모바일 엔드포인트 행동 분석 기반 Policy Engine 구현

팀명: KGL
지도교수: 최윤호
202255655 권태현
202255654 구현서
202255668 이승원

목차

1. 연구 배경 및 목적

1.1 연구 배경

디지털 서비스가 계속 확산되는 가운데에 사용자 인증과 네트워크 접근 통제의 중요성이 더욱 부각되고 있다. 특히 이제 한국에서는 스마트폰을 가지고 있지 않은 사람을 찾기 힘들 정도로 대부분의 인구가 스마트폰을 활용하여 많은 서비스를 제공 받고 있다. 모바일 환경에서 생체 정보, 패턴 입력, 2단계 인증 등 다양한 인증 방식이 도입되고 있지만, 여전히 사용자 도용, 세션 하이재킹, 앱 위변조 등 복합적인 보안 위협에 노출되어 있다.

2025년 5대 사이버 보안 위협 전망(출처 : 안랩 제공)

나날이 발전하는 보안 공격수단으로 인해 기존의 정적인 보안 시스템만으로는 사용자를 보호하기 힘들어졌으며 실시간으로 일어나는 이상 행위를 탐지하고 유연하게 대응하기 어렵다는 문제도 계속 제기되고 있다. 이러한 문제를 해결하기 위해, 사용자의 행동 패턴을 기반으로 이상 행동을 탐지하고, 이를 통해 네트워크 접근을 동적으로 제어할 수 있는 시스템이 필요하다고 생각했다.
한편, 머신러닝 기술의 발전은 이상 탐지 문제를 자동화하고 정교하게 수행할 수 있는 가능성을 열어주었다. 명확한 이상 데이터가 부족한 환경에서도 정상 행동을 학습한 뒤 이상을 판단하는 비지도 학습 방식을 활용하면 현실적으로 동적인 네트워크 접근 시스템을 구축할 수 있을 것이라는 가능성을 보여줬다.
본 연구는 동적인 보안 시스템의 필요성과 기술적 흐름을 바탕으로 기획되었다.

1.2 목적

본 연구의 목적은 모바일 사용자로부터 수집된 로그 데이터를 기반으로, 비지도 학습을 통해 이상 탐지 모델을 구축하고, 이를 기반으로 실시간으로 네트워크 접근을 동적으로 제어하는 시스템을 개발하는 것이다.

구체적으로 다음과 같은 목표를 달성하고자 한다.

1. 모바일 앱을 통해 사용자의 행동 데이터를 자동으로 수집하고 서버로 전송하는 구조 설계
2. 서버에서 수집된 데이터를 저장 및 전처리한 후, 머신러닝 서버에 분석을 요청하는 분산 구조 시스템 설계
3. 이상 탐지 모델을 활용하여 정상/비정상 행위를 판별하는 로직 구현
4. 분석 결과에 따라 실시간으로 네트워크 접근을 제어하고, 관리자가 대시보드에서 상황을 모니터링할 수 있도록 기능 제공

궁극적으로, 기존의 보안 시스템보다 지능적이고 유연한 보안 시스템의 프로토타입을 구현하는 것을 목표로 한다.

2. 문제 정의 및 해결 전략

2.1 사용자 행동 데이터 수집 및 구조화의 어려움

스마트폰 또는 클라우드 기반 시스템에서 사용자의 다양한 행동 또는 이벤트(접속,클릭,다운로드 등)가 로그로 기록되는데 이러한 데이터는 구조화된 형태가 아닌 비정형적이며 포맷이 일정하지 않아 구조화가 어렵다. 예를 들어, 네트워크 요청의 경우 IP, 포트, 리소스 경로 등을 포함하며 터치 이벤트는 좌표와 시간, 압력 등을 포함한다. 이러한 다양한 형태의 데이터를 통합하여 분석 가능한 형태로 바꾸는 것은 어려운 과제이다. 또한 프라이버시 보호를 위해 실제 사용자 데이터를 수집하거나 활용하기에 어려운 경우가 많아 실제 연구에서는 합성(Synthetic)데이터에 의존하는 경우가 많다.

-해결책-

- Isolation Forest & LSTM Auto-Encoder 기반 이상 탐지 모델

요청을 구조적으로 분석하기 위해 데이터를 세가지 범주로 나누어 정형화한다.

행동 로그 속성 분류표

속성 유형	정의	수집 방법
사용자 속성	요청을 보낸 사용자의 신원과 접속 환경에 대한 정보. 예: 누구인지, 어떤 장비로 접속했는지, 접속 위치(IP 기반) 등	- 로그인 정보 - 연결된 장비 정보 - IP 기반 지역 정보
요청 속성	사용자가 어떤 자원에 어떤 작업을 언제, 어떤 방식으로 요청했는지를 나타냄. 예: 접근 대상, 요청 시간, 요청 방식 (GET, POST 등)	- 요청 로그 파싱 - URL 및 요청 방식 분석
통계적 속성	요청 간 시간 간격, 실패율, 요청 빈도 등을 기반으로 사용자의 행동 패턴을 수치화한 특성. 이를 통해 새로운 요청이 기존 패턴과 얼마나 다른지 분석하고 이상 여부를 판단함	- 과거 요청의 시간 간격, 성공률 등 DB 조회 - 세션 단위 통계 계산

- CLUE-LDS

공개 로그 데이터셋으로 5000명 이상의 사용자가 5년간 생성한 실제 클라우드 기반 시스템 로그를 포함하고 있다. 이 데이터들은 JSON포맷으로 되어있으며 전처리만 하면 바로 머신러닝 모델 학습에 활용할 수 있다. 따라서 별도로 데이터를 수집하지 않고 실제 사용자 행동 기반 데이터로 활용할 수 있다.
github:https://github.com/ait-aecid/clue-lds?tab=readme-ov-file

2.2 이상 행동 판단 기준의 모호성

사용자 행동은 시간이 지남에 따라 본질적으로 변화할 수 있으며 정상 행동과 이상 행동의 경계가 항상 명확하지 않다. 예를 들어 평소에는 낮에 접속하던 사용자가 어느 날 밤에 접속하거나, 평소보다 더 빠르게 클릭 또는 반복하는 경우 이는 이상 행동일 수도 아닐 수도 있다.
또한 이상 행동에 대한 라벨(정답 데이터)이 부족하며, 이에 따라 지도학습 기반 모델은 학습에 제약을 받는다.

-해결책-

- Isolation Forest & LSTM Auto-Encoder

정상행동만을 학습하기에 라벨이 적은 상황에서도 활용이 가능한 비지도 학습모델인 Isolation Forest를 사용하고, 시계열 데이터 처리를 위한 모델인 LSTM Auto-Encoder를 사용한다. 정상 행동만 학습을 한 후, 새 요청이 들어왔을 때 이를 재구성하고 오차가 클경우 이상 행동으로 간주한다. 이 방식을 사용하면 사전 라벨 없이도 실시간 이상 탐지가 가능하다.

- CLUE-LDS

공격 시나리오 삽입 기능을 제공해준다. 즉, 정상 로그에 인위적으로 비정상적인 접근 또는 행동을 포함시켜 이상 탐지 모델을 학습하고 테스트할 수 있는 환경을 구성한다. 이로 인해 이상 판단 기준을 정량적으로 실험할 수 있다.
github:https://github.com/ait-aecid/clue-lds?tab=readme-ov-file

2.3 행동 모니터링에서의 개인정보 보호 고려사항

행동 로그에는 IP, 기기 ID등 개인을 식별 할 수 있는 민감 정보가 들어가 있다. 이런 데이터들은 윤리적 기준과 개인정보 보호법에 따라 수집, 공유, 보관 등이 제한된다. 따라서 실시간 행동 모니터링을 하면서도 개인을 식별하지 않는 유의미한 분석이 가능해야 한다.

-해결책-

사용자의 정보를 이름이나 계정이 아닌 숫자 기반 속성 벡터로 변환하여, 데이터 자체에는 식별자가 포함되지 않도록 처리한다. 즉, 실명-ID 없이 익명화된 행동패턴만을 사용한다.
이로써 개인에 대한 식별은 불가능하지만 여전히 분석은 가능하다. 로컬 디바이스에서만 연산을 수행하거나 서버에 전송할 경우 암호화를 적용하는 방법을 사용할 수도 있다.

2.4 모바일 기기의 자원 제약 사항 문제점

스마트폰은 데스크탑에 비해 CPU,메모리, 배터리 지속 시간 등 여러 부분에서 제한되어 지속적인 모니터링, 실시간 모델 추론에 어려움이 있다. 즉, 머신러닝 모델을 모바일에서 직접 학습 또는 실행하는 것은 현실적으로 무리가 있다.

-해결책-

복잡한 연산이 필요한 모델 학습과 판단 과정을 서버 측에서 수행하도록 하고 모바일 기기는 요청 속성만 추출하여 서버에 전송하도록 한다. 이를 통해 모바일 연산 부담을 최소화 시킬 수 있다.
데이터 수집을 5~20초 단위로 분할하여 처리하는 방법을 사용하여 서버 연산부담도 최소화할 수 있을 것이다. 이로서 실시간 인증이 가능하면서도 자원 소모를 최소화 시킬 수 있다.

3. 시스템 구성 및 데이터 흐름도

3.1 전체 시스템 아키텍처

1. 모바일 클라이언트
-사용자의 터치, 위치, 요청 정보를 수집한다.
-전처리 후 벡터화하여 서버로 전송한다.
-가벼운 속성만 추출하여 리소스 소비를 최소화한다.

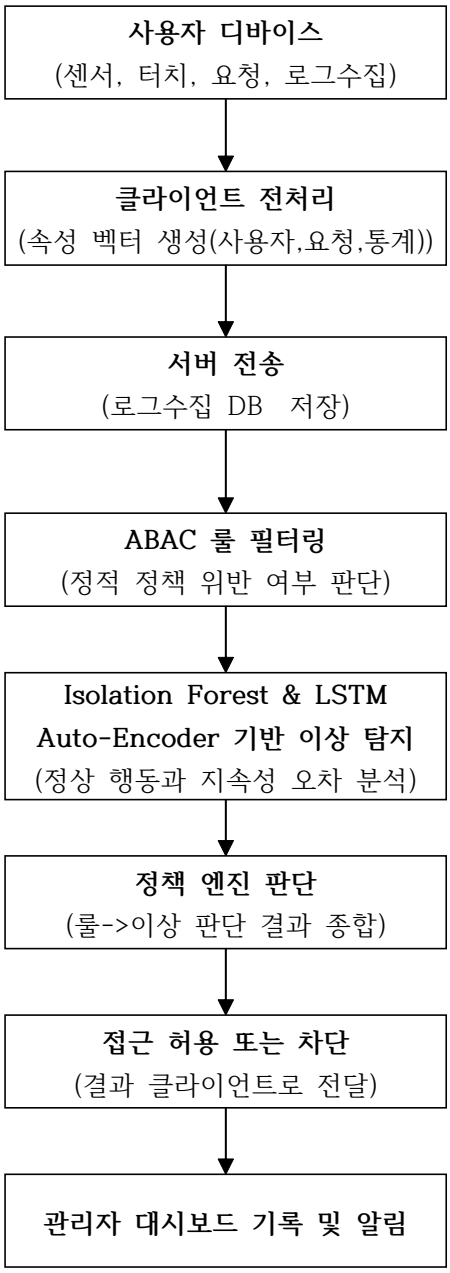
2. 로그 수집 서버
-클라이언트로부터 전송된 속성 데이터를 수신한다.
-DB에 저장하고 실시간 요청 흐름을 관리한다.

3. 행동 분석 및 이상 탐지 모듈(Isolation Forest & LSTM Auto-Encoder)
-사용자 속성, 요청 속성, 통계 속성 입력한다.
-정상 행동만 학습 후 재구성 오차 기반 이상 판단한다.

4. 정책 엔진(ABAC 룰 매칭)
-정적 규칙 기반 필터링: 역할, 위치, 시간대 등을 고려하여 1차 판단한다.
-이상 탐지 결과와 결합하여 최종 접근 허용 여부 결정한다.

5. 대시보드 및 관리자 인터페이스
-감지된 이상 행동을 시각화한다.
-차단 여부, 의심 사용자 현황, 정책 수정 기능을 제공한다.

3.2 데이터 흐름도



3.3 감지된 이상에 대한 대응 메커니즘

시스템이 이상 행동을 탐지했을 때 취할 수 있는 대응 조치를 설계한다.

1. 이상 판단 후 차단 여부 결정
-임계값을 초과한 요청에 대해 경고, 자동 차단 또는 조건부 차단한다.

2. 알림 전송
-대시 보드 및 이메일, 알림 API로 관리자에게 이상 사용자 발생 알림.

3. 로그 추적 기록 강화
-해당 세션의 모든 행동을 고정 로그로 백업한다.

4. 반복적인 이상 요청 발생시
-계정 일시 정지, 경고를 보내거나 추가 인증 요청(FIDO,OTP) 유도한다.

4. 개발 환경 및 기술

4.1 사용 언어 및 프레임워크

Python, Kotlin을 기반으로 하는 서버-클라이언트 구조이며 Django를 통해 REST API를 설계하고 머신러닝 모델과 연결된다.

● 백엔드 서버
-Python + Django
-사용자 행동 데이터 수신 및 처리, 이상 탐지 모델과 연동, API제공

● 머신러닝
-Python(TensorFlow)
-모델 구현 및 학습 수행

● 모바일 클라이언트
-Kotlin(Anodroid Studio)
-사용자 터치/센서 정보 수집 및 서버 전송 에이전트 개발

● 스크립트 및 데이터 전처리
-Python(Pandas, Numpy 등)
-CLUE_LDS 데이터셋 전처리, 속성 벡터 추출 및 저장 처리

4.2 머신러닝 모델링 도구 및 알고리즘

사용할 모델: Isolation Forest & LSTM Auto-Encoder

● 모델 목적
-정상 사용자 행동을 학습하고 이후 요청이 정상 범위 패턴 행동인지 판단

● 사용 목적
-비지도 학습 구조, 라벨없이도 이상 탐지가 가능하다.

● 입력 데이터
-사용자,요청,통계 속성을 벡터로 변환한 입력값

● 출력 결과
-입력을 재구성한 결과와의 오차를 계산한다. 즉, 이상여부 판단함.

● 사용 도구
-데이터 전처리(Python, Pandas, Scikit-learn)
-모델 설계, 학습(TensorFlow)
-시각화 및 평가(Matplotlib, Seaborn, sklearn.metrics 등)

4.3 모델 성능평가

모델이 이상 행동을 잘 탐지하는지에 대해 성능평가를 진행한다.

1. Isolation Forest & LSTM AE로 이상치 점수를 생성
-의심 로그에 대해 이상치 점수를 생성함. 샘플의 이상치 점수를 데이터셋에 저장.

2. 상위 이상치를 가진 샘플 1%~5% 추출
-이상치가 높은 상위 샘플에 대해 수동 검토 진행 -> 라벨링

3. 라벨링된 샘플에 대해 학습 데이터 + 평가데이터셋을 구성
-데이터 셋에 대해 로지스틱 회귀분석을 수행하여 차단이 정당했는지 여부를 평가한다. 평가지표로는 precision, recall, F1-score을 사용한다.

4.4 인프라 구성

● 로그 수집 DB
사용자 요청 로그 및 전처리 속성 저장(SQLite)

● 모델 저장소
학습된 모델 파일 저장(.h5)

● 정책 관리 DB
ABAC 룰셋 저장 및 갱신용 테이블 (ex. 사용자 그룹, 시간 조건 등)

● 대시보드 서버
이상 탐지 결과와 요청 로그를 시각화하는 관리자용 웹페이지

4.5 배포 및 테스트 도구

● API 테스트
-Postman
-Django 서버의 API 응답 및 이상 탐지 결과 확인한다.

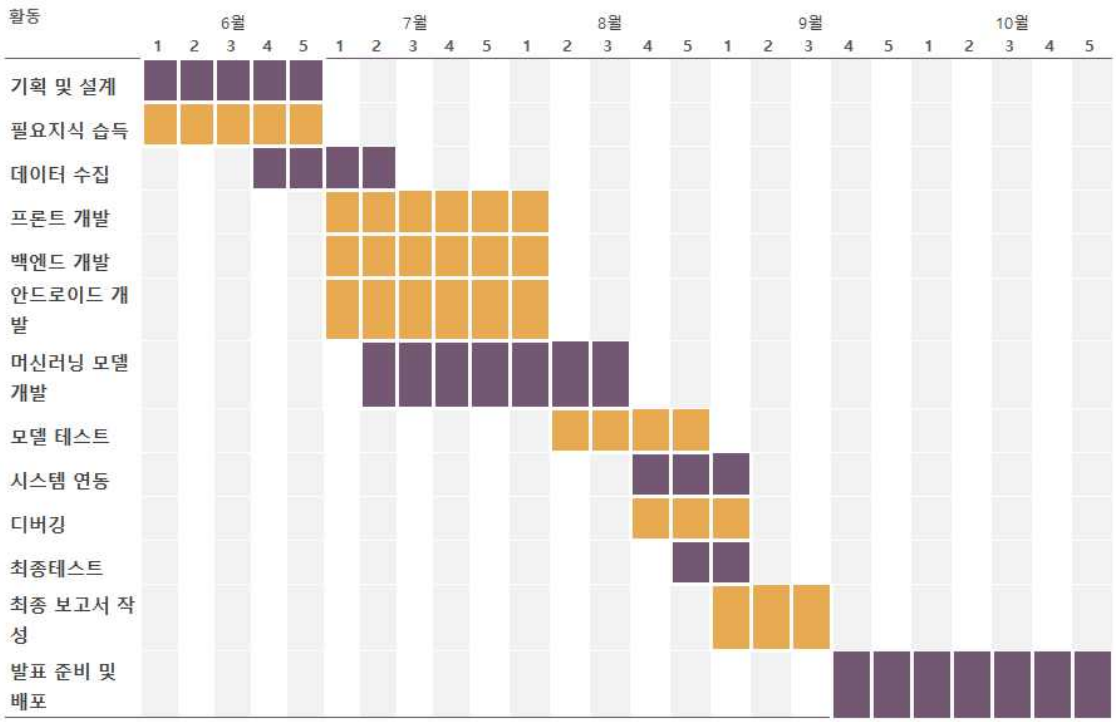
● 모델 추론 테스트
-Jupyter Notebook
-모델의 이상 판단 성능 시뮬레이션 및 시각화

● 배포 환경
-Linux

● 앱 테스트
-Android Emulator + 안드로이드 스마트폰
-터치,센서 수집 앱 테스트 및 서버 통신 검증

● 버전 관리
-Git, GitHub
-코드 관리 및 협업 , 버전 기록용

5. 개발 일정



6. 역할 분담

팀원	**역할**
권태현	-머신러닝 모델 설계 및 학습 -데이터 전처리 -이상 탐지 결과 시각화를 위한 대시보드 UI 개발
구현서	-모바일 어플리케이션 및 UI 개발 -사용자 행동 데이터 수집
이승원	-사용자 행동 데이터 수신 및 처리를 위한 DB 개발 -정책 관리 DB 개발 -이상 탐지 모델과 연동