임의의 이력 신호(135))에서 발생하는 밸런싱된 악의적 신호 및 정상 신호로 이루어진 다양한 공격 시나리오의 트레이닝 세트가 트레이닝 예측 모델에 이용 가능하게 된다. 예를 들어,

온라인 서비스(110)에 대해 활발하게 사용되고 있는 공격 및 악용을 보다 잘 평가하기 위해서 롤링 윈도우(130)의 컨텐츠가 시간에 따라 업데이트됨으로써, 보안 신호(115)를 분석하는 데 사용되는 생산 모델(120)은 지속적으로 재트레이닝되고 및/또는 다른 예측 모델로 대체된다. 따라서, 이 방법(200)은 동작 270 이후에 종결되거나 또는 동작 210으로 돌아가서 롤링 윈도우(130)에 기초해서 트레이닝 데이터세트를 주기적으로 또는 지속적으로 공급하기 위해서, 보안 신호(115)를 계속 수집한다.

[0041] 도 3은 온라인 서비스(110)를 보안하는데 사용하기 위한 예측 모델을 트레이닝시키고 선택하는 예시적인 방법(300)에 포함되는 일반적인 단계를 나타내는 흐름도이다. 방법(300)은 동작 310에서 시작되며, 여기서 방법 200에 따라 개발된 것과 같은, 밸런싱된 악의적 신호와 정상 신호의 트레이닝 데이터세트가 수신된다. 다양한 측면에서, 롤링 윈도우(130)(및 따라서 트레이닝 데이터세트) 또는 사용자 커맨드에 대한 업데이트에 응답해서, 방법(300)은 주기적으로(예를 들어, h시간 마다) 호출된다.

[0042] 동작 320으로 넘어가서, 트레이닝 데이터세트는 평가 서브 세트와 학습 서브 세트로 분할된다. 다양한 측면에서, 트레이닝 데이터세트에 대한 평가 서브 세트의 크기는 달라질 수 있지만, 일반적으로 학습 서브 세트보다 크기가 작다. 예를 들어, 학습 서브 세트는 초기 트레이닝 세트의 3분의 1이 될 수 있고, 따라서, 학습 서브 세트는 초기 트레이닝 세트의 3분의 2가 될 것이다. 당업자라면 트레이닝 데이터세트의 다른 비율이 평가 서브 세트로서 사용되도록 분할될 수도 있다는 것을 이해할 것이다.

[0043] 동작 330에서, 개발 모델(180)을, 온라인 서비스(110)를 보안하기 위해 생산에서(즉, 생산 모델(120)로서) 사용될 잠재적인 예측 모델로서 생성하기 위한 구성 피처가 수신된다. 보안 분석가와 같은 관리 사용자는 보안 신호(115)를 통해서 온라인 서비스(110)에서 청취될 하나 이상의 파라미터를 선택하고, 이들 파라미터를 시험하는데 사용할 피처 타입을 선택한다. 보안 신호(115)는 이벤트 로그, 네트워크 트레이스, 에러 리포트, 특수 이벤트 리스너 리포트, 원자 검출 및 이들의 조합을 포함하지만 이것으로 한정되는 것은 아니며, 선택된 피처의 파라미터는 보안 신호(115)에 포함된 임의의 요소를 포함할 수 있다.

[0044] 예를 들어, 보안 신호(115)가 네트워크 트레이스를 포함하는 경우, 송신기/수신기 어드레스 쌍의 파라미터는 "카운트" 피처 타입에 따라서 선택되어 평가될 수 있으며, 그 결과 이 쌍이 트레이닝 세트 내에서 보여지는 횟수는 대응 피처를 평가할 스코어/값을 증가시킨다. 또 다른 예에서, 보안 신호(115)가 네트워크 트레이스를 포함하는 경우, 송신기/수신기 쌍 사이에서 송신되는 바이트 수의 파라미터는 그 피처를 평가할 스코어/값로서 제공된다. 또 다른 예에서, 상대적인 업로드/다운로드 비율을 나타내기 위해 송신기/수신기 쌍 사이의 전송 밸런스의 파라미터가, 그 피처를 평가하기 위한 값/스코어로서 제공된다. 당업자라면, 비한정의 예로서, 다른 파라미터 및 이들이 예측 모델에 의해 평가되는데 사용될 수 있는 이들 파라미터의 다른 피처 값이 본 출원과 함께 사용되는 것을 상정할 수 있을 것이다.

[0045] 동작(340)으로 넘어가서, 수신한 피처 구성에 기초해서 개발 모델(180)이 생성되고, 하나 이상의 머신 학습 알고리즘을 가진 학습 서브 세트에 따라 개선된다. 각 예측 모델은 특정한 피처 벡터(관리 사용자에 의해 선택된 피처를 명시함)를 수용하도록 생성되며 피처 벡터를 포함하는 여기서 각각의 피처는 계수와 관련된다. 각각의 피처 벡터는 피처 구성에 기초해서 보안 신호(115)로부터 유동적으로 추출된다. 계수의 값은 머신 학습 알고리즘의 다양한 에포크에 걸쳐서 조정되어서, 그 결과 특정 개발 모델(180)이 피처 벡터의 입력을 수신할 때, 학습 서브 세트에 지정된 출력에 매칭되게 악의적 신호 또는 정상 신호의 출력을 신뢰성 있게 생성하도록 다양한 피처 값 사이의 인터렉션이 조정될 수 있다.

[0046] 동작(350)으로 넘어가서, 동작(340)의 트레이닝 데이터세트와 관련하여 개선된 예측 모델이 동작(320)의 트레이닝 데이터세트로부터 분리된 평가 서브 세트에 대해 평가된다. 평가 서브 세트는 신호의 출력이 악의적인지 혹은 정상적인지 여부를 알고 있는 입력(온라인 서비스(110)로부터 수집된 보안 신호(115))을 포함한다. 나아가, 평가 서브 세트의 입력/출력 쌍이 개발 모델(180)을 직접 트레이닝하는 데 사용되지 않았기 때문에, 개발 모델(180)이 미지의 신호가 악의적인 것인지 혹은 정상적인 것인지 여부를 결정하기 위한 일반적인 기능적 규칙을 제공할지 여부에 대한 테스트를 제공한다.

[0047] 특정 개발 모델(180)을 생산 모델(120)로 프로모트할지 여부를 결정하기 위해서 프로모션 임계값이 개발 모델(180)에 적용된다. 프로모션 임계값은 보안 신호(115)로부터 추출된 피처 벡터에 기초해서 신호가 악의적인 것인지 또는 정상적인 것인지 여부를 예측하는데 개발 모델(180)이 얼마나 정확하게 되어야 하는지를 명시한다. 일부 측면에서, 프로모션 임계값은 예를 들어, 적어도 n%의 정확도, 테스트 데이터에 대한 정확도 및 리콜 곡선(precision and recall curves) 하의 영역 등과 같은, 상수로서 설정된다. 다른 측면에서, 프로모션 임계값은 특정한 피처 벡터 또는 공격 타입에 대한 현재 생산 모델(120)의 정확도에 의해 설정되며, 이로써 개발 모델

(180)은 보안 시스템(100)의 생산 모델(120)을 대체하기 위해서 개발 모델(180)은 현재 생산 모델(120)보다 더 정확해야 한다.

[0048] 동작(360)에서, 평가 서브 세트 및 프로모션 임계치에 따라 가장 잘 수행되는 모델(180) 및 재평가된 생산 모델(120)은, 온라인 서비스(110)를 보호하기 위해 보안 시스템(100)에 의한 사용을 위해 프로모션된다. 프로모션 임계값을 더 이상 만족하지 않거나 혹은 개발 모델(180)에 의해서 대체된 생산 모델(120)은, 삭제되거나 혹은 추가적인 트레이닝 혹은 수정을 위해서 및 이후 재평가를 위해서 개발 모델(180)로 디모트될 수 있다. 이후 이 방법(300)은 종료될 수 있다.

[0049] 컴퓨터에서 실행되는 애플리케이션 프로그램과 함께 실행되는 프로그램 모듈의 전반적인 맥락에서 구현예를 설명했지만, 측면들은 당업자라면 다른 프로그램 모듈과 조합해서 구현될 수도 있다는 것을 이해할 것이다. 일반적으로 프로그램 모듈은 루틴, 프로그램, 컴포넌트, 데이터 구조 및 특정 작업을 수행하거나 특정한 추상 데이터 유형을 구현하는 기타 타입의 구조를 포함한다.

[0050] 본 명세서에 개시된 측면 및 기능은, 데스크톱 컴퓨터 시스템, 유선 및 무선 컴퓨팅 시스템, 모바일 컴퓨팅 시스템(예를 들어, 모바일 전화기, 넷북, 태블릿 혹은 슬레이트 형 컴퓨터, 노트북 컴퓨터 및 랩톱 컴퓨터), 핸드-헬드 장치, 멀티프로세서 시스템, 마이크로 프로세서 기반 또는 프로그래밍 가능한 소비자 전자 제품, 미니 컴퓨터, 프린터 및 메인프레임 컴퓨터를 포함하는 다수의 컴퓨팅 시스템을 통해서 동작할 수 있다.

[0051] 나아가, 일 측면에 따르면, 본 명세서에 개시된 측면 및 기능은 분산형 시스템(예를 들어, 클라우드 기반 컴퓨팅 시스템)을 통해서 동작하며, 여기서 애플리케이션 기능, 메모리, 데이터 저장 및 검색 및 다양한 프로세싱 기능은 인터넷 또는 인트라넷과 같은 분산 컴퓨팅 네트워크를 통해 서로 이격되어 동작될 수 있다. 일 특징에 따르면, 다양한 타입의 사용자 인터페이스 및 정보가 온-보드 컴퓨팅 장치 디스플레이를 통해 또는 하나 이상의 컴퓨팅 장치와 관련된 원격 디스플레이 장치를 통해 디스플레이된다. 예를 들어, 사용자 인터페이스 및 다양한 타입의 정보가, 사용자 인터페이스 및 다양한 타입의 정보가 투영되는 벽면 상에 표시되어서 인터렉션된다. 구현예가 실행되는 다수의 컴퓨팅 시스템과의 인터렉션은 키 스트로크 엔트리, 터치 스크린 엔트리, 음성 또는 다른 오디오 엔트리, 제스쳐 엔트리(관련된 컴퓨팅 장치에 컴퓨팅 장치의 기능을 제어하기 위해 사용자 제스처를 캡처해서 해석하는 검출(예를 들어, 카메라) 기능이 구비됨) 등을 포함한다.

[0052] 도 4 및 관련 설명은 예시가 실시되는 다양한 운영 환경에 대한 설명을 제공한다. 그러나, 도 4와 관련하여 예시되고 설명되는 장치 및 시스템은 예시 및 설명을 위한 것이며 본 명세서에 기술된 측면들을 실시하는데 사용될 수 있는 다수의 컴퓨팅 장치 구성을 제한하는 것은 아니다.

[0053] 도 4는 본 개시의 예시가 실시될 수 있는 컴퓨팅 장치(400)의 물리적 컴포넌트(예를 들어 하드웨어)를 도시한 블록도이다. 기본 구성에서, 컴퓨팅 장치(400)는 적어도 하나의 처리 유닛(402) 및 시스템 메모리(404)를 포함할 수 있다. 일 측면에 따라서, 컴퓨팅 장치(400)의 구성 및 타입에 따라서, 시스템 메모리(404)는 휘발성 장치(예컨대, RAM), 비휘발성(예컨대, ROM), 플래시 메모리 또는 이들 메모리의 임의의 조합을 포함할 수 있지만, 이것으로 한정되는 것은 아니다. 일 측면에 따라서, 시스템 메모리(404)는 운영 체제(405) 및 소프트웨어 애플리케이션(4500)을 실행하기에 적합한 하나 이상의 프로그램 모듈(406)을 포함한다. 일 측면에 따라서, 시스템 메모리(404)는 보안 시스템(100), 모델 트레이닝 및 선택 세스템(105) 및 이로써 사용 및 생산되는 임의의 모델을 포함한다. 운영 체제(405)는, 예컨대, 컴퓨팅 장치(400)의 동작을 제어하는데 적합할 수 있다. 또한, 측면들은 그래픽스 라이브러리, 다른 운영 체제, 또는 임의의 다른 애플리케이션 프로그램과 함께 실시될 수 있지만, 어떠한 특정 애플리케이션 또는 시스템으로 제한되는 것은아니다. 이 기본 구성은 도 4에 점선(408) 내에 있는 컴포넌트들로 도시되어 있다. 일 측면에 따라서, 컴퓨팅 장치(400)는 부가적인 특징들 또는 기능을 가질 수 있다. 예를 들어, 컴퓨팅 장치(400)는 또한 예컨대, 자기 디스크, 광학 디스크, 또는 테이프와 같은 부가적인 데이터 저장 장치(착탈식 및/또는 비착탈식)를 포함할 수 있다. 이러한 부가적인 저장부는 도 4에 착탈식 저장 장치(409) 및 비착탈식 저장 장치(410)로 도시되어 있다.

[0054] 전술한 바와 같이, 다수의 프로그램 모듈 및 데이터 파일이 시스템 메모리(404)에 저장된다. 프로그래밍 모듈(406)(예컨대, 보안 시스템(100), 모델 트레이닝 및 선택 시스템(105))은 프로레싱 유닛(402)에서 실행되는 동안, 도 2 및 도 3에 각각 도시된 방법(200, 300)의 스테이지 중 하나 이상을 포함하는 프로세스를 수행할 수 있지만 이것으로 한정되는 것은 아니다. 일 측면에 따라서, 다른 프로그램 모듈이 예시에 따라서 사용되며, 전자 메일 및 컨택 애플리케이션, 워드 프로세싱 애플리케이션, 스프레드시트 애플리케이션, 데이터베이스 애플리케이션, 슬라이드 프리젠테이션 애플리케이션, 드로잉 또는 컴퓨터 지원 애플리케이션 프로그램 등을 포함한다.

[0055] 일 실시예에 따라서, 컴퓨팅 장치(400)는 키보드, 마우스, 펜, 사운드 입력 장치, 터치 입력 장치 등과 같은 하나 이상의 입력 장치(들)(412)를 구비할 수 있다. 일 측면에 따라서 디스플레이, 스피커, 프린터 등과 같은 출력 장치(들)(414)도 포함될 수 있다. 앞서 설명한 장치들은 예시이고 다른 장치도 사용될 수 있다. 컴퓨팅 장치(400)는 다른 컴퓨팅 장치(418)와의 통신을 가능하게 하는 하나 이상의 통신 연결부(416)를 포함할 수 있다. 적절한 통신 연결부(416)의 예는 RF 송신기, 수신기 및/또는 송수신 회로, 범용 직렬 버스(USB), 병렬 및/또는 직렬 포트를 포함하나 이에 국한되지는 않는다.

[0056] 본 명세서에서 사용되는 컴퓨터 판독가능 매체라는 용어는 컴퓨터 저장 매체를 포함한다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 또는 프로그램 모듈과 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 착탈식 및 비착탈식 매체를 포함할 수 있다. 시스템 메모리(404), 착탈식 저장 장치(409) 및 비착탈식 저장 장치(410)는 모두 컴퓨터 판독가능 저장 매체의 예들(즉, 메모리 저장부)이다. 일 측면에 따라서, 컴퓨터 저장 매체는, 제한적인 것은 아니지만, RAM, ROM, EEPROM, 플래시 메모리 또는 기타 메모리 기술, CD-ROM, DVD(digital versatile disk) 또는 다른 광학 저장부, 자기 카세트, 자기 테이프, 자기 디스크 저장부 또는 기타 자기 저장 장치, 또는 정보를 저장하는데 사용될 수 있으며 컴퓨팅 장치(400)에 의해 액세스될 수 있는 임의의 다른 제품을 포함할 수 있다. 일 측면에 따라서, 임의의 그러한 컴퓨터 저장 매체는 컴퓨팅 장치(400)의 일부일 수 있다. 컴퓨터 저장 매체는 반송파 또는 다른 전파 또는 변조된 데이터 신호는 포함하지 않는다.

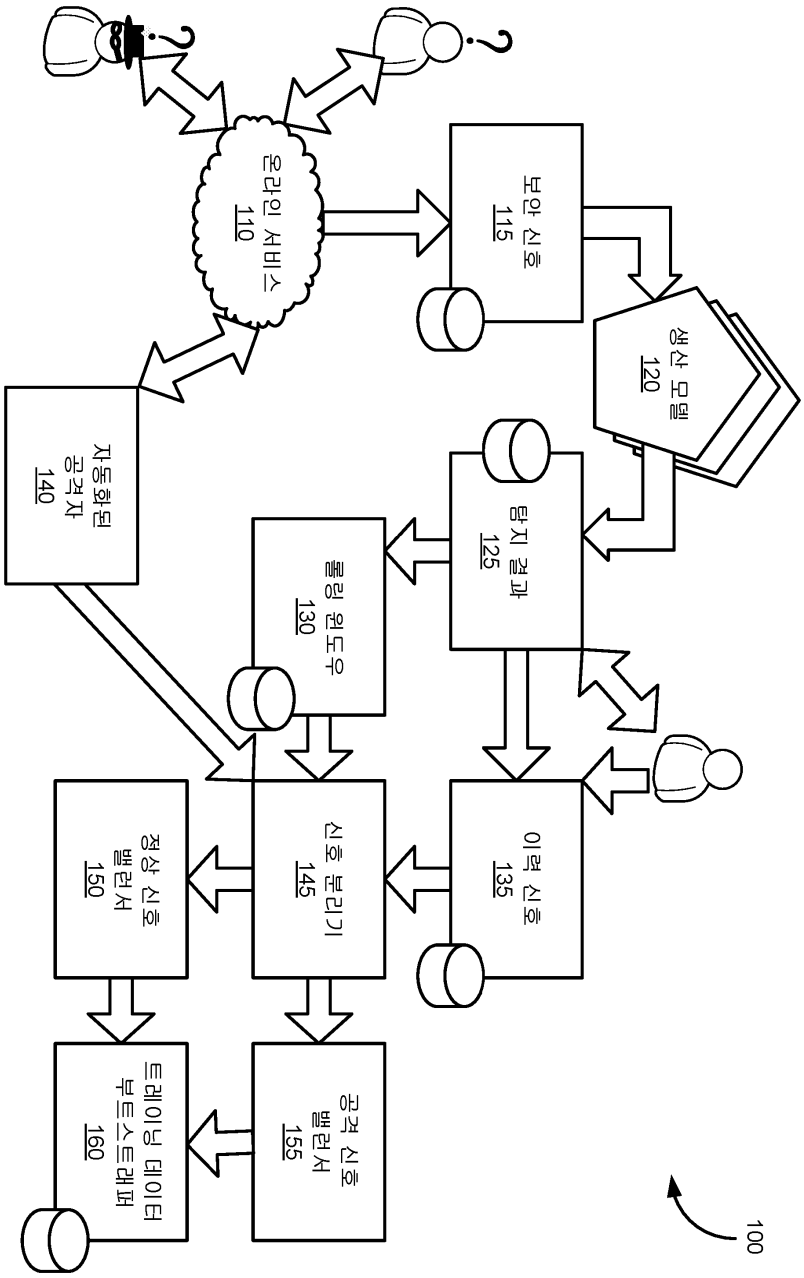
[0057] 일 측면에 따라서, 통신 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈, 또는 반송파나 다른 전송 메커니즘과 같은 변조된 데이터 신호 내의 기타 데이터를 포함할 수 있고, 임의의 정보 전달 매체를 포함한다. 일 측면에 따라서, "변조된 데이터 신호"란 용어는 정보를 신호에 인코드하는 것과 같은 방식으로 설정 또는 변경된 하나 이상의 특징을 갖는 신호를 의미한다. 예를 들어, 제한적인 것은 아니지만, 통신 매체는 유선 네트워크 또는 직접 유선 접속과 같은 유선 매체, 음향, RF, 적외선 및 기타 무선 매체들과 같은 무선 매체를 포함할 수 있다.

[0058] 구현예는 예를 들어 측면들의 실시예에 따른 방법, 시스템 및 컴퓨터 프로그램 제품의 블록도 및/또는 동작 예시를 참조하여 설명되었다. 블록에서의 기능/동작은 임의의 흐름도에 도시된 순서를 벗어나 실행될 수 있다. 예를 들어, 연속적으로 도시되어 있는 두 개의 블록은 관련된 기능/동작에 따라 사실 실질적으로 동시에 실행될 수 있거나 또는 블록들은 때때로 역순으로 실행될 수도 있다.

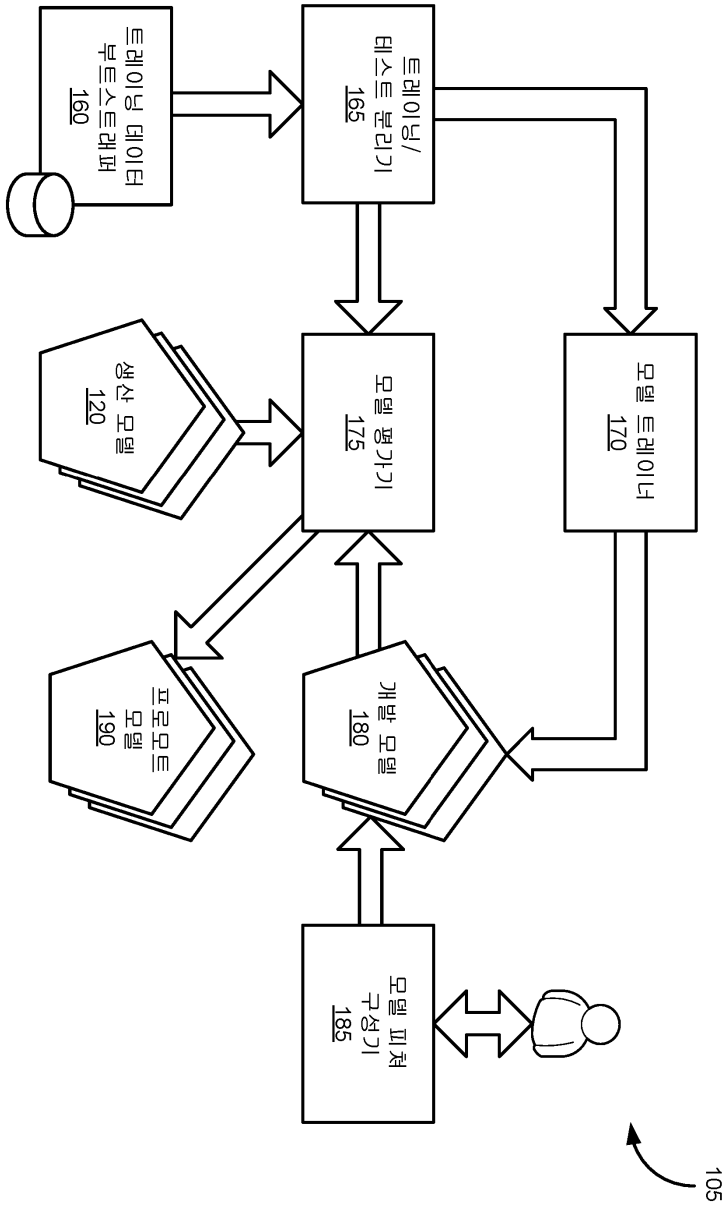
[0059] 본 출원에 제공된 하나 이상의 예시의 설명 및 예시는 청구된 본 발명의 범주를 어떤 식으로든 제한 또는 한정하려는 의도는 없다. 본 출원에서 제공되는 측면, 예시 및 세부사항은 소유권을 전달하고 청구된 발명의 최적의 모드를 구성 및 사용할 수 있도록 하는데 충분한 것으로 간주된다. 구현예는 임의의 측면, 예시, 또는 세부사항에 국한되는 것으로 해석되어서는 안 된다. 함께 또는 개별적으로 도시되고 설명된 것과는 무관하게, (구조적 및 방법론적의) 다양한 특징들은 특정 특징 집합을 갖는 실시예를 구현하기 위해 선택적으로 포함되거나 생략될 수 있다. 본 발명의 설명 및 예시가 제공되었으며, 당업자라면 청구된 발명의 포괄적 범위를 벗어나지 않는 본 출원에서 구현된 일반적인 발명의 개념의 포괄적 측면의 사상 내에 속하는 변형예 및 대안적 실시예를 고안할 수 있다.

도면

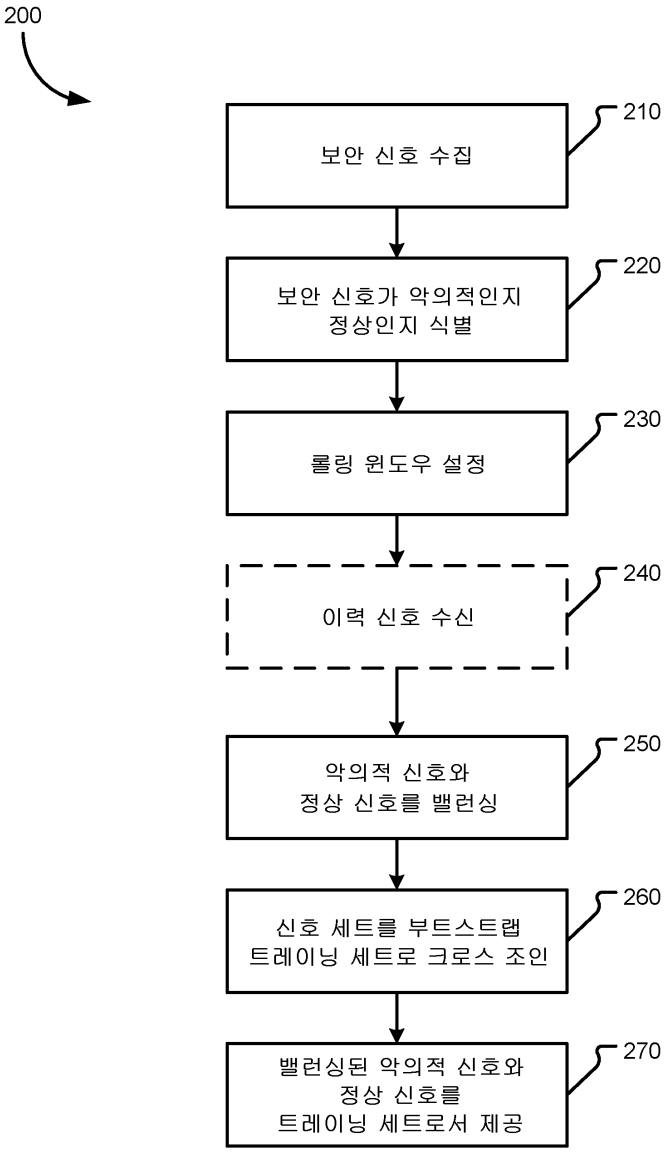
도면1a


온라인 서비스
110
보안 신호
115
생산 모델
120
탐지 결과
125
롤링 윈도우
130
이력 신호
135
신호 분리기
145
정상 신호
밸런서
150
공격 신호
밸런서
155
트레이닝 데이터
부트스트래퍼
160
자동화된
공격자
140
100

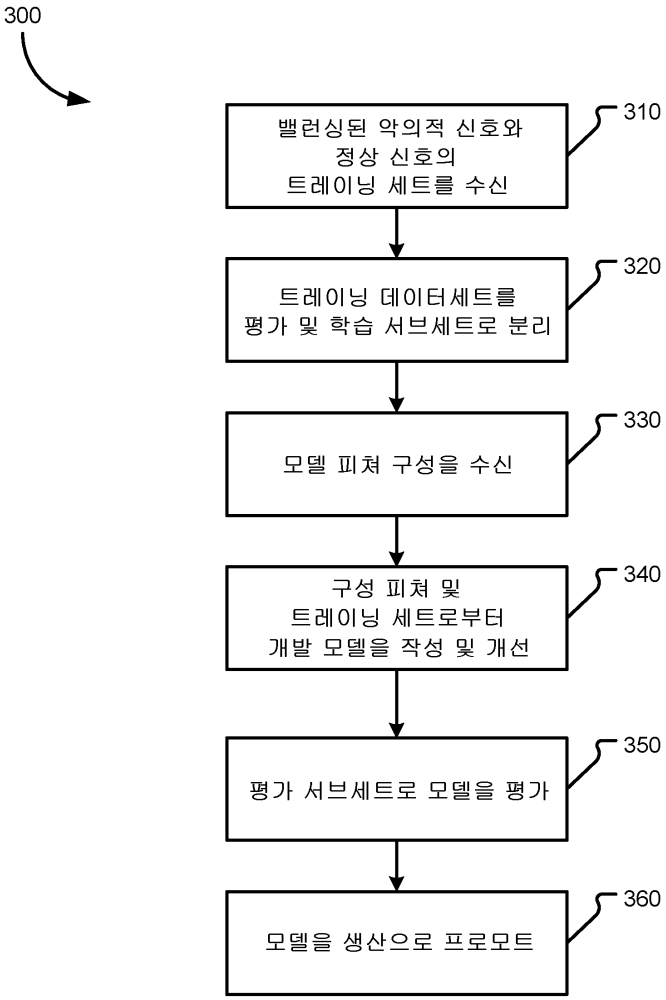
도면1b


트레이닝 데이터 부트스트래퍼 160
트레이닝/테스트 분리기 165
생산 모델 120
모델 평가기 175
모델 트레이너 170
프로모트 모델 190
개발 모델 180
모델 피쳐 구성기 185
105

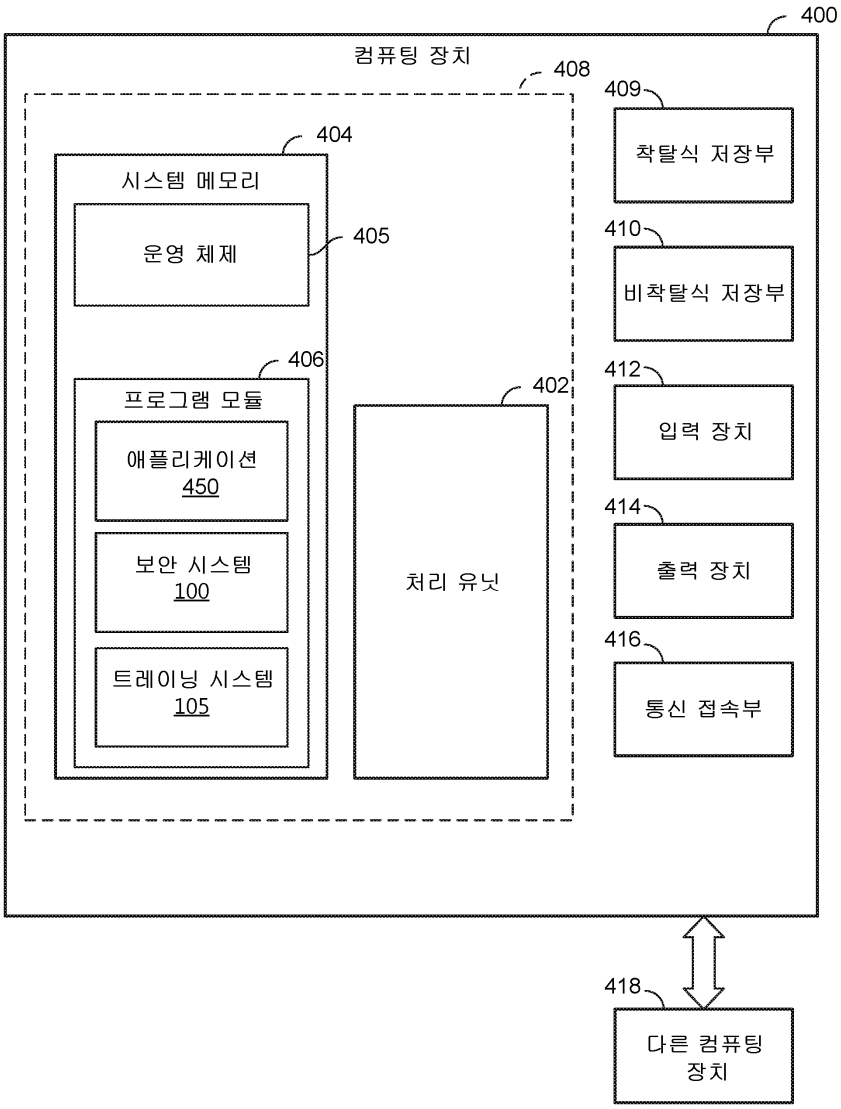
도면2


200
보안 신호 수집
210
보안 신호가 악의적인지
정상인지 식별
220
롤링 윈도우 설정
230
이력 신호 수신
240
악의적 신호와
정상 신호를 밸런싱
250
신호 세트를 부트스트랩
트레이닝 세트로 크로스 조인
260
밸런싱된 악의적 신호와
정상 신호를
트레이닝 세트로서 제공
270

도면3


300
밸런싱된 악의적 신호와
정상 신호의
트레이닝 세트를 수신
310
트레이닝 데이터세트를
평가 및 학습 서브세트로 분리
320
모델 피쳐 구성을 수신
330
구성 피쳐 및
트레이닝 세트로부터
개발 모델을 작성 및 개선
340
평가 서브세트로 모델을 평가
350
모델을 생산으로 프로모트
360

도면4


400
컴퓨팅 장치
408
409
착탈식 저장부
404
시스템 메모리
운영 체제
405
410
비착탈식 저장부
406
프로그램 모듈
402
412
입력 장치
애플리케이션
450
414
출력 장치
보안 시스템
100
처리 유닛
416
통신 접속부
트레이닝 시스템
105
418
다른 컴퓨팅
장치