(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0067426
(43) 공개일자 2025년05월15일

(51) 국제특허분류(Int. Cl.)
G06F 21/56 (2013.01) *G06F 16/242* (2019.01)
G06F 16/25 (2019.01) *G06F 16/26* (2019.01)
G06N 20/00 (2019.01)
(52) CPC특허분류
G06F 21/56 (2013.01)
G06F 16/243 (2019.01)
(21) 출원번호 10-2023-0153544
(22) 출원일자 2023년11월08일
심사청구일자 2023년11월08일

(71) 출원인
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
신동규
서울특별시 서초구 서운로 62, 109동 702호(서초동, 래미안 리더스원)
신동일
서울특별시 강남구 도산대로85길 24, 603호(청담동, 상지리츠빌10차)
박한솔
인천광역시 동구 화수로 38, 6동 901호(송현동, 동부아파트)
(74) 대리인
양성보

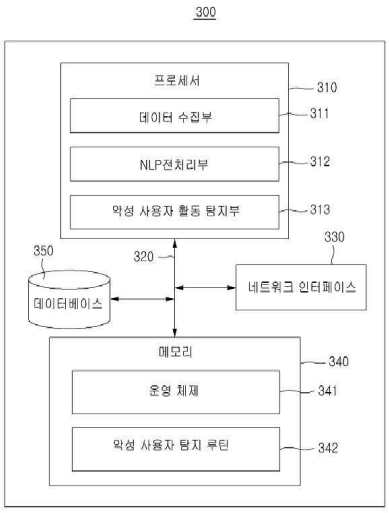
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **악성 사용자 활동 탐지 방법 및 시스템**

(57) 요 약

악성 사용자 활동 탐지 방법 및 시스템이 제시된다. 본 발명에서 제안하는 악성 사용자 활동 탐지 시스템은 클라이언트가 DB API 서버에게 데이터를 요청하고 DB API 서버는 자율 시스템(Autonomous System; AS)의 정보가 담긴 BGP(Border Gateway Protocol) 데이터를 클라이언트에게 전송함으로써 데이터를 수집하는 데이터 수집부, 자율 시스템의 경로를 통해 이상 행동을 탐지하기 위해 상기 수집된 BGP 데이터에 대해 NLP(Natural Language Processing) 전처리 기법을 이용하여 문자형 데이터를 수치형 데이터로 변경하는 NLP 전처리부 및 상기 NLP 전처리 기법을 통해 변경된 수치형 데이터를 AE(Auto-Encoder) 모델에 입력하여 학습함으로써, 자율 시스템의 경로 기록(Trace record)을 미리 저장된 보안사고 사례와 비교 분석하여 정상 데이터와 이상 데이터로 분류하는 악성 사용자 활동 탐지부를 포함한다.

대 표 도 - 도3



(52) CPC특허분류
G06F 16/25 (2019.01)
G06F 16/26 (2019.01)
G06N 20/00 (2021.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711187537
과제번호	2022R1F1A1074773
부처명	교육부
과제관리(전문)기관명	한국연구재단
연구사업명	기본연구
연구과제명	사이버 공간 다차원 모니터링 및 위협 역추적 기술 연구
기 여 율	1/1
과제수행기관명	세종대학교 산학협력단
연구기간	2023.03.01 ~ 2024.02.29

명 세 서

청구범위

청구항 1

클라이언트가 DB API 서버에게 데이터를 요청하고 DB API 서버는 자율 시스템(Autonomous System; AS)의 정보가 담긴 BGP(Border Gateway Protocol) 데이터를 클라이언트에게 전송함으로써 데이터를 수집하는 데이터 수집부;

자율 시스템의 경로를 통해 이상 행동을 탐지하기 위해 상기 수집된 BGP 데이터에 대해 NLP(Natural Language Processing) 전처리 기법을 이용하여 문자형 데이터를 수치형 데이터로 변경하는 NLP 전처리부; 및

상기 NLP 전처리 기법을 통해 변경된 수치형 데이터를 AE(Auto-Encoder) 모델에 입력하여 학습함으로써, 자율 시스템의 경로 기록(Trace record)을 미리 저장된 보안사고 사례와 비교 분석하여 정상 데이터와 이상 데이터로 분류하는 악성 사용자 활동 탐지부

를 포함하는 악성 사용자 활동 탐지 시스템.

청구항 2

제1항에 있어서,

상기 NLP 전처리부는,

토크나이저(Tokenizer)를 통해 문자형 데이터를 수치형 데이터로 변경하기 위해 문자형 데이터의 문자열을 최소 단위로 나누어 시퀀스 분리를 수행하고, 시퀀스 분리를 통해 생성된 분리된 경로(Path) 문자들에 고유한 정수 번호를 부여함으로써 데이터 전처리를 수행하는

악성 사용자 활동 탐지 시스템.

청구항 3

제1항에 있어서,

상기 악성 사용자 활동 탐지부는,

정상 데이터와 이상 데이터에 대한 탐지 정확도를 개선하기 위해 이상 데이터의 수를 증가시켜 정상 데이터와의 비율을 맞추도록 SMOTE(Synthetic Minority Over-sampling Technique)를 이용하는

악성 사용자 활동 탐지 시스템.

청구항 4

제3항에 있어서,

상기 악성 사용자 활동 탐지부는,

상기 SMOTE는 이상 데이터의 샘플을 합성적으로 생성하여 정상 데이터와 이상 데이터의 비율을 맞추기 위해, 임의의 이상 데이터 하나를 중심으로 하여 가장 가까운 이상 데이터를 찾은 후, 상기 중심이 된 이상 데이터와 가장 가까운 이상 데이터 사이에 새로운 데이터를 생성하고, 생성된 데이터를 이상 데이터 클래스에 추가하고, 상기 이상 데이터 클래스에 추가하는 과정을 미리 설정된 조건을 맞출 때까지 반복하는

악성 사용자 활동 탐지 시스템.

청구항 5

제3항에 있어서,

상기 악성 사용자 활동 탐지부는,

AE 모델을 학습하기 전에 데이터의 특성, 편향성, 이상치를 분석하기 위해 데이터의 분포를 시각화하고, 데이터의 군집 구조와 분포를 시각화하여 비 선형적인 데이터 패턴을 포착하기 위해, 비선형 차원 축소 기법인 t-

SNE(t-distributed Stochastic Neighbor Embedded)를 이용하여 이웃 그래프 기반 차원 축소를 수행함으로써, 데이터 간의 유사도에 따라 고차원 데이터를 저차원으로 변환하는

악성 사용자 활동 탐지 시스템.

청구항 6

데이터 수집부를 통해 클라이언트가 DB API 서버에게 데이터를 요청하고 DB API 서버는 자율 시스템(Autonomous System; AS)의 정보가 담긴 BGP(Border Gateway Protocol) 데이터를 클라이언트에게 전송함으로써 데이터를 수집하는 단계;

NLP 전처리부가 자율 시스템의 경로를 통해 이상 행동을 탐지하기 위해 상기 수집된 BGP 데이터에 대해 NLP(Natural Language Processing) 전처리 기법을 이용하여 문자형 데이터를 수치형 데이터로 변경하는 단계; 및

악성 사용자 활동 탐지부가 상기 NLP 전처리 기법을 통해 변경된 수치형 데이터를 AE(Auto-Encoder) 모델에 입력하여 학습함으로써, 자율 시스템의 경로 기록(Trace record)을 미리 저장된 보안사고 사례와 비교 분석하여 정상 데이터와 이상 데이터로 분류하는 단계

를 포함하는 악성 사용자 활동 탐지 방법.

청구항 7

제6항에 있어서,

상기 NLP 전처리부가 자율 시스템의 경로를 통해 이상 행동을 탐지하기 위해 상기 수집된 BGP 데이터에 대해 NLP 전처리 기법을 이용하여 문자형 데이터를 수치형 데이터로 변경하는 단계는,

토크나이저(Tokenizer)를 통해 문자형 데이터를 수치형 데이터로 변경하기 위해 문자형 데이터의 문자열을 최소 단위로 나누어 시퀀스 분리를 수행하고, 시퀀스 분리를 통해 생성된 분리된 경로(Path) 문자들에 고유한 정수 번호를 부여함으로써 데이터 전처리를 수행하는

악성 사용자 활동 탐지 방법.

청구항 8

제6항에 있어서,

상기 악성 사용자 활동 탐지부가 상기 NLP 전처리 기법을 통해 변경된 수치형 데이터를 AE 모델에 입력하여 학습함으로써, 자율 시스템의 경로 기록을 미리 저장된 보안사고 사례와 비교 분석하여 정상 데이터와 이상 데이터로 분류하는 단계는,

정상 데이터와 이상 데이터에 대한 탐지 정확도를 개선하기 위해 이상 데이터의 수를 증가시켜 정상 데이터와의 비율을 맞추도록 SMOTE(Synthetic Minority Over-sampling Technique)를 이용하는

악성 사용자 활동 탐지 방법.

청구항 9

제8항에 있어서,

상기 악성 사용자 활동 탐지부가 상기 NLP 전처리 기법을 통해 변경된 수치형 데이터를 AE 모델에 입력하여 학습함으로써, 자율 시스템의 경로 기록을 미리 저장된 보안사고 사례와 비교 분석하여 정상 데이터와 이상 데이터로 분류하는 단계는,

상기 SMOTE는 이상 데이터의 샘플을 합성적으로 생성하여 정상 데이터와 이상 데이터의 비율을 맞추기 위해, 임의의 이상 데이터 하나를 중심으로 하여 가장 가까운 이상 데이터를 찾은 후, 상기 중심이 된 이상 데이터와 가장 가까운 이상 데이터 사이에 새로운 데이터를 생성하고, 생성된 데이터를 이상 데이터 클래스에 추가하고, 상기 이상 데이터 클래스에 추가하는 과정을 미리 설정된 조건을 맞출 때까지 반복하는

악성 사용자 활동 탐지 방법.

청구항 10

제8항에 있어서,

상기 악성 사용자 활동 탐지부가 상기 NLP 전처리 기법을 통해 변경된 수치형 데이터를 AE 모델에 입력하여 학습함으로써, 자율 시스템의 경로 기록을 미리 저장된 보안사고 사례와 비교 분석하여 정상 데이터와 이상 데이터로 분류하는 단계는,

AE 모델을 학습하기 전에 데이터의 특성, 편향성, 이상치를 분석하기 위해 데이터의 분포를 시각화하고, 데이터의 군집 구조와 분포를 시각화하여 비 선형적인 데이터 패턴을 포착하기 위해, 비선형 차원 축소 기법인 t-SNE(t-distributed Stochastic Neighbor Embedded)를 이용하여 이웃 그래프 기반 차원 축소를 수행함으로써, 데이터 간의 유사도에 따라 고차원 데이터를 저차원으로 변환하는

악성 사용자 활동 탐지 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 머신러닝을 활용한 악성 사용자 활동 탐지 방법 및 시스템에 관한 것이다.

배 경 기 술

[0002] 최근 몇 십 년 동안 인터넷과 디지털 기술의 발전으로 인해 사람들은 사이버 공간을 통해 다양한 활동을 하고 있으며, 이러한 변화로 인해 사람들의 상호작용 역시 대부분 사이버 공간으로 통해 이루어지고 있다. 그러나, 과거1년동안 사이버 공격은 놀라운 속도로 증가하였다. 2021년 3분기부터 2022년 3분기까지의 기간 동안 사이버 공격은 무려 18%나 증가했으며, 국가 간 사이버 공격은 무려 196%나 상승하였다. 상황에 심각성을 인지한 미국 국방부는 사이버 공간을 제5의 전장으로 지정하고 사이버 공격을 대비하고 탐지하기 위해 천문학적인 비용을 투자하고 있다. 그러나, 사이버 공격을 모두 방어하는 것은 현실적으로 불가능한 일이고 매일 새로운 공격 기법들이 탄생하기 때문에, 일부 기관의 정보 보호 시스템들은 60%에서 70%의 공격 탐지율을 보였고, 약 30%의 시스템이 잘못된 탐지 능력을 보여주었다.

선행기술문헌

비특허문헌

[0003] (비특허문헌 0001) [1] Landwehr, N.; Hall, M.; Frank, E. Logistic model trees. Mach. Learn. 2005, 59, 161-205. Breiman, L. Random forests. Mach. Learn. 2001, 45, 5-32.

(비특허문헌 0002) [2] Breiman, L. Random forests. Mach. Learn. 2001, 45, 5-32.

(비특허문헌 0003) [3] Peterson, L.E. K-nearest neighbor. Scholarpedia 2009, 4, 1883.

발명의 내용

해결하려는 과제

[0004] 본 발명이 이루고자 하는 기술적 과제는 특정 사이버 공격 집단을 선정하여 이들의 IP 및 AS에 ML(Machine Learning) 알고리즘을 적용하여 분석한 뒤, 이상 징후를 탐지하기 위한 악성 사용자 활동 탐지 방법 및 시스템을 제공하는데 있다. 본 발명의 실시예에 따르면, 전세계 네트워크의 경로 정보를 포함하고 있는 BGP 데이터를 수집하여 모델이 원활하게 학습할 수 있도록 데이터를 가공하고, 가공이 완료된 데이터를 AE(Auto-Encoder) 모델에 입력하여 이상징후를 탐지한다.

과제의 해결 수단

[0005] 일 측면에 있어서, 본 발명에서 제안하는 악성 사용자 활동 탐지 시스템은 클라이언트가 DB API 서버에게 데이터를 요청하고 DB API 서버는 자율 시스템(Autonomous System; AS)의 정보가 담긴 BGP(Border Gateway

Protocol) 데이터를 클라이언트에게 전송함으로써 데이터를 수집하는 데이터 수집부, 자율 시스템의 경로를 통해 이상 행동을 탐지하기 위해 상기 수집된 BGP 데이터에 대해 NLP(Natural Language Processing) 전처리 기법을 이용하여 문자형 데이터를 수치형 데이터로 변경하는 NLP 전처리부, 상기 NLP 전처리 기법을 통해 변경된 수치형 데이터를 AE(Auto-Encoder) 모델에 입력하여 학습함으로써, 자율 시스템의 경로 기록(Trace record)을 미리 저장된 보안사고 사례와 비교 분석하여 정상 데이터와 이상 데이터로 분류하는 악성 사용자 활동 탐지부를 포함한다.

[0006] 상기 NLP 전처리부는 토크나이저(Tokenizer)를 통해 문자형 데이터를 수치형 데이터로 변경하기 위해 문자형 데이터의 문자열을 최소 단위로 나누어 시퀀스 분리를 수행하고, 시퀀스 분리를 통해 생성된 분리된 경로(Path) 문자들에 고유한 정수 번호를 부여함으로써 데이터 전처리를 수행한다.

[0007] 상기 악성 사용자 활동 탐지부는 정상 데이터와 이상 데이터에 대한 탐지 정확도를 개선하기 위해 이상 데이터의 수를 증가시켜 정상 데이터와의 비율을 맞추도록 SMOTE(Synthetic Minority Over-sampling Technique)를 이용한다.

[0008] 상기 악성 사용자 활동 탐지부는 상기 SMOTE는 이상 데이터의 샘플을 합성적으로 생성하여 정상 데이터와 이상 데이터의 비율을 맞추기 위해, 임의의 이상 데이터 하나를 중심으로 하여 가장 가까운 이상 데이터를 찾은 후, 상기 중심이 된 이상 데이터와 가장 가까운 이상 데이터 사이에 새로운 데이터를 생성하고, 생성된 데이터를 이상 데이터 클래스에 추가하고, 상기 이상 데이터 클래스에 추가하는 과정을 미리 설정된 조건을 맞출 때까지 반복한다.

[0009] 상기 악성 사용자 활동 탐지부는 AE 모델을 학습하기 전에 데이터의 특성, 편향성, 이상치를 분석하기 위해 데이터의 분포를 시각화하고, 데이터의 군집 구조와 분포를 시각화하여 비 선형적인 데이터 패턴을 포착하기 위해, 비선형 차원 축소 기법인 t-SNE(t-distributed Stochastic Neighbor Embedded)를 이용하여 이웃 그래프 기반 차원 축소를 수행함으로써, 데이터 간의 유사도에 따라 고차원 데이터를 저차원으로 변환한다.

[0010] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 악성 사용자 활동 탐지 방법은 데이터 수집부를 통해 클라이언트가 DB API 서버에게 데이터를 요청하고 DB API 서버는 자율 시스템의 정보가 담긴 BGP 데이터를 클라이언트에게 전송함으로써 데이터를 수집하는 단계, NLP 전처리부가 자율 시스템의 경로를 통해 이상 행동을 탐지하기 위해 상기 수집된 BGP 데이터에 대해 NLP 전처리 기법을 이용하여 문자형 데이터를 수치형 데이터로 변경하는 단계, 악성 사용자 활동 탐지부가 상기 NLP 전처리 기법을 통해 변경된 수치형 데이터를 AE 모델에 입력하여 학습함으로써, 자율 시스템의 경로 기록을 미리 저장된 보안사고 사례와 비교 분석하여 정상 데이터와 이상 데이터로 분류하는 단계를 포함한다.

발명의 효과

[0011] 본 발명의 실시예들에 따르면 머신러닝을 사용하여 BGP 데이터를 기반으로 사이버 공간의 이상 행동을 탐지하고, 데이터를 정상 혹은 이상 데이터로 분류하기 위해 KNN, NB, RF, AE 등 다양한 머신러닝 알고리즘을 활용할 수 있다. 또한, 알고리즘에 원활한 학습을 진행하기 위해 Tokenizer를 사용하여 문자형 데이터들을 수치형 데이터들로 변환하고, 모델의 분류 성능을 향상시키기 위해 SMOTE 샘플링 기법을 활용한다. 본 발명의 실시예들에 따르면 머신러닝을 사용하여 BGP 데이터를 기반으로 이상 행동을 탐지할 수 있는 프로세스를 제공한다. 본 발명의 실시예들에 따르면 실시간 사이버 이상 행동을 탐지하고 기존 악의적인 사용자들이 경로 변경을 탐지하여 지속적인 감시를 수행하는데 사용될 수 있다. 또한, 본 발명의 실시예에 따른 사용된 전처리 방법은 기존 BGP 데이터에서는 사용되지 않았던 NLP기반 Tokenizer 기법을 적용하고, SMOTE를 사용하여 모델의 성능을 향상시킬 수 있다.

도면의 간단한 설명

[0012] 도 1은 본 발명의 일 실시예에 따른 RF 모델을 설명하기 위한 도면이다.

도 2는 본 발명의 일 실시예에 따른 AE 구조를 설명하기 위한 도면이다.

도 3은 본 발명의 일 실시예에 따른 악성 사용자 활동 탐지 시스템의 구성을 나타내는 도면이다.

도 4는 본 발명의 일 실시예에 따른 악성 사용자 활동 탐지 방법을 설명하기 위한 흐름도이다.

도 5는 본 발명의 일 실시예에 따른 데이터 수집 과정을 설명하기 위한 도면이다.

도 6은 본 발명의 일 실시예에 따른 정상 데이터와 이상 데이터 수집 과정을 설명하기 위한 도면이다.

도 7은 본 발명의 일 실시예에 따른 이상 데이터 탐지 모델을 설명하기 위한 도면이다.

도 8은 본 발명의 일 실시예에 따른 BGP 데이터 전처리 과정을 설명하기 위한 도면이다.

도 9는 본 발명의 일 실시예에 따른 전처리가 완료된 데이터에 대하여 파이 차트를 통해 데이터 분포를 나타낸 도면이다.

도 10은 본 발명의 일 실시예에 따른 원본 데이터셋과 SMOTE의 분포를 t-SNE를 통해 나타낸 도면이다.

도 11은 본 발명의 일 실시예에 따른 전처리 전후의 데이터셋에 대한 ROC 커브와 AUROC 값을 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0013] 최근 인터넷과 디지털 기술의 발전으로 다양한 활동이 사이버 공간에서 이루어졌으나, 그 만큼 사이버 공격 또한 급증하여, 그 어느 때보다 사이버 공격을 탐지하고 예방하는 것이 중요해졌다. 따라서, 본 발명에서는 머신러닝을 이용하여 BGP(Border Gateway Protocol) 데이터를 통해 사이버 공간에서의 이상 행동을 탐지하고 분류하는 방법을 제안한다. 본 발명의 난제 중 하나는 자율 시스템(Autonomous System; AS)의 경로를 통해 이상 행동을 탐지해야 하는데 머신러닝이 학습할 수 없는 문자형 데이터라는 것이다. 이 문제를 해결하기 위해 NLP(Natural Language Processing) 전처리 기법인 Tokenizer를 통해 문자형 데이터를 수치형 데이터로 변경하였다. 이는 단순히 문자를 숫자를 바꾼 것이 아니라, 데이터의 패턴과 특성을 그대로 담아내었다. 그런 다음 불균형 데이터 문제를 해결하기 위해 SMOTE을 사용했다. 예를 들어, K-Nearest Neighbor, Naive Bayes, Random Forest, Autoencoder 등 다양한 머신러닝 알고리즘을 사용하여 모델을 학습시켰고, 성능 또한 크게 개선하였다. 본 발명의 실시예에 따른 실험 결과 AE 모델과 사용했을 때, F1-Score 0.99로 가장 우수한 성능을 보여주었다. 또한, AUROC(Area Under the Receiver Operating Characteristic Curve)를 통해 모든 모델이 90%이상의 우수한 성능을 보였다. 본 발명의 실시예에 따르면, BGP 데이터를 통해 악의적인 사용자들의 사이버 이상 행동을 실시간으로 탐지 및 감시할 수 있으며, 사이버 보안 향상에 이바지할 수 있을 것으로 기대된다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0015] BGP는 인터넷의 주요 라우팅 프로토콜 중 하나로, 여러 개의 자율 시스템(Autonomous System; AS) 사이에서 라우팅 정보를 교환하고 인터넷 상의 통신 경로를 결정하는데 사용된다. AS는 관리자에 의해 자율적으로 운영되는 네트워크 집합을 의미하며, 각 AS는 고유한 번호인 AS 번호를 가지고 있다.

[0016] BGP는 네트워크의 경로와 접근 가능성 정보를 교환하기 위해 주로 사용된다. 이러한 정보는 다양한 AS들 사이를 거쳐 목적지까지 가는 최적의 경로를 결정하는데 사용된다. BGP는 트랜지션-자홉 방식으로 동작하며, 라우팅 테이블을 업데이트하고 광범위한 네트워크 변화에 대응하는 데 사용되는 다양한 속성과 메커니즘을 제공한다.

[0017] 이러한 BGP 데이터는 대규모의 라우팅 정보를 포함하고 있으며, 전 세계의 다양한 AS들 간에 라우팅 정보가 교환되기 때문에 네트워크 동작에 매우 중요한 역할을 한다. 그러나 이러한 BGP 데이터의 복잡성과 크기 때문에 이를 효과적으로 분석하고 이상 상황을 탐지하는 것에 어려움이 있다.

[0018] BGP 데이터로 이상을 탐지하기 위해서는 데이터 기반으로 비정상 AS 경로를 탐지해야 한다. 따라서, 본 발명에서는 종래기술들의 한계점을 극복하기 위해 데이터 수집, 데이터 전처리, 알고리즘 성능 검증 등을 개선하여 보다 탄력적인 사이버 이상 탐지 프로세스를 제안한다.

[0019] 본 발명의 실시예에 따르면, ML 모델들로 이상징후를 탐지하기 전에 먼저 데이터의 고유한 특성을 파악하고 데이터를 증강하였다. 그 후, RF(Random Forest), LMT(Logistic model trees), KNN(K-Nearest Neighbors) 등의 머신러닝 알고리즘을 사용하여 악성 활동을 탐지한다.

[0020] 본 발명의 실시예에 따르면, 실시간 데이터인 BGP 데이터를 가지고 정상과 이상 활동을 실시간으로 이상활동을 탐지하고자 한다. 또한 기존 BGP를 활용한 종래기술에서 제시하지 못한 ML 모델들의 다양성, 평가지표 제시를 통해 종래기술과 다른 방식을 적용하고자 한다.

[0022] 도 1은 본 발명의 일 실시예에 따른 RF 모델을 설명하기 위한 도면이다.

[0023] RF(Random Forest)는 여러 개의 결정 트리를 결합하여 학습하는 앙상블 방법이다. 도 1에서 알 수 있듯이, RF는 여러 개의 결정 트리를 생성한 다음, 랜덤한 데이터 샘플들을 각 트리에 독립적으로 학습시킨다. 이렇게 함으로써 각 트리는 서로 다른 데이터 특성들을 학습하게 되고, 이들의 결과를 조합하여 과적합을 방지하고 보다 정확하고 안정적인 예측을 할 수 있도록 한다.

[0024] RF의 앙상블 방법은 동일한 알고리즘을 사용하더라도 랜덤한 데이터를 사용하기 때문에 각 트리마다 서로 다른 특성을 학습한다. 대표적인 방법으로는 Bagging과 Boosting이 있다.

[0025] Bagging은 데이터셋에서 무작위로 복원 추출을 통해 여러 개의 샘플 데이터 집합을 생성하고 각각의 데이터를 독립적으로 결정 트리에 학습시킨다. 학습된 모든 결정 트리의 예측 결과를 평균 값을 구하거나 투표를 통해 최종 예측을 수행한다. 위 기법은 다양한 샘플 데이터와 특성을 다루어 예측의 정확성과 안정성을 향상시키는데 도움이 된다.

[0026] Boosting은 초기에 모든 데이터에 대한 동일 가중치를 부여하고 학습을 진행한다. 예측 결과를 평가하여 잘못 분류된 데이터에 더 높은 가중치를 부여하고 다시 학습을 진행한다. 위 과정을 통해 결정 트리는 편향된 학습을 진행하기 때문에 오류를 보완하고 정확성을 높일 수 있다.

[0027] 앞서 설명된 바와 같이, RF의 가장 큰 특성은 랜덤한 데이터 샘플들을 만들어서 각 트리에 다른 특성을 가지도록 학습을 유도한다. 이러한 특성은 일반화 성능을 향상시켜 과적합을 방지한다. 또한 데이터를 랜덤하게 샘플링 하는 것은 편향된 데이터를 결정 트리에 입력하여도 모델을 더욱 강화한다. 이는 대량의 데이터를 효율적으로 처리하며, 모델의 편향을 회피하여 과적합 문제를 방지하고 더욱 정확한 예측 능력을 향상시켜 변동성을 줄일 수 있다.

[0028] KNN(*K-Nearest Neighbors*)은 데이터셋에서 이상 데이터를 탐지하기 위해 "정상" 데이터들의 집합으로부터 얼마나 멀리 떨어져 있는지를 측정하여 이상 데이터를 분류한다. KNN의 이상 탐지는 데이터간의 거리를 기반으로 수행된다. 예를 들어, 주어진 데이터 포인트와 가장 가까운 K개의 이웃 데이터 포인트를 먼저 찾은 다음, 해당 데이터 포인트와 이웃(다시 말해, 데이터들)들 사이의 거리를 측정한다. 만약 한 데이터 포인트를 선택하고 K=5로 설정하면, 이 데이터 포인트와 가장 가까운 5개의 이웃 데이터 포인트를 찾아서 그들과의 거리를 측정한다. 그리고 이 거리가 멀면, 해당 데이터 포인트는 이상 값으로 간주된다. 이렇게 여러 데이터 포인트과의 거리를 비교하면서 이상값을 찾는 것이 KNN 이상 탐지 방법이다. 작은 K값은 이상 값은 민감하게 탐지하고, 큰 K값은 이상 값을 보수적으로 탐지한다. KNN의 수식은 식(1)과 같다.

[0029] $$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n}(x_{i,k}, x_{j,k})^2} \quad (1)$$

[0030] NB(Naive Bayes)는 기계 학습의 분류 알고리즘 중 하나로, 확률 기반의 간단하고 효과적인 모델이다. 이 모델은 주어진 데이터의 특성을 기반으로 데이터를 분류하는데 'naive'라는 이름처럼 특성들이 서로 독립적이라고 가정한다. 이는 실제 데이터에서는 항상 성립하지 않을 수 있지만, NB는 이런 간단한 가정을 통해 계산이 편리해지고 속도가 빨라진다. 학습할 때, 주어진 학습 데이터를 사용하여 각 클래스에 대한 특성들의 확률 분포를 추정한다. 데이터들의 평균 값과 분산 정도를 파악하여, 각 클래스에 속할 확률을 계산하여 높은 확률의 클래스에 데이터를 분류하는 것이 NB의 특징이다. NB는 계산이 간단하고 데이터의 차원이 높아도 빠른 속도로 연산이 가능하지만 가정이 간단하다는 한계로 인해 특성들 사이의 상호작용을 고려하지 못하여 정확도가 떨어질 수 있다.

[0032] 도 2는 본 발명의 일 실시예에 따른 AE 구조를 설명하기 위한 도면이다.

[0033] AE(AutoEncoder)는 데이터의 특징을 추출하여 분류하는 알고리즘 중 하나이다. 도 2와 같이 데이터를 압축하는 인코더(Encoder)와 압축된 데이터를 복원하는 디코더(Decoder)로 구성되어 있다. 인코더는 고차원 데이터를 입력 받아 저차원으로 압축한 후, 잠재 공간에 매핑을 한다. 디코더는 저차원을 압축된 데이터들은 다시 원래 차원으로 복구하여 재구성하는 학습 방법을 가지고 있다. 여기서, 입력 데이터와 재구성 데이터를 비교하여 오류를 최소화하는 것이 AE 학습의 핵심이다. 식(2)는 입력 데이터 X를 저차원 데이터 h로 압축하고 매핑하는 것은 의미한다.

[0034] $$h = f(X) = \sigma(WX + b) \quad (2)$$

[0035] 위 수식에서 W는 가중치 행렬, b는 편향, σ는 활성화 함수를 의미한다. 일반적으로 학습에 사용하는 활성화 함수는 시그모이드 함수(sigmoid function) 또는 하이퍼볼릭탄젠트 함수(hyperbolic tangent function)를 사용한다. 압축된 데이터들을 다시 복원하여 재구성하는 수식은 식(3)과 같다.

[0036] $$r = g(h) = \sigma(W'h + b') \quad (3)$$

[0037] 식(3)에서 압축된 데이터의 W'는 가중치, b'는 편향, σ는 활성화 함수, r은 복원(reconstruction)을 의미한다. 앞서 구했던 h 와 r을 통해 복원 에러(reconstruction error)를 구하는데 AE는 복원 에러를 줄이도록 학습을 진행한다. 해당 값은 구하는 것은 식(4)와 같다.

[0038] $$L(X, r) = \|X - r\|_2^2 \quad (4)$$

[0039] 위 수식에서, $\|\cdot\|$는 L_2 norm을 의미한다. L_2 norm이란 벡터의 길이 또는 크기를 측정하는 방법 중 하나로, 각 원소의 제곱을 합한 후 그 값의 제곱근을 취하는 방법이다. 만약 $L(X,r)$의 값이 크다면 제대로 복원되지 않아 이상 데이터로 분류하고, $L(X,r)$이 0이거나 매우 작다면 정상 데이터로 분류하도록 학습이 진행된다.

[0041] 도 3은 본 발명의 일 실시예에 따른 악성 사용자 활동 탐지 시스템의 구성을 나타내는 도면이다.

[0042] 본 실시예에 따른 악성 사용자 활동 탐지 시스템(300)은 프로세서(310), 버스(320), 네트워크 인터페이스(330), 메모리(340) 및 데이터베이스(350)를 포함할 수 있다. 메모리(340)는 운영체제(341) 및 악성 사용자 활동 탐지 루틴(342)을 포함할 수 있다. 프로세서(310)는 데이터 수집부(311), NLP 전처리부(312) 및 악성 사용자 활동 탐지부(313)를 포함할 수 있다. 다른 실시예들에서 악성 사용자 활동 탐지 시스템(300)은 도 3의 구성요소들보다 더 많은 구성요소들을 포함할 수도 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다. 예를 들어, 악성 사용자 활동 탐지 시스템(300)은 디스플레이나 트랜시버(transceiver)와 같은 다른 구성요소들을 포함할 수도 있다.

[0043] 메모리(340)는 컴퓨터에서 판독 가능한 기록 매체로서, RAM(random access memory), ROM(read only memory) 및 디스크 드라이브와 같은 비소멸성 대용량 기록장치(permanent mass storage device)를 포함할 수 있다. 또한, 메모리(340)에는 운영체제(341)와 악성 사용자 활동 탐지 루틴(342)을 위한 프로그램 코드가 저장될 수 있다. 이러한 소프트웨어 구성요소들은 드라이브 메커니즘(drive mechanism, 미도시)을 이용하여 메모리(340)와는 별도의 컴퓨터에서 판독 가능한 기록 매체로부터 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독 가능한 기록 매체는 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록 매체(미도시)를 포함할 수 있다. 다른 실시예에서 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록 매체가 아닌 네트워크 인터페이스(330)를 통해 메모리(340)에 로딩될 수도 있다.

[0044] 버스(320)는 악성 사용자 활동 탐지 시스템(300)의 구성요소들간의 통신 및 데이터 전송을 가능하게 할 수 있다. 버스(320)는 고속 시리얼 버스(high-speed serial bus), 병렬 버스(parallel bus), SAN(Storage Area Network) 및/또는 다른 적절한 통신 기술을 이용하여 구성될 수 있다.

[0045] 네트워크 인터페이스(330)는 악성 사용자 활동 탐지 시스템(300)을 컴퓨터 네트워크에 연결하기 위한 컴퓨터 하드웨어 구성요소일 수 있다. 네트워크 인터페이스(330)는 악성 사용자 활동 탐지 시스템(300)을 무선 또는 유선 커넥션을 통해 컴퓨터 네트워크에 연결시킬 수 있다.

[0046] 데이터베이스(350)는 악성 사용자 활동 탐지를 위해 필요한 모든 정보를 저장 및 유지하는 역할을 할 수 있다. 도 3에서는 악성 사용자 활동 탐지 시스템(300)의 내부에 데이터베이스(350)를 구축하여 포함하는 것으로 도시하고 있으나, 이에 한정되는 것은 아니며 시스템 구현 방식이나 환경 등에 따라 생략될 수 있고 혹은 전체 또는 일부의 데이터베이스가 별개의 다른 시스템 상에 구축된 외부 데이터베이스로서 존재하는 것 또한 가능하다.

[0047] 프로세서(310)는 기본적인 산술, 로직 및 악성 사용자 활동 탐지 시스템(300)의 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(340) 또는 네트워크 인터페이스(330)에 의해, 그리고 버스(320)를 통해 프로세서(310)로 제공될 수 있다. 프로세서(310)는 데이터 수집부(311), NLP 전처리부(312) 및 악성 사용자 활동 탐지부(313)를 위한 프로그램 코드를 실행하도록 구성될 수 있다. 이러한

프로그램 코드는 메모리(340)와 같은 기록 장치에 저장될 수 있다.

[0048] 데이터 수집부(311), NLP 전처리부(312) 및 악성 사용자 활동 탐지부(313)는 도 4의 단계들(410~430)을 수행하기 위해 구성될 수 있다.

[0049] 악성 사용자 활동 탐지 시스템(300)은 데이터 수집부(311), NLP 전처리부(312) 및 악성 사용자 활동 탐지부(313)를 포함할 수 있다.

[0050] 본 발명의 실시예에 따른 데이터 수집부(311)는 클라이언트가 DB API 서버에게 데이터를 요청하고 DB API 서버는 자율 시스템의 정보가 담긴 BGP 데이터를 클라이언트에게 전송함으로써 데이터를 수집한다.

[0051] 본 발명의 실시예에 따른 NLP 전처리부(312)는 자율 시스템의 경로를 통해 이상 행동을 탐지하기 위해 상기 수집된 BGP 데이터에 대해 NLP 전처리 기법을 이용하여 문자형 데이터를 수치형 데이터로 변경한다.

[0052] 본 발명의 실시예에 따른 NLP 전처리부(312)는 토크나이저(Tokenizer)를 통해 문자형 데이터를 수치형 데이터로 변경하기 위해 문자형 데이터의 문자열을 최소 단위로 나누어 시퀀스 분리를 수행하고, 시퀀스 분리를 통해 생성된 분리된 경로(Path) 문자들에 고유한 정수 번호를 부여함으로써 데이터 전처리를 수행한다.

[0053] 본 발명의 실시예에 따른 악성 사용자 활동 탐지부(313)는 상기 NLP 전처리 기법을 통해 변경된 수치형 데이터를 AE 모델에 입력하여 학습함으로써, 자율 시스템의 경로 기록을 미리 저장된 보안사고 사례와 비교 분석하여 정상 데이터와 이상 데이터로 분류한다.

[0054] 본 발명의 실시예에 따른 악성 사용자 활동 탐지부(313)는 정상 데이터와 이상 데이터에 대한 탐지 정확도를 개선하기 위해 이상 데이터의 수를 증가시켜 정상 데이터와의 비율을 맞추도록 SMOTE(Synthetic Minority Over-sampling Technique)를 이용한다.

[0055] 본 발명의 실시예에 따른 악성 사용자 활동 탐지부(313)는 상기 SMOTE는 이상 데이터의 샘플을 합성적으로 생성하여 정상 데이터와 이상 데이터의 비율을 맞추기 위해, 임의의 이상 데이터 하나를 중심으로 하여 가장 가까운 이상 데이터를 찾은 후, 상기 중심이 된 이상 데이터와 가장 가까운 이상 데이터 사이에 새로운 데이터를 생성하고, 생성된 데이터를 이상 데이터 클래스에 추가하고, 상기 이상 데이터 클래스에 추가하는 과정을 미리 설정된 조건을 맞출 때까지 반복한다.

[0056] 본 발명의 실시예에 따른 악성 사용자 활동 탐지부(313)는 AE 모델을 학습하기 전에 데이터의 특성, 편향성, 이상치를 분석하기 위해 데이터의 분포를 시각화하고, 데이터의 군집 구조와 분포를 시각화하여 비 선형적인 데이터 패턴을 포착하기 위해, 비선형 차원 축소 기법인 t-SNE(t-distributed Stochastic Neighbor Embedded)를 이용하여 이웃 그래프 기반 차원 축소를 수행함으로써, 데이터 간의 유사도에 따라 고차원 데이터를 저차원으로 변환한다.

[0058] 도 4는 본 발명의 일 실시예에 따른 악성 사용자 활동 탐지 방법을 설명하기 위한 흐름도이다.

[0059] 제안하는 악성 사용자 활동 탐지 방법은 데이터 수집부를 통해 클라이언트가 DB API 서버에게 데이터를 요청하고 DB API 서버는 자율 시스템의 정보가 담긴 BGP 데이터를 클라이언트에게 전송함으로써 데이터를 수집하는 단계(410), NLP 전처리부가 자율 시스템의 경로를 통해 이상 행동을 탐지하기 위해 상기 수집된 BGP 데이터에 대해 NLP 전처리 기법을 이용하여 문자형 데이터를 수치형 데이터로 변경하는 단계(420), 악성 사용자 활동 탐지부가 상기 NLP 전처리 기법을 통해 변경된 수치형 데이터를 AE 모델에 입력하여 학습함으로써, 자율 시스템의 경로 기록을 미리 저장된 보안사고 사례와 비교 분석하여 정상 데이터와 이상 데이터로 분류하는 단계(430)를 포함한다.

[0060] 단계(410)에서, 데이터 수집부를 통해 클라이언트가 DB API 서버에게 데이터를 요청하고 DB API 서버는 자율 시스템의 정보가 담긴 BGP 데이터를 클라이언트에게 전송함으로써 데이터를 수집한다.

[0061] 단계(420)에서, NLP 전처리부가 자율 시스템의 경로를 통해 이상 행동을 탐지하기 위해 상기 수집된 BGP 데이터에 대해 NLP 전처리 기법을 이용하여 문자형 데이터를 수치형 데이터로 변경한다.

[0062] 본 발명의 실시예에 따르면, 토크나이저(Tokenizer)를 통해 문자형 데이터를 수치형 데이터로 변경하기 위해 문자형 데이터의 문자열을 최소 단위로 나누어 시퀀스 분리를 수행하고, 시퀀스 분리를 통해 생성된 분리된 경로(Path) 문자들에 고유한 정수 번호를 부여함으로써 데이터 전처리를 수행한다.

[0063] 단계(430)에서, 악성 사용자 활동 탐지부가 상기 NLP 전처리 기법을 통해 변경된 수치형 데이터를 AE 모델에 입력하여 학습함으로써, 자율 시스템의 경로 기록을 미리 저장된 보안사고 사례와 비교 분석하여 정상 데이터와 이상 데이터로 분류한다.

[0064] 본 발명의 실시예에 따르면, 정상 데이터와 이상 데이터에 대한 탐지 정확도를 개선하기 위해 이상 데이터의 수를 증가시켜 정상 데이터와의 비율을 맞추도록 SMOTE(Synthetic Minority Over-sampling Technique)를 이용한다.

[0065] 본 발명의 실시예에 따르면, 상기 SMOTE는 이상 데이터의 샘플을 합성적으로 생성하여 정상 데이터와 이상 데이터의 비율을 맞추기 위해, 임의의 이상 데이터 하나를 중심으로 하여 가장 가까운 이상 데이터를 찾은 후, 상기 중심이 된 이상 데이터와 가장 가까운 이상 데이터 사이에 새로운 데이터를 생성하고, 생성된 데이터를 이상 데이터 클래스에 추가하고, 상기 이상 데이터 클래스에 추가하는 과정을 미리 설정된 조건을 맞출 때까지 반복한다.

[0066] 본 발명의 실시예에 따르면, AE 모델을 학습하기 전에 데이터의 특성, 편향성, 이상치를 분석하기 위해 데이터의 분포를 시각화하고, 데이터의 군집 구조와 분포를 시각화하여 비 선형적인 데이터 패턴을 포착하기 위해, 비선형 차원 축소 기법인 t-SNE(t-distributed Stochastic Neighbor Embedded)를 이용하여 이웃 그래프 기반 차원 축소를 수행함으로써, 데이터 간의 유사도에 따라 고차원 데이터를 저차원으로 변환한다.

[0068] 도 5는 본 발명의 일 실시예에 따른 데이터 수집 과정을 설명하기 위한 도면이다.

[0069] 본 발명의 일 실시예에 따른 사이버 공간의 이상 징후를 탐지하기 위해 BGP 데이터를 수집하는 과정을 도 5에 도시하였다.

[0070] 먼저, 클라이언트(510)가 DB API 서버(520)에게 데이터를 요청하고 DB API 서버(520)는 DB 저장부(530)에 저장된 AS의 정보가 담긴 BGP 데이터를 클라이언트(510)에게 전송함으로써 데이터를 수집한다.

[0072] 도 6은 본 발명의 일 실시예에 따른 정상 데이터와 이상 데이터 수집 과정을 설명하기 위한 도면이다.

[0073] 도 5에서 설명된 바와 같이 수집된 BGP 데이터의 AS 경로 기록(Trace record)(610)을 데이터베이스(620)에 미리 저장된 보안사고 사례와 비교 분석(611)하여 정상 데이터와 이상 데이터로 분리한다(621). 정상 데이터(631)는 사이버 공격 기록이 없는 AS들을 말하며, 이상 데이터(632)는 사이버 공격을 자주 시도한 AS를 의미한다. 앞서 설명된 바와 같이 수집된 데이터는 표 2에 나타나 있으며, 데이터 속성에 대한 설명은 표 3에 나타내었다.

[0074] <표 2>

Timestamp	Peer IP	Peer AS	path	Location	Label
2022-09-16	212.66.96.212	20912	A-> C -> D	Italy	1
2022-09-16	12.0.1.63	7018	U-> F -> E	US	0
2022-09-16	37.139.139.17	57866	H-> U -> A	Netherland	1
2022-09-17	194.153.0.2	5413	L -> C -> A	Australia	0
2022-09-17	198.129.33.85	293	U -> A -> D	US	0
...	...	...	...	...	...

[0075]

[0076] <표 3>

Column	Definition
Timestamp	라우팅 데이터를 송수신한 시간을 표시하는 시간 정보이다. 일반적으로 날짜와 시간 정보를 조합하여 이벤트가 발생한 정확한 시점을 표시한다
Peer IP	네트워크 환경에서 통신을 주고받는 두 개의 장치 또는 시스템 간의 IP 주소를 가리킨다. 각 장치는 고유한 IP 주소를 갖기 때문에 이를 통해 소유자를 식별할 수 있다.
Peer AS	네트워크 환경에서 통신을 주고받을 때, 통신을 송신한 AS 의 번호를 가리킨다. AS 번호도 IP와 마차가지로 고유한 식별 번호를 가져 소유자를 알 수 있다.
Path	데이터 패킷이 어떤 AS 를 걸쳐서 목적지까지 도착했는지에 대한 정보를 알 수 있다.
Location	Peer AS 의 소유국의 정보를 알 수 있다.
Label	정상과 이상 데이터를 구분할 수 있다. 0: Normal data 1: abnormal data

[0077]

[0079] 도 7은 본 발명의 일 실시예에 따른 이상 데이터 탐지 모델을 설명하기 위한 도면이다.

[0080] 본 발명의 실시예에 따른 이상 데이터 선정 기준은 자주 공격을 시도한 Peer AS의 경로가 불규칙적으로 변화한 경우를 이상 데이터를 선정하였다. 예를 들어, 경로의 경우 평소에는 'A - > B -> C' 순으로 경로를 이용하던 의심 대상이 갑자기 'A -> D -> E -> F'로 변경되면 이상 데이터로 분류한다.

[0081] 정상 데이터 및 이상 데이터들에 대하여 AE 모델 구조를 기반으로 사이버 공간의 이상징후 탐지 실험을 진행하였다. 정상 데이터 및 이상 데이터는 AE 모델에 입력되어 압축 및 복원 단계를 걸치게 된다. 이 때, 복원이 되면 정상 데이터로 분류하고 복원되지 않으면 이상 데이터로 분류하게 된다.

[0083] 도 8은 본 발명의 일 실시예에 따른 BGP 데이터 전처리 과정을 설명하기 위한 도면이다.

[0084] 모델들이 학습하기 위해서는 수치형 데이터만 존재한다. 그러나, 표 2에 나타낸 바와 같이, 경로(Path)와 위치(Location) 같은 경우는 문자형 데이터이기 때문에 모델들이 학습을 진행할 수 없다. 이러한 문제점을 해결하기 위해 전처리 과정을 진행한다. 도 8을 참조하면, 본 발명의 실시예에 따른 Tokenizer를 통해 문자형 데이터를 수치형 데이터로 변경하는 과정을 도시하였다.

[0085] 먼저, 문자열을 최소 단위로 나누어 시퀀스 분리를 수행한다. 예를 들어 'A -> B ->F->H'라는 경로가 있다면 이를 ['A', 'B', 'F', 'H']와 같이 분리된다. 마지막으로 분리된 경로 문자들에게 고유한 정수 번호를 부여해 줌으로써 데이터 전처리가 완료된다.

[0087] 도 9는 본 발명의 일 실시예에 따른 전처리가 완료된 데이터에 대하여 파이 차트를 통해 데이터 분포를 나타낸 도면이다.

[0088] 도 9(a)는 원본 데이터셋을 나타내는 도면이고, 도 9(b)는 전처리가 완료된 데이터를 본 발명의 실시예에 따른 파이 차트(pie chart)를 통해 데이터 분포를 나타낸 도면이다. 도 9를 참조하면, 심각한 데이터 분균형이 존재하며, 이상 데이터의 수가 너무 부족하다는 것을 알 수 있다. 이는 이상 데이터의 패턴을 제대로 학습하지 되지 않아 정상 데이터와 이상 데이터를 제대로 탐지하지 못하는 결과를 초래한다. 따라서, 본 발명에서는 이상 데이터의 수를 증가시켜 정상 데이터와의 비율을 맞추기 위해 SMOTE(Synthetic Minority Over-sampling Technique)를 사용한다.

[0089] 본 발명의 실시예에 따른 SMOTE는 모델들이 이상 데이터의 탐지능력을 향상시키는 기법이다. 이상 데이터의 샘플을 합성적으로 생성하여 정상 데이터와 이상 데이터의 균형을 맞추는 방식으로 동작한다. 이 알고리즘은 이상 데이터 하나를 중심으로 잡은 다음, 가장 가까운 이상 데이터를 찾는다. 그런 다음 중심이 된 이상 데이터와 가장 가까운 이상 데이터 사이에 새로운 데이터를 생성하고, 생성된 데이터를 이상 데이터 클래스에 추가한다. 위 과정은 사용자가 설정한 데이터 필요를 맞출 때까지 계속 반복된다. SMOTE는 이상 탐지, 사진 인식 등 다양한

분야에서 널리 사용되고 있다. SMOTE의 사용 결과는 Figure 7의 오른쪽 그림을 통해 알 수 있다.

[0091] 도 10은 본 발명의 일 실시예에 따른 원본 데이터셋과 SMOTE의 분포를 t-SNE를 통해 나타낸 도면이다.

[0092] 도 10(a)는 원본 데이터셋을 나타내는 도면이고, 도 10(b)는 SMOTE의 분포를 t-SNE를 통해 나타낸 도면이다.

[0093] 모델을 학습하기 전에 데이터의 특성, 편향성, 이상치 등을 파악하여 데이터가 어떻게 구성되어 있는지 이해하기 위해 데이터의 분포를 파악해야 한다. 데이터의 분포를 이해하는 가장 좋은 방법은 시각화를 통해 데이터를 이해하는 것이다. 시각화 하는 대표적인 방법은 차원 축소를 통해 고차원 데이터를 저차원으로 변환해주는 PCA(Principal Component Analysis)를 사용하는 것이다. 그러나, PCA는 선형 분석을 사용하기 때문에, 비 선형적인 패턴이나 복잡한 군집 형태를 잘 처리하지 못하는 한계점이 존재한다. 특히, 데이터가 비선형 경계를 가지는 군집화 문제에 적용할 경우, PCA로 차원 축소를 수행하며 데이터의 분포를 완벽하게 파악하지 못할 수 있다. 이러한 문제를 해결하기 위해 t-SNE(t-distributed Stochastic Neighbor Embedded)를 사용한다. t-SNE는 비선형 차원 축소 기법으로, 이웃 그래프 기반 차원 축소를 사용하기 때문에, 데이터 간의 유사도를 고려하여 고차원 데이터를 저차원으로 변환한다. 또한, 데이터의 군집 구조와 분포를 시각화 하여 비 선형적인 데이터 패턴을 잘 포착하고 보존한다. 도 10(b)를 참조하면, 데이터 분포가 이상 데이터는 오른쪽에 군집되어 있으며, 그 중 몇 개는 정상 데이터와 섞여 있는 모습을 볼 수 있다. 그러나, 도 10(a)를 참조하면, 원본 데이터는 이상 데이터가 거의 존재하지 않는 모습을 보인다. t-SNE을 통해 이상 데이터가 적은 원본 데이터에 대한 전처리 과정을 걸쳐 적절하게 이상 데이터가 분포되어 있음을 알 수 있다.

[0095] 도 11은 본 발명의 일 실시예에 따른 전처리 전후의 데이터셋에 대한 ROC 커브와 AUROC 값을 나타내는 도면이다.

[0096] 도 11(a)는 원본 데이터셋을 나타내는 도면이고, 도 11(b)는 본 발명의 실시예에 따라 전처리된 데이터셋에 대한 ROC 커브와 AUROC 값을 나타내는 도면이다.

[0097] 본 발명의 실시예에 따른 평가 모델에는 KNN, NB, RF, AE를 사용하였으며, 성능을 측정하는 지표로 Precision, Recall, F1-Score를 이용한다. 그 결과, 전처리된 데이터셋(SMOTE)이 원본 데이터셋보다 압도적으로 높은 성능을 보여주는 것을 알 수 있다. 특히 AE 경우 F1-score 0.89에서 0.95으로 크게 증가했고, 대체적으로 모든 지표가 개선되었음을 알 수 있다.

[0099] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0100] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

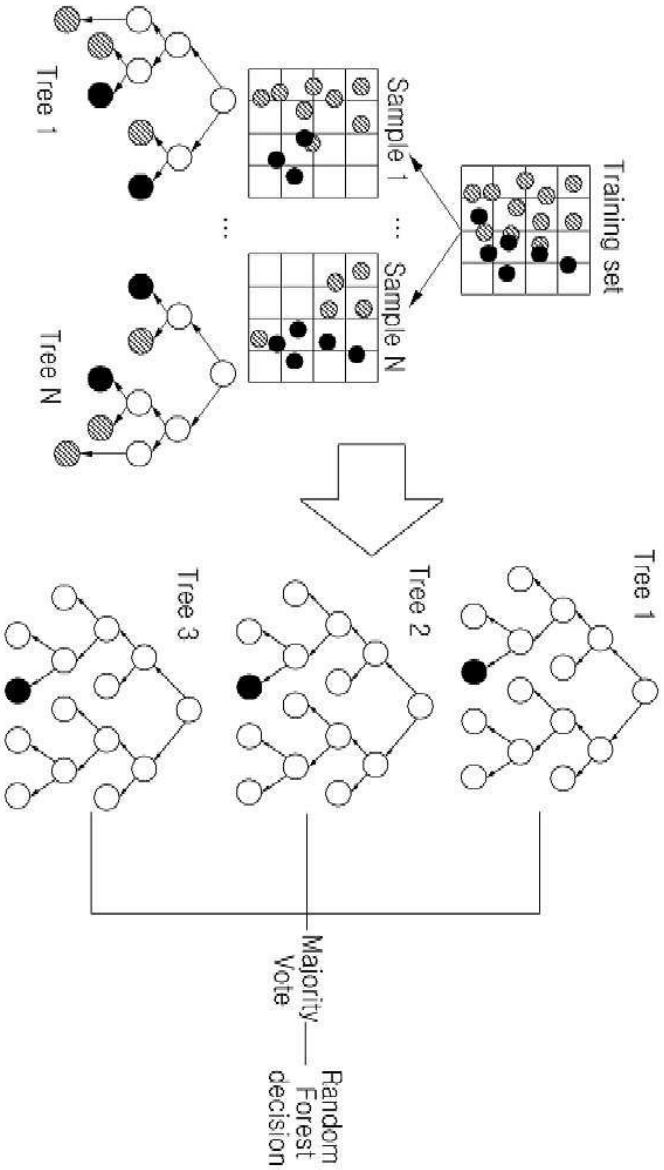
[0101] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0102] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

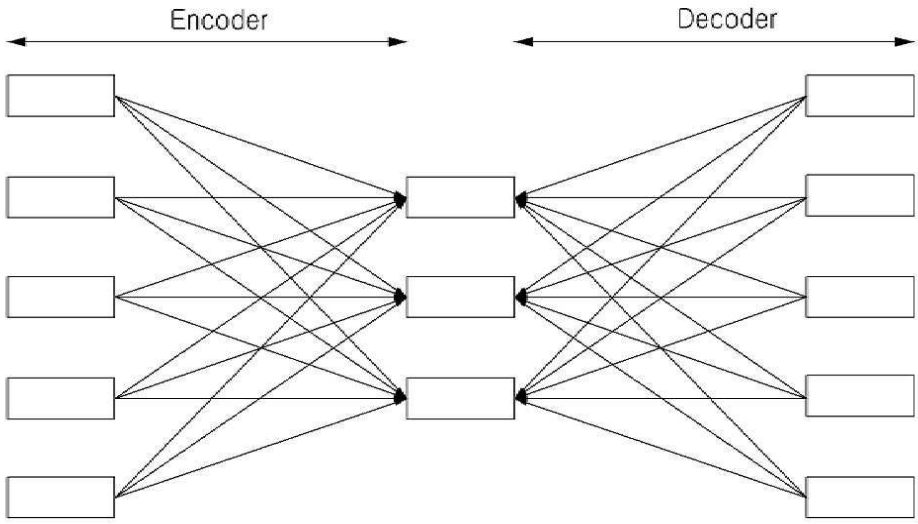
[0103] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

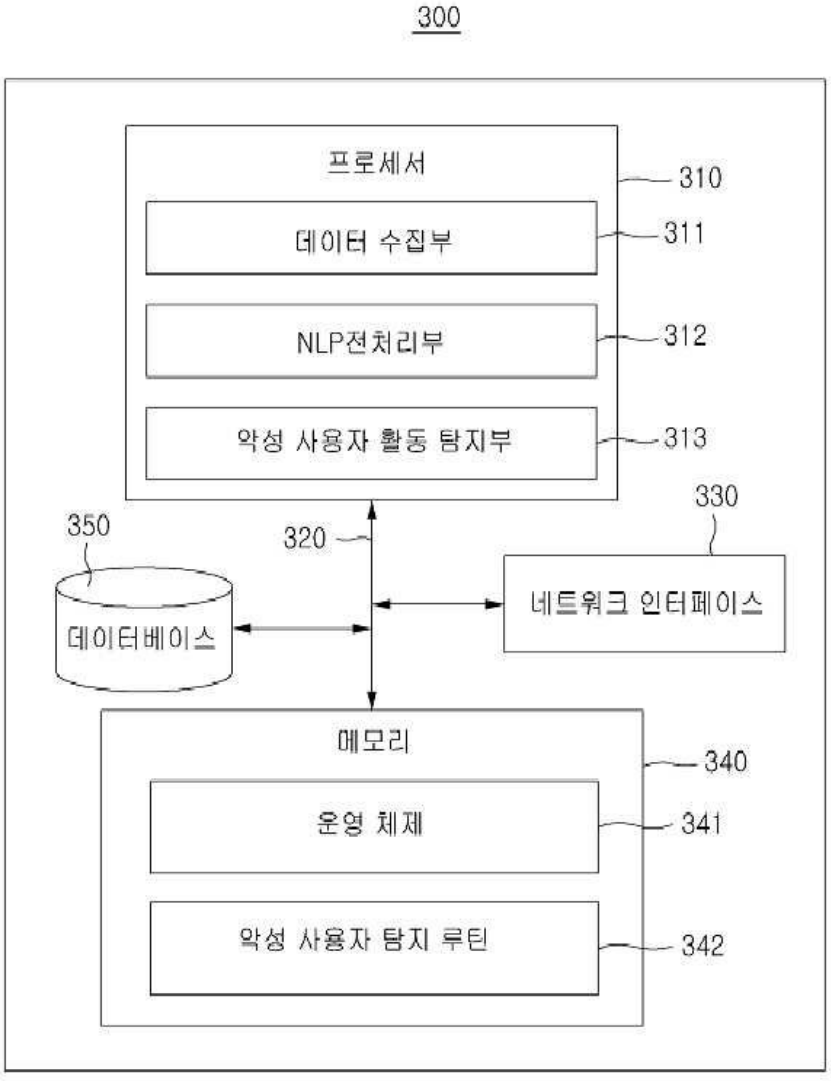
도면1


Tree 1
Sample 1
Training set
Sample N
Tree N
Tree 1
Tree 2
Tree 3
Majority Vote
Random Forest decision

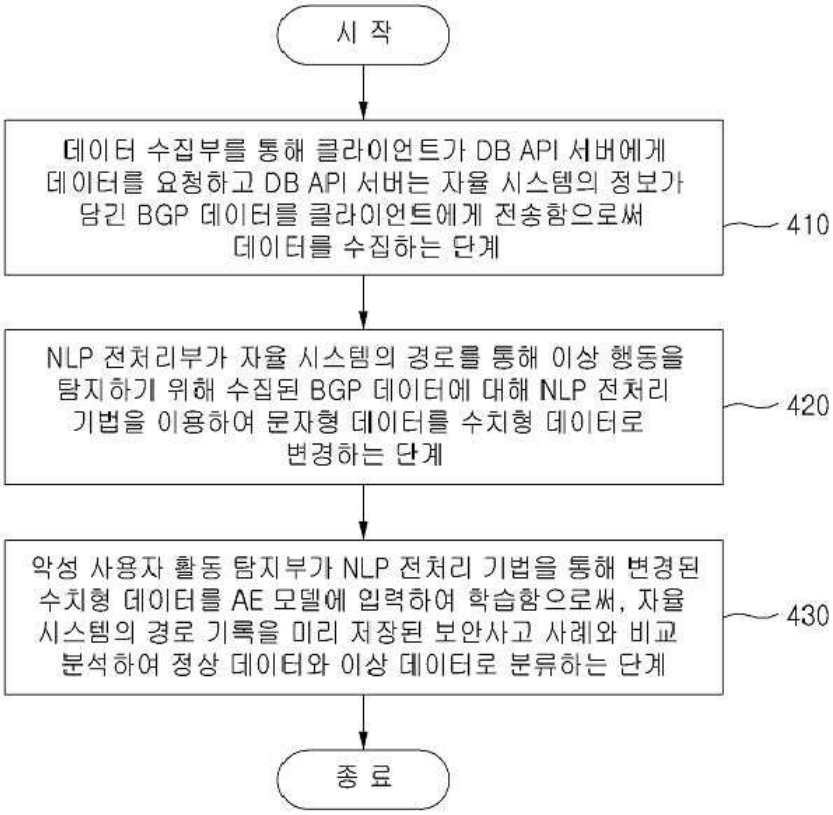
도면2


Encoder
Decoder

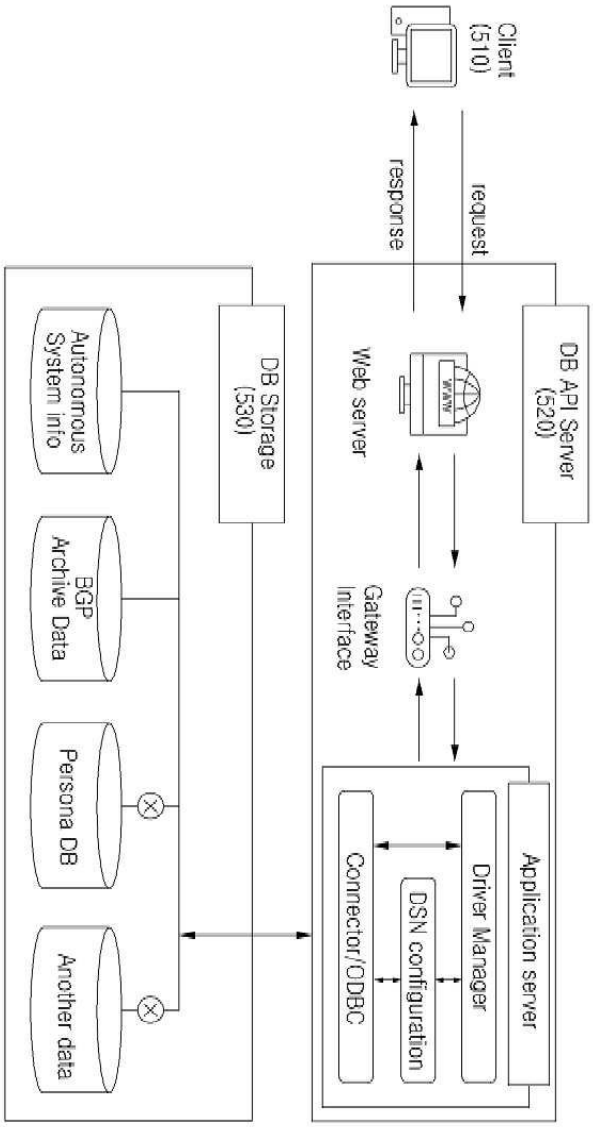
도면3


300
프로세서
310
데이터 수집부
311
NLP전처리부
312
악성 사용자 활동 탐지부
313
330
350
320
네트워크 인터페이스
데이터베이스
메모리
340
운영 체제
341
악성 사용자 탐지 루틴
342

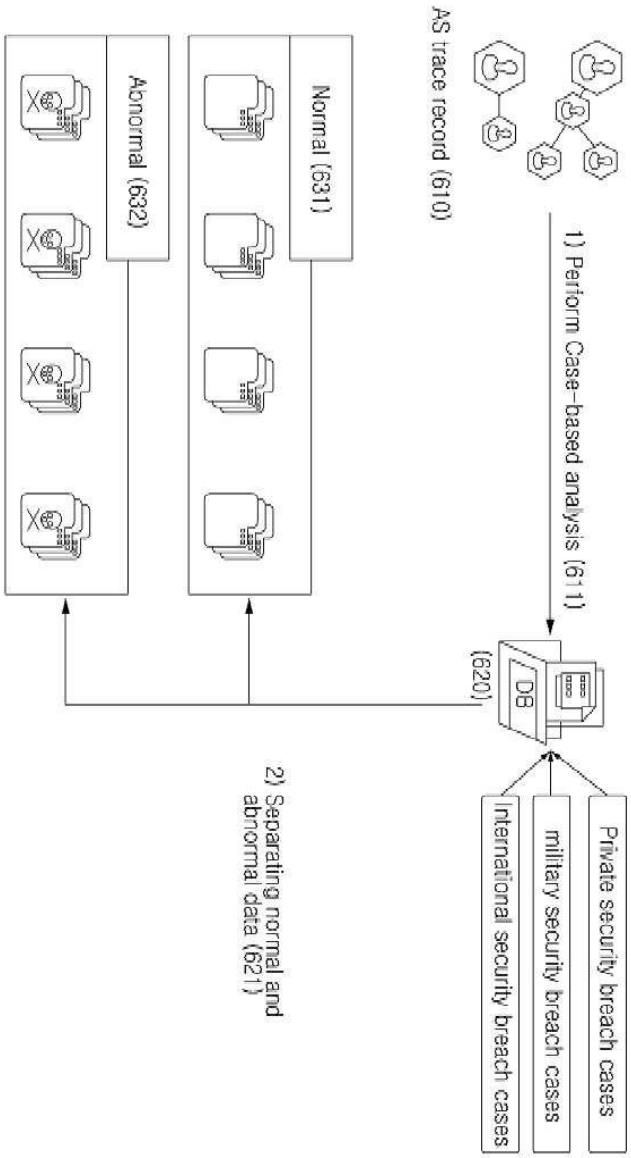
도면4


시 작
데이터 수집부를 통해 클라이언트가 DB API 서버에게 데이터를 요청하고 DB API 서버는 자율 시스템의 정보가 담긴 BGP 데이터를 클라이언트에게 전송함으로써 데이터를 수집하는 단계
410
NLP 전처리부가 자율 시스템의 경로를 통해 이상 행동을 탐지하기 위해 수집된 BGP 데이터에 대해 NLP 전처리 기법을 이용하여 문자형 데이터를 수치형 데이터로 변경하는 단계
420
악성 사용자 활동 탐지부가 NLP 전처리 기법을 통해 변경된 수치형 데이터를 AE 모델에 입력하여 학습함으로써, 자율 시스템의 경로 기록을 미리 저장된 보안사고 사례와 비교 분석하여 정상 데이터와 이상 데이터로 분류하는 단계
430
종 료

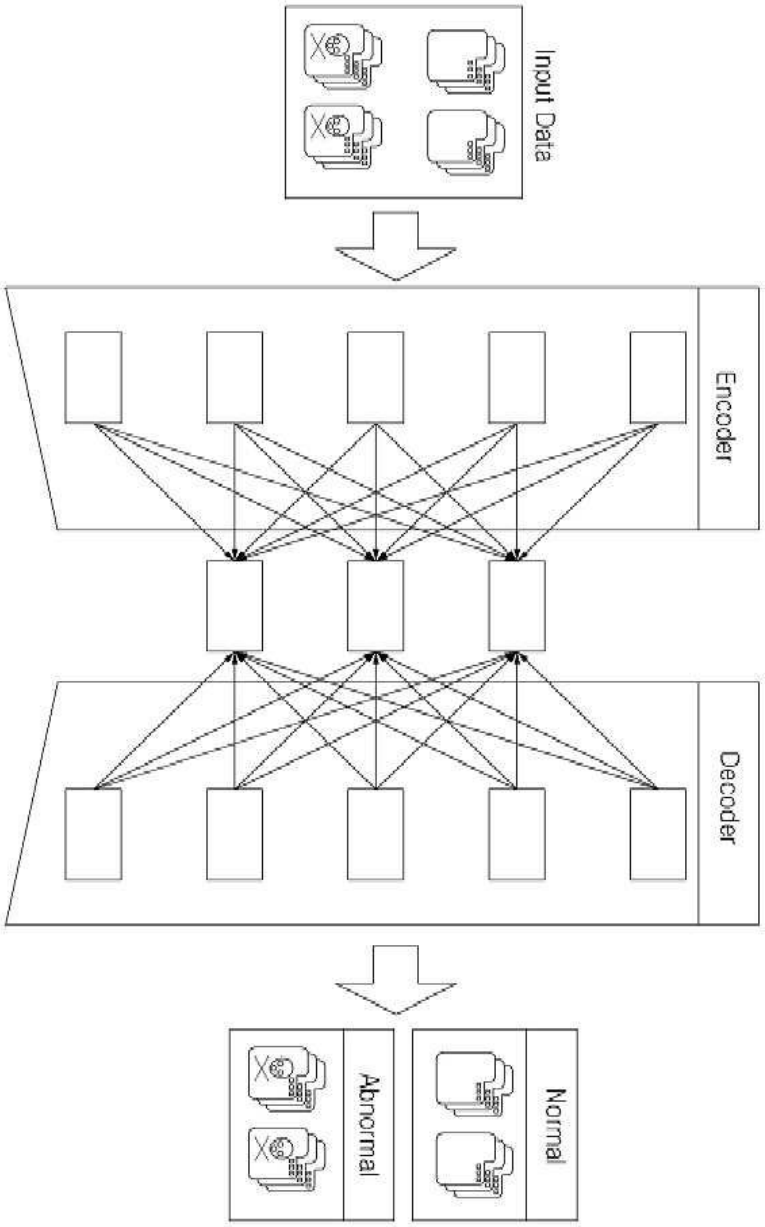
도면5


Client
(510)
response
request
DB API Server
(520)
Web server
Gateway
Interface
Application server
Driver Manager
DSN configuration
Connector/ODBC
DB Storage
(530)
Autonomous
System info
BGP
Archive Data
Persona DB
Another data

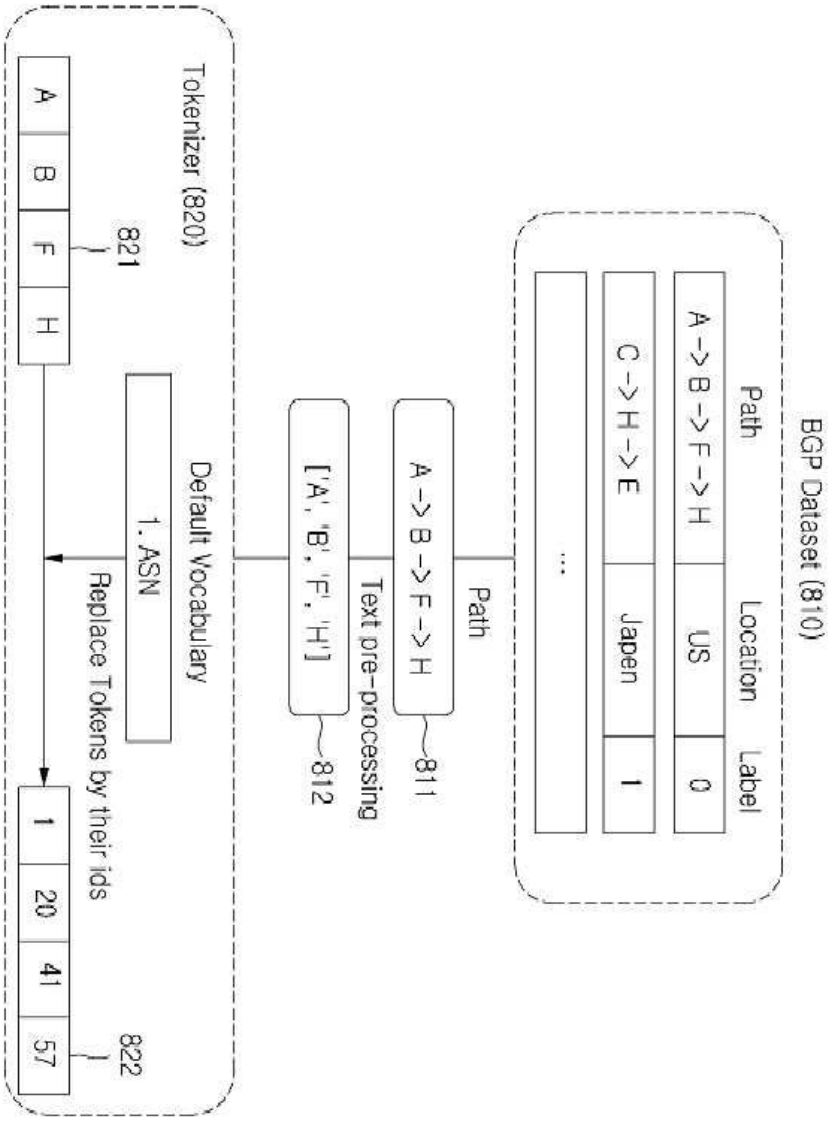
도면6


AS trace record (610)
1) Perform Case-based analysis (611)
DB
(620)
Private security breach cases
military security breach cases
International security breach cases
2) Separating normal and abnormal data (621)
Normal (631)
Abnormal (632)

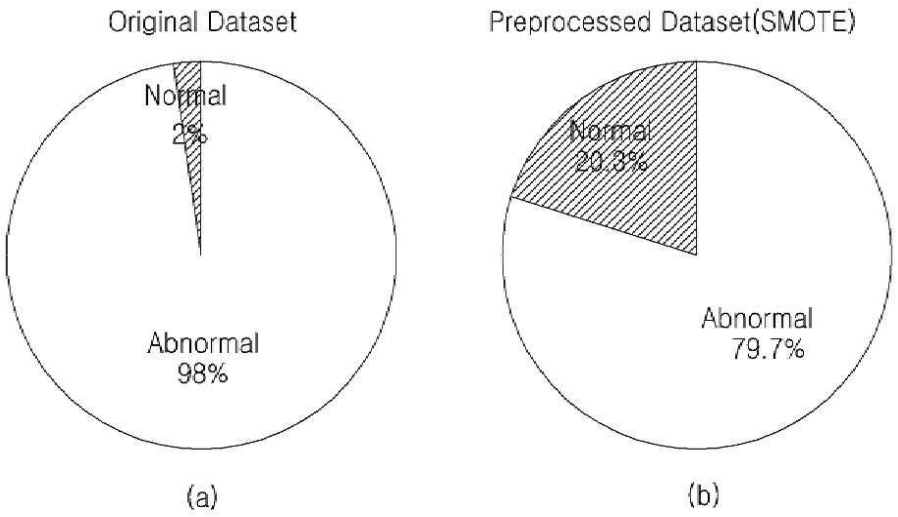
도면7



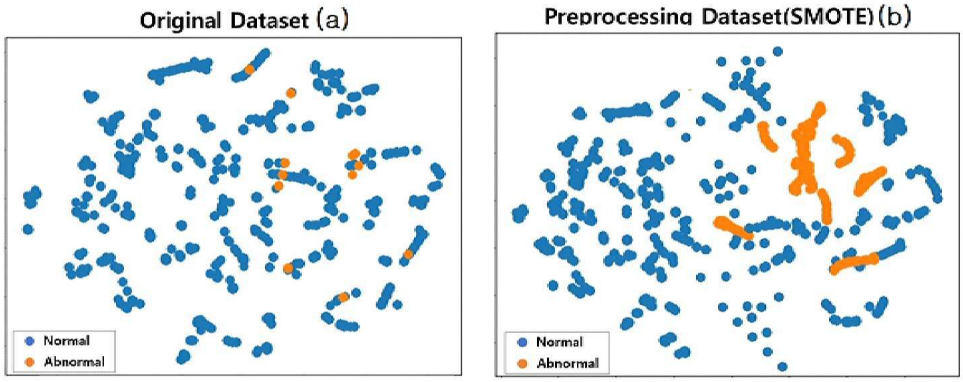
도면8


BGP Dataset (810)
Path
Location
Label
A -> B -> F -> H
US
0
C -> H -> E
Japen
1
...
Path
A -> B -> F -> H
811
Text pre-processing
['A', 'B', 'F', 'H']
812
Tokenizer (820)
Default Vocabulary
1. ASN
821
A
B
F
H
Replace Tokens by their ids
1
20
41
57
822

도면9



도면10



도면11

