(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년07월15일
(11) 등록번호 10-2422307
(24) 등록일자 2022년07월13일

(51) 국제특허분류(Int. Cl.)
H04W 4/029 (2018.01) *H04W 4/60* (2018.01)
H04W 88/18 (2019.01)
(52) CPC특허분류
H04W 4/029 (2020.05)
H04W 4/60 (2018.02)
(21) 출원번호 10-2021-0026858
(22) 출원일자 2021년02월26일
심사청구일자 2021년02월26일
(56) 선행기술조사문헌
KR1020170139776 A*
KR102182675 B1*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
충남대학교산학협력단
대전광역시 유성구 대학로 99 (궁동, 충남대학교)
(72) 발명자
원유재
대전광역시 유성구 온천로 45, 알파동 805호(봉명동)
전석준
충청북도 청주시 흥덕구 2순환로1375번길 20(가경동, 가경 푸르지오)
(뒷면에 계속)
(74) 대리인
특허법인 공간

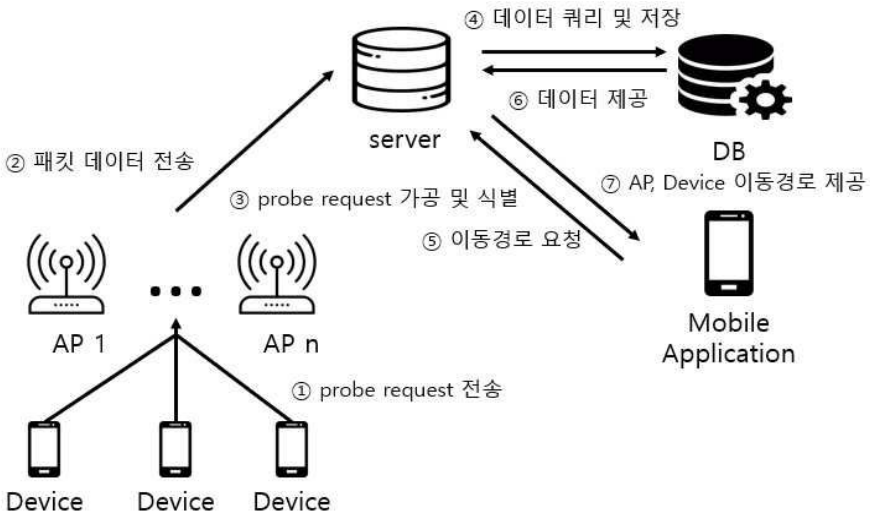
전체 청구항 수 : 총 1 항 심사관 : 선동국

(54) 발명의 명칭 **무선단말을 이용한 사용자 이동경로 추적 시스템**

(57) 요 약

본 발명에 의한, 무선단말을 이용한 사용자 이동경로 추적 시스템은, 무선 단말 식별 및 분류 모듈을 포함하는 서버, 사용자 정의 펌웨어가 설치된 AP, 무선단말을 포함하고, AP에서는 무선단말의 패킷을 수집하여 서버로 전송하고, 서버에서는, 단말 식별 및 분류 시스템이 동작하여 수집한 무선 패킷에서 프로브 리퀘스트 프레임과 WLAN 매니지먼트 프레임을 필터링하여 저장한 뒤 이를 이용하여 단말 식별 및 분류 모델을 생성 및 학습하고, 그 결과를 이용하여 무선단말을 식별하고, 그 무선단말의 식별결과를 이용하여 각 무선단말의 이동경로를 추적하는 것을 특징으로 한다.

대 표 도 - 도1



(52) CPC특허분류
H04W 88/18 (2019.01)
(72) 발명자
정하영
대전광역시 서구 도안동로 123, 1706동 2304호(도안동, 도안리슈빌)
김용환
충청북도 옥천군 군북면 이백3길 15-16
조영우
대전광역시 유성구 농대로17번길 24, 302호(궁동)

이 발명을 지원한 국가연구개발사업

과제고유번호	2019001343
과제번호	2019-0-01343
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	융합보안핵심인재양성사업
연구과제명	융합보안핵심인재양성사업
기 여 율	1/1
과제수행기관명	충남대학교 산학협력단
연구기간	2020.04.01 ~ 2023.12.31

공지예외적용 : 있음

명 세 서

청구범위

청구항 1

무선 단말 식별 및 분류 모듈을 포함하는 서버, 사용자 정의 펌웨어가 설치된 AP, 무선단말을 포함하고,

AP에서는 무선단말의 패킷을 수집하여 서버로 전송하고,

서버에서는, 단말 식별 및 분류 시스템이 동작하여 수집한 무선 패킷에서 프로브 리퀘스트 프레임과 WLAN 매니지먼트 프레임을 필터링하여 저장한 뒤 이를 이용하여 단말 식별 및 분류 모델을 생성 및 학습하고, 그 결과를 이용하여 무선단말을 식별하고,

그 무선단말의 식별결과를 이용하여 각 무선단말의 이동경로를 추적하고,

프로브 리퀘스트 프레임과 WLAN 매니지먼트 프레임을 필터링하여 저장할 때, 프로브 리퀘스트 프레임에서 사용할 데이터 필드는, 첫 번째 패킷 수신 이후 패킷이 수신된 시간에 대한 데이터 필드(frame.time_relative) , 해당 Probe request에 부여된 Number에 관한 데이터 필드(wlan.seq), Probe request의 목적지 AP의 SSID에 관한 데이터 필드(wlan.ssid), Probe request의 전체 길이에 관한 데이터 필드(frame.len)로 구분되고, 해당 Probe request에 부여된 Number에 관한 데이터 필드(wlan.seq)는 1 순환 주기마다 4096의 배수를 곱하여 계속 증가하는 값으로 가공되고,

WLAN Management Frame에서 사용할 데이터는, MPDU(MAC Protocol Data Unit)의 밀도값에 관한 데이터(wlan.ht.ampduparam.mpdudensity), Wi-Fi 카드의 물리적인 기능과 관련된 정보 데이터(wlan.ht.capabilities), STBC(Space Time Block Code) 수신 값 데이터(wlan.ht.capabilities.rxstbc), STBC(Space Time Block Code) 송신 값 데이터(wlan.ht.capabilities.txstbc)를 포함하고,

디바이스의 고유 지문을 추출하여 무선 통신 단말을 식별 및 분류하여 화이트리스트를 구축할 때 생성할 특징(Feature)는 Sequence Number 증가율과 패킷의 길이이고,

첫 번째 패킷 수신 이후 패킷이 수신된 시간에 대한 데이터 필드(frame.time_relative), 해당 Probe request에 부여된 Number에 관한 데이터 필드(wlan.seq)를 Sequnce Number 증가율 측정에 사용하고,

Probe request의 목적지 AP의 SSID에 관한 데이터 필드(wlan.ssid), Probe request의 전체 길이에 관한 데이터 필드(frame.len)를 패킷의 길이 측정에 사용하고,

MPDU(MAC Protocol Data Unit)의 밀도값에 관한 데이터(wlan.ht.ampduparam.mpdudensity), Wi-Fi 카드의 물리적인 기능과 관련된 정보 데이터(wlan.ht.capabilities), STBC(Space Time Block Code) 수신 값 데이터(wlan.ht.capabilities.rxstbc), STBC(Space Time Block Code) 송신 값 데이터(wlan.ht.capabilities.txstbc)를 무선 단말 분류 및 식별을 위해 사용하고,

무선 단말 분류를 할 때, 먼저 단말의 Feature를 모델을 통해 검증하는데, 모델을 통해 분류된 단말과 같을 예측 확률이 60퍼센트 이상일 경우 해당 단말과 같은 종류로 판단하여 같은 Label을 지정하고, 예측 확률이 60퍼센트 미만인 경우 이전에 학습한 적 없는 단말로 간주하여 새로운 Label을 할당하는 것을 특징으로 하는 무선단말을 이용한 사용자 이동경로 추적 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 무선단말을 이용한 사용자 이동경로 추적 시스템에 관한 것으로, 보다 상세하게는, AP를 활용하는, 무선단말을 이용한 사용자 이동경로 추적 시스템에 관한 것이다.

배 경 기 술

[0002] 최근 코로나 19로 인해, 사람들의 이동경로를 추적할 필요가 커지고 있는데, 무선단말(예들 들면, 스마트폰)을

이용한 이동경로 추적이 효과적인 경우가 많다.

[0003] 종래에도 무선단말을 이용한 이동경로 추적 기술은 개발되어 있고, 특허문헌 1 내지 3은 그 예이다.

[0004] 특허문헌 1은 이동 단말의 위치 측정 시스템 및 그 방법에 관한 것이고, 특허문헌 2는 단말의 위치를 추정하는 방법 및 장치에 관한 것이고, 특허문헌 3은 실내 사용자의 위치를 추적하여 재난 관제를 제공하는 장치 및 방법에 관한 것이다.

[0006] 그러나 종래의 이동경로 추적 시스템은 비용에 비해 정확도와 효율성이 높지 않았다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 한국 공개특허공보 10-2010-0124649 A (공개일: 2010. 11. 29.)

(특허문헌 0002) 한국 등록특허공보 10-1711791 B1 (등록일: 2017. 2. 23.)

(특허문헌 0003) 한국 공개특허공보 10-2020-0049340 A (공개일: 2020. 5. 8.)

발명의 내용

해결하려는 과제

[0009] 본 발명이 해결하고자는 하는 과제는, 정확도와 효율성이 높은, 무선단말을 이용한 사용자 이동경로 추적 시스템 을 제공하는 것이다.

과제의 해결 수단

[0011] 본 발명에 의한, 무선단말을 이용한 사용자 이동경로 추적 시스템은, 무선 단말 식별 및 분류 모듈을 포함하는 서버, 사용자 정의 펌웨어가 설치된 AP, 무선단말을 포함하고, AP에서는 무선단말의 패킷을 수집하여 서버로 전송하고, 서버에서는, 단말 식별 및 분류 시스템이 동작하여 수집한 무선 패킷에서 프로브 리퀘스트 프레임과 WLAN 매니지먼트 프레임을 필터링하여 저장한 뒤 이를 이용하여 단말 식별 및 분류 모델을 생성 및 학습하고, 그 결과를 이용하여 무선단말을 식별하고, 그 무선단말의 식별결과를 이용하여 각 무선단말의 이동경로를 추적하는 것을 특징으로 한다.

발명의 효과

[0013] 본 발명에 의한, 무선단말을 이용한 사용자 이동경로 추적 시스템은 정확도와 효율성이 높다.

도면의 간단한 설명

[0015] 도 1은 무선신호 기반 사용자 이동 경로 추적 시스템의 전체 구조를 도식화한 그림

도 2는 디렉토리의 계층 구조의 개념도

도 3은 무선 단말 분류 과정의 개념도

도 4는 무선 단말 식별 과정의 개념도

도 5는 단말 모드의 작동방식을 예시한 도면

도 6은 AP 모드의 작동방식을 예시한 도면

도 7은 무선 단말 식별 및 분류 시스템의 정확도 검증시 구성한 실험 환경의 개념도

도 8은 devN.csv 파일을 저장한 예

도 9는 featured.csv 파일로 저장한 예

발명을 실시하기 위한 구체적인 내용

[0016] 학교, 호텔, 카페, 병원 등을 포함한 각종 공용 시설과 가정에 무선 AP(Access Point, 액세스 포인트, 무선 LAN에서 기지국 역할을 하는 소출력 무선기기)가 널리 보급된 높은 수준의 무선통신망 인프라가 확보되어 있다.

[0017] 이에 따라 무선 통신 단말과 AP간에 주고받는 무선신호를 통해 대상을 식별하는 기술의 연구가 진행되었다. 그러나 하나의 단말이 여러 AP에서 감지된 경우 같은 단말로 식별하기 난해한 문제가 있기 때문에 이와 관련된 기술의 개발이 요구되고 있다.

[0019] 무선 네트워크는 유선 네트워크에 비해 주위 환경의 영향을 많이 받으며, 이에 따라 수집한 특징값에 잡음으로 인한 오차가 발생하는 경우 정확한 식별이 어렵다는 문제가 있다. 기존의 연구는 이를 해결하기 위해 고가의 무선신호 수집 장비를 사용하므로 범용적으로 사용되는 기존의 무선통신망 환경에서는 활용하기 어렵다. 선행연구인 기계학습을 이용한 무선 통신 디바이스 식별 시스템[서정훈, 기계학습을 이용한 무선 통신 디바이스 식별 시스템, 2020, 1-3p.]에서는 이 단점을 보완하기 위해 저가의 무선 NIC와 기계학습을 통한 핑거프린팅 기법을 이용하였다. 그러나 선행 연구에서는 학습할 무선 통신 디바이스의 화이트리스트를 구축한 뒤 단일 AP에서만 무선 통신 디바이스의 신호 데이터를 수집 및 식별함을 전제로 진행하였다. 그러므로 선행 연구에서 사용한 식별 방식으로는 이동 경로를 추적하는 데 필수적인 다수 AP에서의 동작을 보장하지 않는다고 할 수 있다. 따라서 본 발명에서 제안하는 시스템은 다수의 AP와 다수의 단말이 존재하는 환경에서 NIC와 기계학습을 이용해 각 AP에서 단말을 식별한 뒤 여러 AP에서 감지된 같은 단말들을 같은 단말끼리 분류하고, 이렇게 분류된 데이터를 모바일 애플리케이션 상에 출력하여 사용자가 단말의 이동 경로를 추적하는 것을 돕는다.

[0020] 제안하는 시스템은 AP에 사용자 정의 펌웨어를 설치하여 무선신호를 수집하고 이를 서버로 전달한 뒤 수집한 신호를 가공하여 개개의 AP마다 수집한 무선 통신 디바이스별로 고유한 지문을 추출한다. 그리고 잡음 및 오차에 대한 내성을 강화하고 같은 단말끼리 분류하기 위하여 추출한 지문을 기계학습 알고리즘의 Feature로 사용하여 무선 단말 분류 모델을 구축한다. 구축한 모델은 저장하여 이후 무선신호를 수집해 모델에 입력함으로써 단말 분류 결과를 알 수 있다. 최종적으로 분류된 결과와 감지된 위치, 감지된 시간의 정보를 데이터베이스에 저장하고 이를 모바일 애플리케이션 상에 표시함으로써 모바일 앱 이용자가 단말의 위치를 추적하기 쉽도록 돕는다.

[0021] 제안하는 시스템은 다양한 무선 네트워킹 방식 중 가장 보편적인 Wi-Fi를 대상으로 하며 본 출원에서 언급되는 무선 통신 디바이스는 스마트폰을 의미한다. 또한, 제안하는 시스템은 다수의 Wi-Fi 환경에서 어떠한 단말이 다른 시간에 다른 위치, 즉 다른 AP 상에서 감지되더라도 두 단말을 같은 단말로 분류해내는 것을 목표로 한다. 단말을 분류하는 것은 특정한 단말의 시간에 따른 이동 경로를 추적하는 데에 필수적인 요소이므로 본 출원에서는 무선 단말을 정확히 분류해내고 그 정확도를 개선하는 데에 주안점을 둔다.

[0023] Wi-Fi 스캔이란 IEEE 802.11 프로토콜을 이용하는 무선 통신망에 무선 단말이 참여하기 위해 호환 가능한 무선 통신망을 찾는 행위를 의미한다. Wi-Fi 스캔에는 수동 스캐닝과 능동 스캐닝 유형이 존재한다.

[0025] 1) 수동 스캐닝

[0026] 수동 스캐닝은 무선 단말의 입장에서 수동적으로 주위 AP를 스캔하는 방법으로, 무선 단말이 능동적으로 AP를 찾기 위한 행동을 취하지 않고도 AP가 자기 자신을 알리기 위해 전송하는 신호를 스캔하여 주위 AP를 인식하는 방법이다. 이 과정에서 AP가 보내고 무선 단말이 스캔하는 신호를 Beacon frame이라고 한다.

[0028] 2) 능동 스캐닝

[0029] 능동 스캐닝은 무선 단말의 입장에서 능동적으로 주위 AP를 스캔하는 방법이다. 능동 스캐닝에서 무선 단말은 근처 AP에 신호를 Broadcast하고 이 신호를 받은 AP들이 응답하는 신호를 단말에 직접 보내면 응답 신호를 수신한 무선 단말이 주위에 어떤 AP가 존재하는지 인식할 수 있다. 이 과정에서 무선 단말이 Broadcast하는 신호를 Probe request frame(프로브 리퀘스트 프레임), AP가 응답하는 신호를 Probe response frame(프로브 리스판스 프레임)이라고 한다.

[0030] 무선 단말이 자신의 정보를 담은 Probe request 패킷을 주위에 Broadcast하고 이를 수신한 AP는 자신의 정보를 담은 Probe response를 무선 단말에 직접 전송하여 이를 수신한 단말은 자신 주위에 어떠한 AP들이 존재하는지 인식할 수 있다.

[0031] Probe request frame은 관리 프레임으로서 단말의 SSID, MAC address, Channel 등의 정보를 포함하고 있으며 이러한 정보는 무선 단말이 AP를 인식 또는 연결하고자 할 때 사용된다.

[0032] Probe request(프로브 리퀘스트)는 추가적으로 Null probe request(널 프로브 리퀘스트)와 Directed probe request(디렉티드 프로브 리퀘스트)로 나누어진다. Null probe request는 무선 단말이 주위의 모든 AP를 탐지하

기 위해 사용하며 Directed probe request는 특정 AP로만 Probe request를 전송하는 경우에 사용한다. 둘을 구분하는 방법은 목적지 SSID의 존재 여부로, Frame body 내부의 SSID field는 Null probe request의 경우 null이며 Directed probe request 경우 해당 AP의 SSID값을 갖는다.

[0034] WLAN 관리 프레임은 Wi-Fi 스캔에서 주변을 스캔한 뒤 근처의 네트워크와 단말을 식별하는 과정에서 사용되는 프레임이다. 정당한 사용자인지를 검증하는 인증 절차, 이러한 인증 절차를 통과한 사용자와 프로토콜을 협의하고 결합하는 과정에서 사용되고 있다.

[0036] 기존의 위치 추적 기술 중 대표적인 기술이 GPS로 통신위성과 단말 간의 통신을 통해 이루어지는 위치 추적 기술이다. GPS는 경로 안내나 측정 등의 분야에서 널리 사용되는 기술이지만, 도심에서는 야외에 비해 정확도가 떨어지는 문제가 존재한다. 따라서 도심지나 실내에서는 무선 통신 디바이스가 발생하는 신호를 통해 단말의 위치를 추적하는 연구가 다양하게 진행되었다.

[0037] BLE beacon은 Bluetooth 4.0 이전의 기존 블루투스 기술의 커버리지와 데이터 전송 속도를 희생하여 전력 소모를 줄였다. 이처럼 낮은 커버리지를 가지기 때문에 BLE beacon 기반 위치 추적 기술 연구는 대부분 실내에서 다수의 Beacon을 설치하여 다양한 지점에서 감지되는 신호의 강도를 수집한 후 계산을 통해 단말의 위치를 알아내는 방식을 선택하고 있다. BLE beacon 기반 위치 추적 기술에 주로 사용되는 RSSI 값은 사람의 움직임이나 문의 열고 닫힘 등 주변 환경에 따라서 변하기 쉬우므로 직접 이용하기 어려운 문제가 있으며, BLE beacon은 BLE beacon 프로토콜 자체의 보안성 문제를 내재하고 있다. 이 때문에 BLE beacon 통신은 스푸핑(Spoofing)과 클로닝(Cloning)의 위협에 노출되어 있다.

[0039] 본 발명에서는 Beacon에 비해 더 넓은 커버리지를 가지고, 전송 속도가 빠른 Wifi를 사용함으로써 넓은 지역에서 많은 양의 패킷을 빠르게 수집하여 측정된 단말의 식별 정확도를 높인다.

[0040] 본 발명에서 제안하는 시스템은 AP에서 무선 단말의 패킷을 수집하고 Fingerprinting을 통해 디바이스의 고유 지문을 추출하여 무선 통신 단말을 식별 및 분류하여 Labeling 한 뒤 이를 기반으로 화이트리스트를 구축한다. 이를 위하여 필요한 Feature는 Sequence Number 증가율, 패킷의 길이, wlan.ht.ampduparam.mpdudensity, wlan.ht.capabilities, wlan.ht.capabilities.rxstbc, wlan.ht.capabilities.txstbc 필드이다. 시스템은 기계학습 분류 알고리즘인 Random Forest를 이용하여 단말을 학습하고 분류한 후 이를 모바일 애플리케이션에 출력하여 앱 사용자의 경로 추적을 돕는다.

[0042] 본 출원에서 제안하는 무선 신호 기반 사용자 이동 경로 추적 시스템은 크게 두 가지 시스템으로 구성된다. 첫째는 무선 단말 식별 및 분류 시스템이다. 이 시스템은 수집한 패킷으로부터 프로브 리퀘스트 프레임(Probe Request Frame)을 추출하고, 추출한 데이터로부터 단말의 고유 지문을 생성한 뒤 랜덤 포리스트(Random Forest) 알고리즘을 이용하여 단말을 식별 분류한다. 무선 단말 식별 및 분류 시스템은 데이터를 수집 및 가공하여 단말 식별 및 분류를 위한 지문을 생성하는 피처(Feature) 추출 모듈과 단말의 지문을 활용하여 단말을 식별 및 분류하는 식별 및 분류 모듈로 구성되어 있다. 둘째는 사용자 이동 경로 추적 시스템이다. 이 시스템은 모바일 환경에서 구동되며 AP의 커버리지, 식별된 단말이 감지된 위치 등의 정보를 시각화하여 표시한다. 도 1은 무선 신호 기반 사용자 이동 경로 추적 시스템의 전체 구조를 도식화한 그림이다.

[0044] 무선신호 기반 사용자 이동 경로 추적 시스템은 먼저 사용자 정의 펌웨어가 설치된 AP에서 무선 단말의 패킷을 수집하여 서버로 전송한다. 서버에서는 단말 식별 및 분류 시스템이 동작하여 수집한 무선 패킷에서 프로브 리퀘스트 프레임(Probe request frame)과 WLAN 매니저먼트 프레임(WLAN Management Frame)을 필터링하여 저장한 뒤 이를 참조하여 데이터 전처리를 진행한다. 제안하는 시스템에서 수집한 피쳐(Feature) 중 wlan.seq(Sequence number) 필드는 추가적인 가공 과정을 거쳐 사용한다. 다음으로 생성된 지문을 바탕으로 라벨링(Labeling)을 진행한다. 전처리가 완료된 이후 추출된 Feature를 식별 및 분류 모듈로 전달하고, 모듈은 이를 기반으로 단말 식별 및 분류 모델을 생성 및 학습한다. 모델이 생성된 이후 새로이 수집된 패킷이 서버로 전송되면 상기한 학습 과정과 같은 전처리 과정을 거친 후 식별 및 분류 모델을 통해 식별을 수행한 뒤 Labeling한다.

[0046] 본 발명의 피쳐(Feature) 추출 모듈은 AP로부터 단말이 송신한 패킷을 전송받으며 받은 패킷을 전처리하고 지문을 추출하는 역할을 수행한다.

[0047] 1) 패킷 수집

[0048] 본 발명의 시스템은 본래 AP에 사용자 정의 펌웨어를 설치하여 AP와 단말 간에 주고받는 패킷을 수집하여 서버로 전달받는다. 그러나 해당 펌웨어의 개발에 대한 연구는 본 출원의 연구와 동시에 진행되었기 때문에 본 출원의 실시예에서는 무선 NIC를 사용하여 패킷을 직접 수집하였다. 수집 도구로는 와이어샤크(Wireshark)를 이용하

였으며 수집 과정에서는 AP의 커버리지 내 모든 패킷을 수집하여 .pcapng 형식으로 저장하도록 하였다.

[0050] 2) 패킷 가공

[0051] 패킷 가공 모듈에서는 가장 먼저 pcapng 파일 내부에 존재하는 무선 단말들의 데이터를 단말별로 구분하여 각 단말에 해당하는 디렉토리 내부에 csv 파일 형식으로 저장한다. 해당 과정을 수행하기 위해서 우선 Tshark 도구를 이용하여 pcapng 파일 내부의 Probe Request Frame, WLAN Management Frame 데이터만을 필터링하여 .csv 파일 형식으로 저장한다. 다음으로 필터링한 Probe Request Frame, WLAN Management Frame 내에서 이후 사용할 데이터만을 추출하여 단말별로 그룹화하는 과정을 수행한다. Probe Request Frame에서 사용할 데이터 필드는 총 4가지로 다음의 표 1과 같다.

표 1

[0052]

Field	설명	사용 목적
frame.time_relative	- 첫 번째 패킷 수신 이후 패킷이 수신된 시간	- Sequnce Number 증가율 측정에 사용
wlan.seq	- 해당 Probe request에 부여된 Number	- Sequnce Number 증가율 측정에 사용
wlan.ssid	- Probe request의 목적지 AP의 SSID	- 패킷 길이 측정에 사용
frame.len	- Probe request의 전체 길이	- 패킷 길이 측정에 사용

[0053] 이 중 wlan.seq는 추가적인 가공이 필요한 Field이다. wlan.seq 는 단말이 발생시키는 Probe request에 부여된 12비트 정수를 의미하며 0부터 12비트 이진수의 최댓값인 4095 사이를 순환하므로 증가율을 측정하기 위해서는 계속 증가하는 값으로 변환할 필요가 있다. 따라서 1 순환 주기마다 4096의 배수를 곱하여 계속 증가하는 값으로 가공한다.

표 2

[0054]

Field	설명	사용 목적
wlan.ht.ampduparam.mpdudensity	- MPDU(MAC Protocol Data Unit)의 밀도값	- 무선 단말 분류 및 식별을 위해 사용
wlan.ht.capabilities	- Wi-Fi 카드의 물리적인 기능과 관련된 정보	
wlan.ht.capabilities.rxstbc	- STBC(Space Time Block Code) 수신 값	
wlan.ht.capabilities.txstbc	- STBC(Space Time Block Code) 송신 값	

[0055] 위의 표 2는 WLAN Management Frame에서 사용할 데이터들을 정리한 것이다. 이 역시 총 4가지로 이러한 필드들을 그룹화하여 단말을 구분한 뒤 devN 하위에 time_seperated 디렉토리를 생성한 후, devN.csv 파일을 10분 단위로 분할하여 devN_시간_10n 단위로 저장한다. 이는 학습에 필요한 feature를 최대한 많이 얻기 위함이다.

[0056] 도 2는 디렉토리의 계층 구조의 개념도이다. 즉, 이해를 돕기 위해 복잡한 디렉토리의 계층 구조를 한눈에 볼 수 있도록 시각화한 것이다.

[0057] dev1부터 devN까지의 디렉토리 내부의 데이터들이 모두 분할이 완료되면 Feature 추출을 시작한다.

[0059] 3) Feature 추출

[0060] AP에서 무선 단말의 패킷을 수집하고 Fingerprinting을 통해 디바이스의 고유 지문을 추출하여 무선 통신 단말을 식별 및 분류하여 Labeling 한 뒤 이를 기반으로 화이트리스트를 구축한다. 이를 위하여 생성할 Feature는 Sequence Number 증가율과 패킷의 길이이고, 직접 사용할 Feature는 wlan.ht.ampduparam.mpdudensity, wlan.ht.capabilities, wlan.ht.capabilities.rxstbc, wlan.ht.capabilities.txstbc 이다. 각 무선 단말의 필터링된 csv 파일로부터 각 Feature를 추출하는 방법은 다음과 같다.

[0062] - Sequence Number 증가율 : Sequence Number 증가율은 단위 시간당 패킷의 Sequence number 증가량을 의미한다. 선행 연구에 따르면 frame.time_relative 필드의 값을 단위 시간으로 하고 순환하지 않도록 가공한

wlan.seq field를 sequence number라고 할 때 단위 시간 대비 sequence number는 일차 직선 형태로 나타나며 단말마다 다른 값을 가진다. 따라서 그래프의 기울기를 추정하기 위해 Linear regression을 500회 진행하여 Feature 값으로 삼는다. 본 시스템에서는 1시간 분량의 패킷당 1개의 Sequence Number 증가율을 추출해 추출한 값들의 평균을 실제 Feature로 사용한다.

[0064] - 패킷의 길이 : Probe request에는 Directed와 Null probe request의 두 가지 프레임이 존재하는데, Null probe request는 무선 단말이 주위의 모든 AP를 탐지하기 위해 사용하며 Directed probe request는 특정 AP로만 Probe request를 전송하는 경우에 사용한다. Frame body 내부의 SSID field는 Null probe request의 경우 null이며 Directed probe request 경우 해당 AP의 SSID값을 갖는다. 따라서 Probe request의 frame.len 값은 Null probe request인지 Directed probe request인지에 따라 달라진다. AP의 SSID에 따라 길이가 달라지나 SSID의 종류는 무한히 많기 때문에 같은 Probe request여도 frame.len 내부의 값이 다를 수 있다. 이를 해결하기 위해 Directed probe request 패킷을 사용하지 않는 경우 기계학습에 사용할 패킷의 개수 감소로 인한 식별 정확도 감소로 이어질 수 있다. 따라서 패킷의 길이를 Feature로 사용하기 위해서는 Directed probe request의 길이를 Null probe request와 같도록 맞추는 가공 과정이 필요하다. 따라서 frame.len의 길이와 wlan.ssid의 길이의 차를 Feature 값으로 삼는다.

[0066] - wlan.ht.ampduparam.mpdudensity, wlan.ht.capabilities, wlan.ht.capabilities.rxstbc, wlan.ht.capabilities.txstbc : WLAN Management Frame의 필드로써, 가공하지 않고 필드의 값을 그대로 Feature로 사용한다.

[0068] 추출이 완료된 데이터는 앞서 그림 2에서 볼 수 있듯 devN 디렉토리 아래의 featured 디렉토리 내부에 featured.csv 형식으로 저장된다. featured.csv 파일은 이후 무선 단말 식별 및 분류 모듈로 전송되어 Labeling, 식별, 분류 과정에 사용한다.

[0070] 본 발명의 무선 단말 식별 및 분류 모듈은 기계학습 알고리즘인 랜덤 포리스트(Random Forest)를 이용하는 모듈이다. Random Forest 학습 모델이 존재하지 않는 경우 기존의 데이터를 라벨링(Labeling) 한 후 모델을 생성하고, 이후로부터는 피처U(Feature) 값을 통해 단말을 Labeling 함으로써 분류 기능을 수행한다. Feature와 Label을 통해서는 이미 학습하여 모델에 존재하는 인가 단말과 모델에 존재하지 않는 비인가 단말을 식별하는 기능을 수행한다.

[0072] 1) 단말 식별 모델 생성

[0073] 식별 및 분류 모듈은 Feature 추출 모듈로부터 생성된 devN의 featured.csv 파일들을 참조하여 (length, Sequence Number 증가율)의 묶음과 1~N까지의 정수를 대응시켜 Labeling하고 패킷의 길이와 Sequence Number 증가율을 Feature로 삼는 Random Forest 모델을 생성한다.

[0075] 2) 단말 식별 및 분류

[0076] 도 3은 무선 단말 분류 과정의 개념도이다. 먼저 단말의 Feature를 모델을 통해 검증한다. 모델을 통해 분류된 단말과 같을 예측 확률이 60퍼센트 이상일 경우 해당 단말과 같은 종류로 판단하여 같은 Label을 지정한다. 예측 확률이 60퍼센트 미만인 경우 이전에 학습한 적 없는 단말로 간주하여 새로운 Label을 할당한다.

[0077] 도 4는 무선 단말 식별 과정의 개념도이다. 먼저 단말의 Feature와 Label을 모델을 통해 검증한다. 이때 모델은 입력 받은 Feature를 통해 자신이 학습한 단말 중 가장 유사하다고 판단되는 단말의 Label과 예측 확률을 제시한다. 모델이 제시한 Label과 분류 과정에서 할당된 Label이 다르다면 비인가 단말로 식별한다. 반대로 모델이 제시한 Label과 분류 과정에서 할당된 Label이 같고, 예측 확률이 60% 이상이면 인가된 단말로 식별한다. 식별 및 분류된 결과는 서버 데이터베이스에 저장하여 이동 경로 추적 모바일 애플리케이션에서 사용할 수 있도록 한다.

[0078] 본 말명의 이동 경로 추적 모바일 애플리케이션은 무선 단말 식별 및 분류 시스템에서 데이터베이스에 저장한 식별 및 분류 결과와 이들이 감지된 AP, AP의 커버리지를 시각적으로 표시하는 애플리케이션이다. 애플리케이션에는 기본적으로 지도가 출력되며 단말과 AP 두 가지 모드가 존재한다. 도 5는 단말 모드의 작동방식을 예시한 도면이다.

[0079] 단말 모드에서는 목록보기 버튼을 터치하여 감지된 단말들을 시간순으로 나열한 목록으로 진입할 수 있으며 목록에서 특정한 단말을 터치한 경우 해당 단말이 감지된 AP의 정보, AP의 위치, 단말의 정보와 감지된 시간이 지도상에 출력된다.

[0080] 도 6은 AP 모드의 작동방식을 예시한 도면이다. AP 모드에서는 지도상에 AP와 커버리지가 출력되며 AP를 터치할 경우 해당 AP에서 감지된 단말들을 시간순으로 나열한 목록으로 진입할 수 있다. AP모드 역시 목록에서 특정한 단말을 터치한 경우 해당 단말이 감지된 AP의 정보, AP의 위치, 단말의 정보가 지도상에 출력된다.

[0082] 본 출원에서 제안한 무선 단말 식별 및 분류 시스템의 정확도를 검증하기 위해 진행한 실험과 그 결과를 설명하면 다음과 같다.

[0084] 도 7은 무선 단말 식별 및 분류 시스템의 정확도 검증시 구성한 실험 환경의 개념도이다. 실험에는 AP가 3대, 단말은 5대를 사용하였으며 일반적인 인가 단말 4대와 MAC 주소를 위조한 비인가 단말 1대로 구성하였다. AP에서 패킷을 전달받는 역할을 대체하기 위하여 모니터 모드가 지원되는 무선 NIC를 사용하여 무선 신호를 Wireshark tool로 수집하였으며 시스템이 동작할 PC를 1대 사용하였다.

[0085] 실험 시나리오는 다음과 같다. 먼저 2대의 AP(AP1, AP2)와 5대의 무선 단말(단말 B ~ 단말 F, 단말 E는 단말 B의 MAC 주소로 위장한 비인가 단말)이 존재하는 환경에서 단말 B, C, D, F는 AP1에 연결하고, 단말 E는 AP2에 연결하여 24시간 동안 무선 패킷을 수집한다. 수집된 패킷 파일에서 본 연구에서 사용하는 Probe Request Frame과 WLAN Management Frame을 필터링하여 .csv 파일로 따로 저장한다. 이후 필터링한 데이터의 그룹화를 진행하여 무선 단말들의 데이터를 분류하고 디렉토리별로 저장한 뒤 데이터를 10분 단위로 분할한다.

[0087] 다음으로 Labeling을 진행하는데, 모델을 생성하기 전에는 디렉토리별로 구분된 단말들을 차례대로 Labeling 한다. 다음으로 수집한 무선 패킷에서 Feature를 추출하여 단말 식별 모듈에 쓰일 학습용 데이터를 생성한 뒤 이를 학습하여 단말 식별 및 분류 모델을 생성한다. 모델 생성이 완료되면 학습용 데이터 생성 시와 같은 환경에서 단말 C, D, F는 AP1에 연결하고 단말 B를 AP1에 연결하고 단말 E는 전원을 끈 상태에서 8시간 동안 패킷을 수집한다. 이후 단말 C, D, F의 전원 및 연결 상태를 유지하며 단말 B의 전원을 끄고 단말 E의 전원을 켜 AP1에 연결하고 무선 패킷을 수집한다. 마지막으로 B, C, D, F의 전원 및 연결 상태를 유지하며 단말 E를 AP2에 연결하여 테스트용 데이터를 수집한다. 마지막으로 수집된 테스트 데이터를 시스템에 입력하여 식별 및 분류 정확도를 측정한다.

[0089] 학습 및 테스트 데이터 생성을 위한 무선 패킷의 수집은 Wireshark 를 실행하여 수행한다. 수집된 패킷을 res/pcapng 디렉토리 아래의 data.pcapng 파일에 저장하고 Probe request와 WLAN Management Frame만을 필터링하여 res/pcapng_csv/learn 아래에 probe.csv 파일로 저장한다.

[0090] 생성된 probe.csv 파일을 단말별로 적절하게 나누어 저장하기 위해서는 wlan.ht.ampduparam.mpdudensity, wlan.ht.capabilities, wlan.ht.capabilities.rxstbc, wlan.ht.capabilities.txstbc 필드와 length 필드들을 이용하여 그룹화를 수행한다.

[0091] 각 디렉토리 내부에는 devN.csv 파일이 생성되며, devN.csv 파일 내부에는 각 단말만의 데이터가 저장된다.

[0092] 단말별로 나누어 저장하는 과정이 끝나면 도 8과 같이 devN.csv 파일을 10분 간격으로 나누어 res/packet/probe/devN/time_seprated 아래에 저장한다.

[0093] 도 8은 devN.csv 파일을 저장한 예이다.

[0095] 마지막으로 도 0처럼 데이터 가공 과정을 거쳐 res/packet/probe/devN/featured 아래에 featured.csv 파일로 저장한다. 도 9는 featured.csv 파일로 저장한 예이다.

[0096] dev1 ~ devN까지의 모든 디렉토리에 대하여 fearuted.csv 파일이 생성되면 학습 데이터 가공이 종료된다.

[0098] 무선 단말 식별 및 분류를 위해 학습용 데이터로 Random forest 모델을 생성 및 학습시키고 테스트 데이터를 통해 단말 분류 결과를 확인하였다. 아래의 표 3과 표 4는 각각 학습 데이터, 테스트 데이터에 대한 분류 실험 결과를 나타내고 있다.

표 3

[0099]

dev1	dev2	dev3	dev4
단말 D	단말 C	단말 B	단말 E (B로 위장한 단말)
		단말 F	

[0100] 학습용 데이터 분류 결과 단말 D, C, E를 정확하게 분류해내었으며 이에 더하여 Spoofing 단말인 단말 E와 기존

단말인 단말 B는 완벽히 분류를 수행하였다. 그러나 단말 B와 단말 F는 Feature값이 매우 유사하여 같은 단말인 dev 3으로 판단되어 분류하지 못하였다.

표 4

[0101]

dev1	dev2	dev5	dev4
단말 D	단말 C	단말 B	단말 E (B로 위장한 단말)
		단말 F	

[0102] 테스트 데이터 분류 결과 역시 학습용 데이터와 유사하게 단말 D, C, E를 정확하게 분류해내었으며 이에 더하여 Spoofing 단말인 단말 E와 기존 단말인 단말 B는 완벽히 분류를 수행하였다. 그러나 테스트 데이터의 단말 B, F는 같은 단말로 분류되었으나 레이블 할당과정에서 예측 확률값이 60% 미만으로 dev 3이 아닌 dev 5로 새로운 레이블이 할당되었다. 다음으로 생성된 모델로부터 식별 테스트를 진행하였다.

[0103] 다음의 표 5는 단말 식별 테스트를 위하여 5대의 단말로부터 24시간 동안 수집한 데이터로부터 생성된 총 418개의 테스트 데이터를 단말별로 분류한 것이다.

표 5

[0104]

단말	패킷 개수	수집시간
B	144	24시간
C	48	
D	33	
E	49	
F	144	

[0106] 다음의 표 6은 전체 무선 단말에 대한 식별 테스트 결과를 정리한 것이다. dev 1, 2, 4는 모두 정확하게 예측하였음을 볼 수 있다. 이는 TP(True Positive)로써 실제 참인 정답을 참이라고 예측한 경우이다. 그리고 dev4와 dev5에 대하여 dev5로 라벨링 된 단말들과 dev4로 라벨링 된 단말은 다른 단말이지만, 식별 모델은 dev5의 데이터들을 dev4로 예측하였다. 하지만 알고리즘 절차상 라벨링된 dev 번호가 같지 않고 예측 확률이 60% 미만으로 관찰되어 unathourized 단말로 식별하였다. 이 같은 경우 TN(True Negative)로써 거짓인 정답을 거짓이라고 잘 예측한 경우이다. 그러나 dev6은 레이블 할당과정에서 데이터 1개가 예측 확률이 60% 미만이여서 새로이 Label을 할당한 것으로 모델이 잘못 식별했음을 알 수 있다. 이는 FN(False Negative)인 경우로 실제 참인 정답을 거짓이라고 잘못 예측한 결과이다.

표 6

[0107]

	TP(True Positive)	식별 모델		
	TN(True Negative)	dev1	dev2	dev4
	FN(False Negative)			
예측 결과	dev1	33	-	-
	dev2	-	48	-
	dev4	-	-	48
	dev5	-	-	144
	dev6	-	-	1

표 7

[0109]

전체 데이터 수	TP(인가 식별 정답)	TN(비인가 식별 정답)	정확도
274	129	144	0.99

[0110] 위의 표 7은 전체 무선 단말에 대한 식별 테스트의 정확도를 정리한 것이다. 정확도를 "(인가 식별 정답 개수 +

비인가 식별 정답 개수) / 전체 데이터 수"로 정의하였을 때, 시스템의 식별 정확도는 (129+144) / 274 = 0.99, 99%임을 확인할 수 있다.

[0112] 본 출원에서는 좀 더 넓은 범위에서 사용자의 위치를 추적할 수 있도록 무선신호 기반 사용자 이동 경로 추적 기술을 제안하였다. 제안한 시스템은 무선 단말 분류 및 식별 서브 시스템과 이동경로 추적 모바일 애플리케이션으로 나뉘어 있으며 서브시스템은 Feature 추출 모듈과 Random Forest 알고리즘을 이용한 식별 및 분류 모듈로 구성되어 있다.

[0113] 그리고 제안된 시스템이 정상적으로 동작하며 여러 개의 AP에서 같은 무선 단말끼리 정확히 분류 및 식별할 수 있는지 확인하기 위해 실험을 진행하였다. 2대의 AP와 다수의 무선 단말과 MAC 주소를 위장한 비인가 무선 단말이 같이 존재하는 환경에서 학습 데이터를 수집하여 Random Forest 모델을 생성 및 학습하고, 기존의 무선 단말과 MAC주소를 위조한 무선 단말이 다른 AP에 접속하는 과정을 거친 테스트 데이터를 수집하여 결과를 확인하였다. 실험 결과 무선 단말 식별 정확도는 99%, 단말 분류 결과는 60%를 달성하였다. 이는 선행 연구인 무선 단말 식별 연구와 비교하였을 때 학습 과정에서 비인가 무선 단말을 학습하더라도 이를 식별해낼 수 있는 장점이 있으며 다수의 AP에서 감지된 같은 단말을 분류해낼 수 있다는 의의가 있다. 또한, 학습 전처리 과정에서 MAC 주소를 이용한 전처리 대신 무선 단말들의 Feature 데이터를 그룹화를 이용하였기에 보안상의 이점이 존재한다.

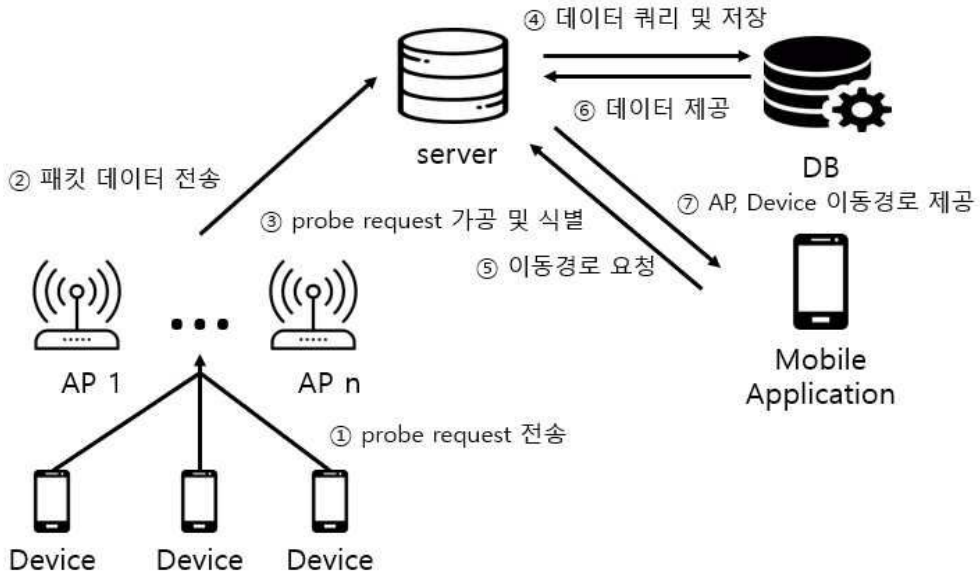
[0114] 본 출원에서 제안하는 시스템에서는 식별 및 분류의 정확성을 위해 일정한 양 이상의 데이터가 모이지 않으면 기계학습시 Feature로 사용하지 않는다. 이는 패킷을 빨리 수집할수록 식별 및 분류가 완료되기까지의 시간이 짧아짐을 의미하므로 이를 해결하기 위해 더 많은 양의 패킷 데이터를 수집할 필요가 있다. 실험 결과 무선 단말이 송신하는 Probe request의 주기는 단말마다 고유한 일정 주기를 가지므로 이를 해결하기 위해서는 더 많은 AP를 설치하여 수집되는 DIrected probe request의 양을 늘리는 방향으로 연구를 진행할 수 있다.

[0116] AP만으로 무선단말에서 보내는 정보를 수집할 수 없지만, 본 발명에서는 AP내부에 커스텀 펌웨어를 설치하여 이를 해결하였다. 또는 무선 단말이 AP를 인식하기 위한 와이파이 스캐닝 방법 중 능동 스캐닝에 속하는 프로브 리퀘스트(Probe Request)신호를 수집하기 때문에, 와이파이 기능만 켜져 있다면 무선단말이 AP와 연결되어 있지 않더라도 AP가 해당 무선단말을 파악할 수 있는 장점이 있다.

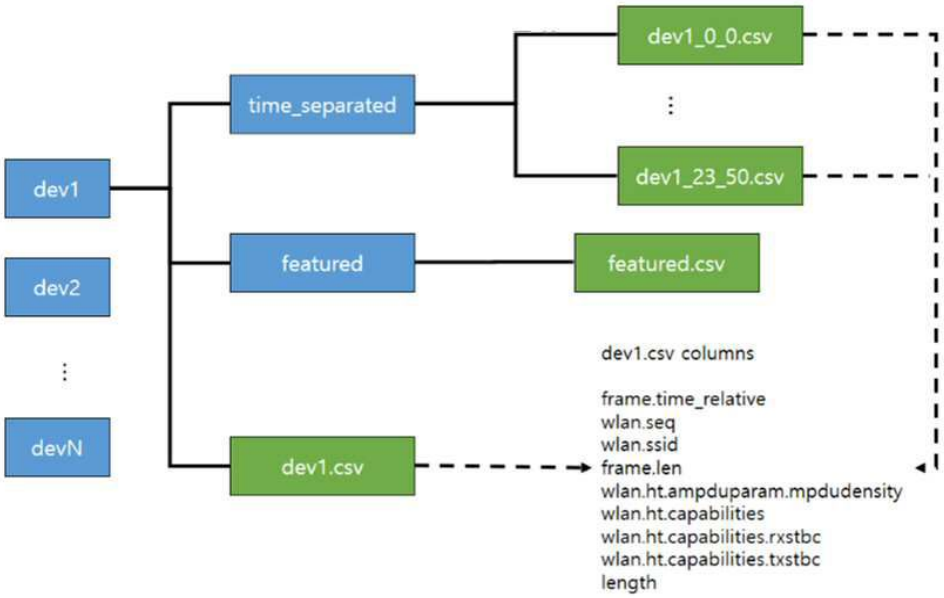
[0118] WiFi, BLE beacon 등의 무선신호를 기반으로 한 기존의 실내 위치 추적 시스템은 좁은 공간에서 사용자의 위치를 실시간으로 추적하는 기술이다. 따라서 기존의 위치 추적 시스템과 본 시스템을 결합한다면 AP의 커버리지 내에 들어온 단말의 위치를 더욱 정확하게 추적해낼 수 있다. 나아가서 802.11 프레임 외에 BLE 등의 무선 패킷으로부터 추가적인 지문 추출을 통해 분류 정확도를 높일 수 있다면 넓은 범위와 좁은 범위를 아우르는 무선 단말 위치 추적 시스템으로 활용할 수 있을 것으로 예상된다.

도면

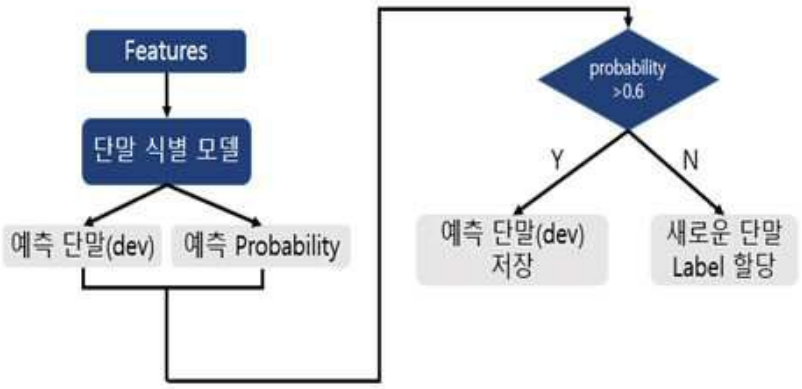
도면1



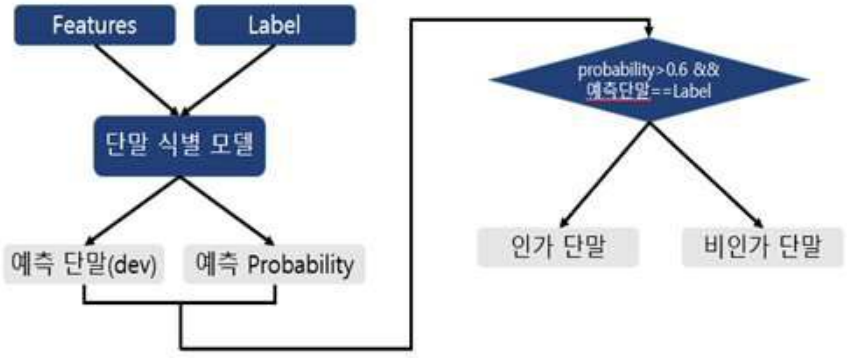
도면2


dev1_0_0.csv
time_separated
dev1_23_50.csv
dev1
featured
featured.csv
dev2
dev1.csv columns
frame.time_relative
wlan.seq
wlan.ssid
devN
dev1.csv
frame.len
wlan.ht.ampduparam.mpdudensity
wlan.ht.capabilities
wlan.ht.capabilities.rxstbc
wlan.ht.capabilities.txstbc
length

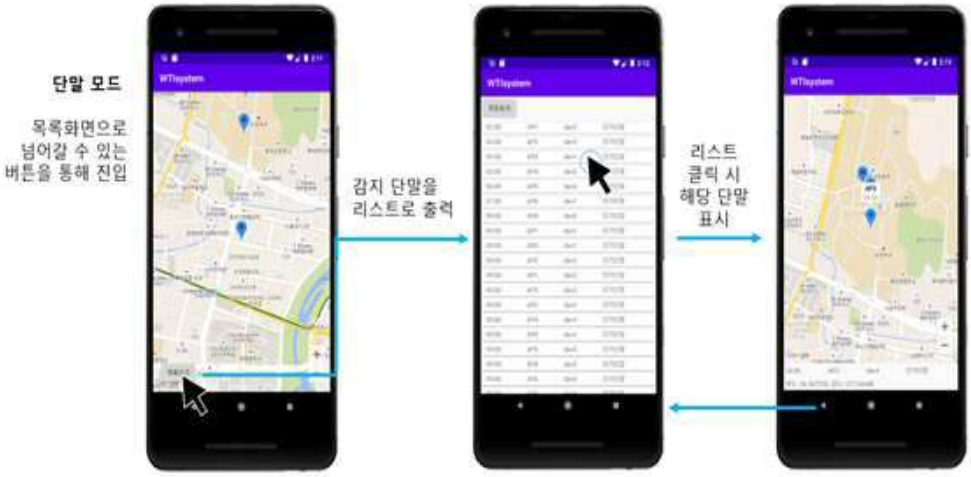
도면3


Features
단말 식별 모델
예측 단말(dev)
예측 Probability
probability
>0.6
Y
N
예측 단말(dev)
저장
새로운 단말
Label 할당

도면4


Features
Label
단말 식별 모델
예측 단말(dev)
예측 Probability
probability>0.6 &&
예측단말==Label
인가 단말
비인가 단말

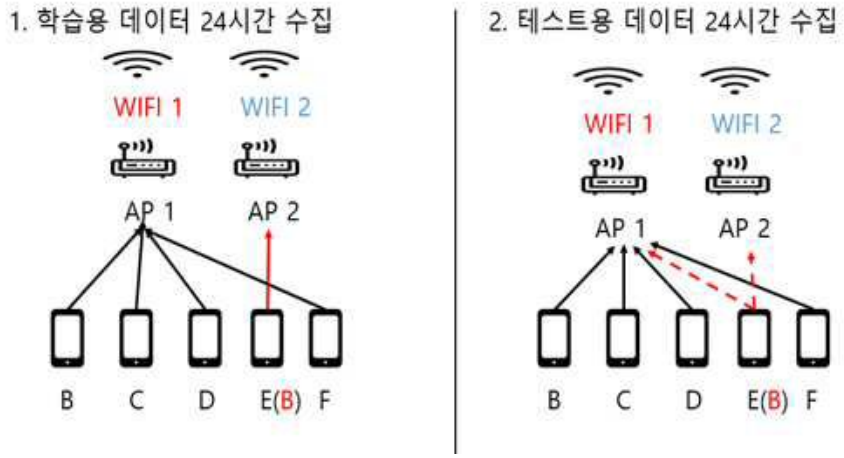
도면5


단말 모드
목록화면으로 넘어갈 수 있는 버튼을 통해 진입
감지 단말을 리스트로 출력
리스트 클릭 시 해당 단말 표시

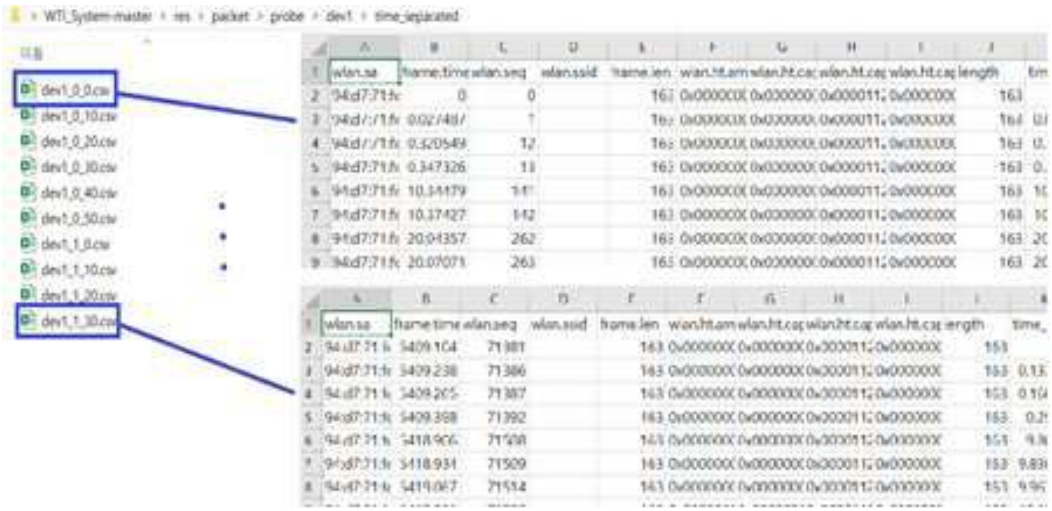
도면6


AP모드
지도에 출력된 AP를 클릭해서 진입
해당AP에서 감지된 단말들
감지시간에 따라 분류된 확장형 리스트
리스트 클릭 시 해당 단말의 이동경로를 지도에 표시

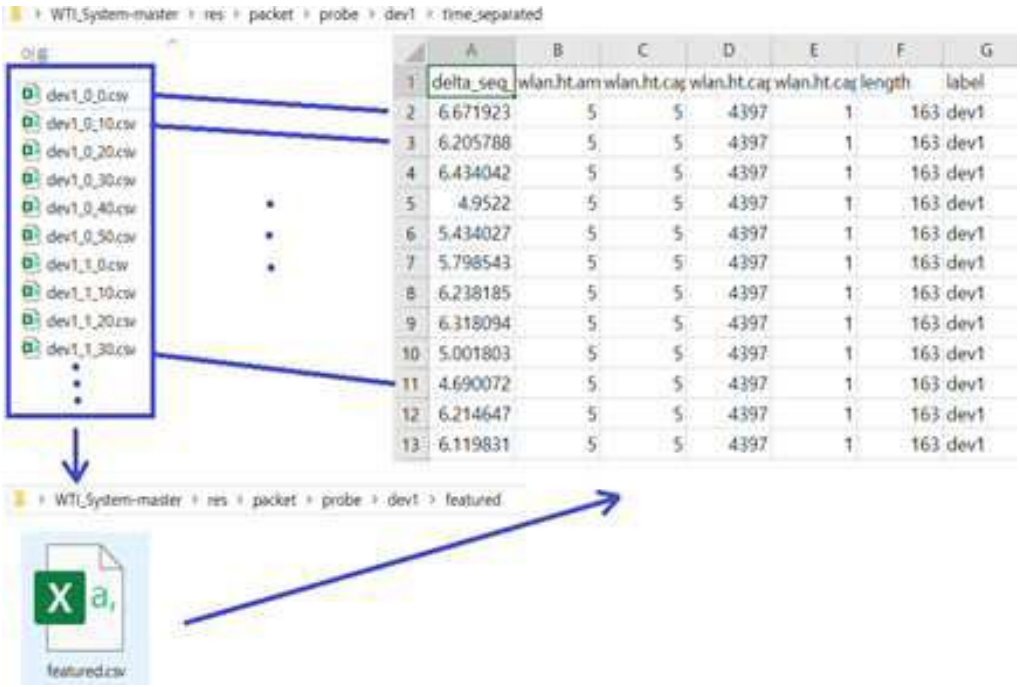
도면7


1. 학습용 데이터 24시간 수집
WIFI 1
WIFI 2
AP 1
AP 2
B
C
D
E(B)
F
2. 테스트용 데이터 24시간 수집
WIFI 1
WIFI 2
AP 1
AP 2
B
C
D
E(B)
F

도면8


WTI_System-master › res › packet › probe › dev1 › time_separated
dev1_0_0.csv
dev1_0_10.csv
dev1_0_20.csv
dev1_0_30.csv
dev1_0_40.csv
dev1_0_50.csv
dev1_1_0.csv
dev1_1_10.csv
dev1_1_20.csv
dev1_1_30.csv

도면9


WTI_System-master › res › packet › probe › dev1 › time_separated
dev1_0_0.csv
dev1_0_10.csv
dev1_0_20.csv
dev1_0_30.csv
dev1_0_40.csv
dev1_0_50.csv
dev1_1_0.csv
dev1_1_10.csv
dev1_1_20.csv
dev1_1_30.csv
delta_seq wlan.ht.am wlan.ht.ca wlan.ht.ca wlan.ht.ca length label
6.671923 5 5 4397 1 163 dev1
6.205788 5 5 4397 1 163 dev1
6.434042 5 5 4397 1 163 dev1
4.9522 5 5 4397 1 163 dev1
5.434027 5 5 4397 1 163 dev1
5.798543 5 5 4397 1 163 dev1
6.238185 5 5 4397 1 163 dev1
6.318094 5 5 4397 1 163 dev1
5.001803 5 5 4397 1 163 dev1
4.690072 5 5 4397 1 163 dev1
6.214647 5 5 4397 1 163 dev1
6.119831 5 5 4397 1 163 dev1
WTI_System-master › res › packet › probe › dev1 › featured
featured.csv