Why abandon Randomized MAC-Address : Wi-Fi Probe Request 기반 유동인구 분석 방법

셔키르현 오포호노브 · 이재현 · 문준영*

㈜제로웹

Why abandon Randomized MAC-Address : An Analysis of Wi-Fi Probe Request for Crowd Counting

Shokirkhon Oppokhonov · Jae-Hyun Lee · Jun-young Moon*

ZEROWEB

E-mail : chodak1290@gmail.com / jack2002jack@zeroweb.kr / suho10@korea.ac.kr

요 약

실시간으로 유동인구를 계측하는 기술은 다양한 산업 분야에서 공간 밀집도에 대한 통찰력을 제공하여 더 좋은 서비스 환경을 만들어준다. 이에 따라 여러 기업과 학계에서는 특정 공간의 유동인구 데이터를 계측하기 위해 오랫동안 다양한 연구를 시도해왔으며, 최근에는 스마트시티와 디지털트윈의 일환으로 Wi-Fi 신호를 활용한 유동인구 분석 사업화가 더욱 활발한 추세이다. 본 논문에서는 유동인구 데이터를 활용하는 일반 수요자(비전문가)의 관점에서 유동인구 계측 시스템에 대해 쉽게 이해할 수 있도록 설명한다. 구체적으로는 사람들이 소지한 스마트폰으로부터 수집되는 MAC-address 값 기반의 유동인구 추정 방법을 제시하는데, 추정값의 정확도를 분석하기 위해 Real MAC-address와 Random MAC-address 값을 구분한 뒤, Real MAC-address가 추출된 실제 스마트폰 기기 수와 CCTV 화면에 집계된 사람 수를 비교하는 실험을 진행한다. 그 결과 두 데이터 간의 유사한 상관 계수가 나타났다. 이러한 결과에 근거하여 MAC-address 분류를 통한 Wi-Fi Probe Request 기반 유동인구 분석 방법을 제시한다.

ABSTRACT

Estimation of the presence of people in real time is extremely useful for businesses in providing better services. Many companies and researchers have attempted various researches in order to count the number of floating population in specific space. Recently, as part of smart cities and digital twins, commercialization of measuring floating populations using Wi-Fi signals has become active in the public and private sectors. This paper explains the floating population measuring system from the perspective of general consumers(non-experts) who uses current population data. Specifically, it presents a method of estimating the floating population based on MAC-address values collected from smartphones. By distinguishing Real MAC-address and Random MAC-address values, we compare the estimated number of smartphone devices and the actual number of people caught on CCTV screens to evaluate the accuracy of the proposed method. And it appeared to have a similar correlation between the two datas. As a result, we present a method of estimating the floating population based on analyzing Wi-Fi Probe Requests

키워드

Wi-Fi Probe Request, Real MAC, Random MAC, People Counting, Crowd Counting

Ⅰ. 서 론

과거의 유동인구 계측은 단순히 보안상의 목적으로 이루어졌다. [1], [2] 그러나 최근에는 스마트시티, 빅데이터, 디지털트윈에 대한 도입과 관심의

* corresponding author

증가함에 따라 공공과 민간의 관리자나 주최측에서 유동인구에 대한 통계를 지속적으로 확인하고자 하는 수요가 증가하고 있다.

유동인구 데이터는 공간에 대한 점유율 탐지를 통해 지역의 주거/상권변화와 이벤트 성과분석 및 마케팅 연구뿐만 아니라, 사람 수에 비례하여 효율적으로 난방, 냉각, 조명, 차양 및 환기 시스템을 컨트롤하는 등 최적화된 에너지 환경 시스템을 구현할 수 있고 재난 상황 발생 시 신속한 대처를 할 수 있도록 도움을 준다. [3], [4], [5]

본 연구에서는 상기 예시와 같이 다양한 산업군에서 활용될 수 있는 유동인구 데이터를 추출하기 위해 스마트폰의 MAC-address(이하 MAC)를 활용한 유동인구 계측 방법을 제안한다.

일반적으로 Wi-Fi 통신 모듈이 장착된 스마트폰과 같은 휴대용 전자장치는 Wi-Fi Probe Request 동작으로 주변에 존재하는 AP(Access Point) 정보를 찾는다. 휴대용 전자장치(스마트폰)와 AP(공유기)는 상호 네트워크 연결을 위해 각자의 MAC(기기 고유주소) 값과 RSSI(신호세기) 등의 데이터를 송출한다. 일련의 과정에서 두 장치가 직접 연결되지 않더라도 암호인증 과정 없이 상호 기기에 대한 정보를 수집한다.

2010년대 이후로 많은 사람들이 Wi-Fi 인터페이스가 탑재되어있는 스마트폰을 휴대하고 있는 것에 착안하여, Wi-Fi Probe Request를 활용한 스마트폰 MAC 계측으로 유동인구를 분석하고 있다. 그러나 Android 8 및 iOS 8 이상의 운영체제를 갖는 모바일 기기에서는 사용자 개인정보 보호의 일환으로 기본적으로 Randomized MAC-address (MAC 주소 무작위화)가 설정되어 있기 때문에 유동인구 계측의 정확성을 위해서는 별도의 Randomized MAC(이하 Rand-MAC) 처리 알고리즘이 요구된다.

따라서 Wi-Fi 신호를 활용하여 유동인구를 분석하는 시스템은 정확도 고도화를 위한 지속적인 알고리즘 연구가 진행되어야 한다. 그러나 계속해서 변화하는 개인정보 정책에 맞추어 알고리즘을 수정하는 것은 쉽지 않기 때문에, 기존에 설치되어있는 Wi-Fi 기반 유동인구 시스템은 시간이 지날수록 정확도가 낮아지는 실정이다. 더욱이 공공이나 민간에서의 일반 수요자들은 유동인구 분석을 전문기관에 전적으로 의뢰하고, 화면상에 구현되는 데이터 정보만을 활용하기 때문에 유동인구 분석 전문기관에서 앞서 서술한 내용 등의 기술적인 전문지식을 제공하지 않으면, 일반 사용자들은 어떠한 이유로 유동인구 데이터가 불안정해지는지 알 수 없다. 따라서 본 논문에서는 전문지식이 없는 일반 사용자의 눈높이에 맞추어 유동인구 계측 시스템에 대해 설명하고, 실험을 통한 결과물로서 MAC 주소를 구분하여 Wi-Fi 신호를 기반으로 유동인구를 계측하는 방법을 제시한다.

본 논문은 다음과 같이 구성된다. Ⅱ장에서는 Wi-Fi 기반 유동인구 계측 관련 연구를 검토한다. Ⅲ장에서는 전통시장에서의 현장 테스트 시나리오를 베이스로 Wi-Fi 스캐닝 시스템의 구성과 MAC 주소 분류 방식을 설명한다. Ⅳ장에서는 실험 결과를 정리한다. 그리고 Ⅴ장에서 연구 결과를 요약하면서 논문을 마무리 한다.

Ⅱ. 관련 연구

2.1. Wi-Fi 기반 유동인구 분석 기술

Wi-Fi 기반의 유동인구 계측과 관련한 연구는 각각의 고유 특성을 지닌 세 가지의 대표적 기술이 존재한다.

가. 주변의 Wi-Fi 무선 신호세기를 직접 스캔하여 사전에 Mapping된 POI(Point of intersting) 정보로 실내에서의 위치를 감지하고 움직임을 계측하는 Wi-Fi Fingerprint 기술.

나. 인체의 움직임이 특정 무선 신호의 채널 환경 변화에 쉽게 영향을 주는 것에 착안한 기술로써 Wi-Fi 인프라가 구축되어있는 곳에 Wi-Fi 모듈이 탑재된 전자기기를 소지하지 않더라도 전파 흐름의 변화에 따라 유동인구를 감지하고 추정하는 Electronic frog eye 기술[6].

다. 스마트폰이 Wi-Fi를 사용하기 위해 주변의 AP를 탐색하는 동안 인식되는 MAC 주소 정보로 유동인구를 추정하는 Wi-Fi Probe Request 기술.

본 연구에서는 세 번째로 열거한 Wi-Fi Probe Request 기술을 통해 실험을 진행한다.

2.2. 유동인구 분석 기술 특징

유동인구 계측에는 앞서 직접 서술한 대표적인 세 가지 Wi-Fi 기반 계측기법 이외에도 수동 계측기법, 영상 기반 계측기법 [7], 적외선 센서 기반 계측기법 [8], [9] 및 Wi-Fi Traffic 기반 계측기법 [10] 등 공간적 특성과 활용목적에 따라 다양한 유동인구 계측기술이 존재한다. 그러나 각각의 기술마다 장단점이 명확하다.

일례로 수동 계측기법은 가장 높은 정확도를 갖지만 정의한 이름 그대로 사람이 직접 계측해야 하기에 전담 인력이 필요하다. 24시간 동안 계측공간의 입구에서 서서 한 명씩 들어오고 나가는 사람들을 계측하는 것은 막대한 유지 비용이 들어가는 큰 단점이 있다.

적외선 센서 기반 계측기법은 비교적 설치가 쉽고 센서 모듈이 저렴하다는 장점이 있다. 그러나 태양광 등의 외부 노이즈를 차단하기 위해 송신부와 수신부로 이루어지는 한 쌍의 센서를 공간 입구에 설치해야 하는 구조적 한계가 있다. 이로 인해 문 앞에 한 사람이 오래 머무르거나 장애물이 있을 때, 혹은 여러 사람이 동시에 문을 지나갈 때 다수의 움직임을 하나의 움직임으로 계측하는 큰

단점으로 연결된다.

지능형 CCTV 등의 영상 기반 계측기법은 적외선 센서와 달리 동시다발적으로 진입하는 사람들에 대한 계측에서는 자유롭지만, 어두운 공간에서의 객체 인식, 벽으로 가려지는 공간의 계측이 어렵다는 한계가 있으며, 다른 계측기법에 비교하여 고화질의 카메라 및 이미지 프로세싱에 필요한 모듈 가격 등 구축 비용이 높다는 단점이 있다.

본 연구에서 활용하는 Wi-Fi Probe Request 기법은 앞서 열거한 기술과 비교했을 때 작은 벽이나 장애물에 대한 노이즈 영향이 적으며, 높은 곳에 설치하지 않아도 넓은 공간을 측정할 수 있고 상대적으로 구축 비용이 저렴하다는 점에서 큰 이점을 지닌다. 그러나 이 기술 역시 앞서 서론에 명시한 바와 같이, 개인정보 보호 강화에 따른 Rand-MAC의 발생과 계측 공간에 위치하는 인원이 스마트폰에서 Wi-Fi 모듈을 사용하지 않는 등의 변수로 인하여 정확도가 떨어지는 단점이 존재한다.

III. 실험환경 구성

3.1. 현장 테스트 설계

Wi-Fi probe request 기반으로 계측된 값과 실제 존재하는 유동인구 데이터 간의 정확도를 측정하기 위해 부산광역시 해운대구에 위치한 전통시장인 반송큰시장을 테스트 장소로 삼았다.

반송큰시장은 시장의 주요출입로마다 CCTV가 설치되어있으며, 해운대구청과 ㈜제로웹의 협업으로 2019년부터 반경 내의 스마트폰을 모니터링 할 수 있는 Wi-Fi 스캐너가 서로 겹치지 않게 최소 30m 간격을 두고 구축되어있다. 이는 하나의 스마트폰이 두 개 이상의 Wi-Fi 스캐너에서 카운팅 되지않는 환경이 구성됨으로써 Rand-MAC이 야기하는 중복 데이터를 사전에 차단한다. 또한 전통시장의 통신 불안정성을 대비하여 4개의 메인 스캐너 외에 3개의 예비 스캐너를 추가 배치하였다. 본 연구에서 활용한 반송큰시장의 CCTV와 Wi-Fi 스캐너 구축 Layout은 그림 1과 같다.

그림 1. 반송큰시장 CCTV, Wi-Fi 스캐너 Layout

3.2. Wi-Fi 스캐너 적합성 확인

테스트에 활용되는 Wi-Fi 스캐너는 하드웨어적 계측 오차를 줄이고자 3가지 종류의 스캐너를 준비하여 같은 장소에서 사전 비교 테스트를 진행하였다.

그림 2. 실험 전 비교를 위한 Wi-Fi 스캐너 3종

그림 2의 우측 스캐너는 2016년도에 개발된 Wi-Fi 신호와 블루투스 신호를 통해 인근 스마트폰의 MAC 주소를 카운팅하는 센서인 REALSTEP SPOT이다.

그림 2의 중간 스캐너는 Raspberry_pi_3 보드에 ESP-32 통신 모듈과 미세먼지 환경센서를 추가 탑재하여 REALSTEP SPOT과 동일한 방식으로 MAC 주소를 카운트하면서 설치장소 주변의 미세먼지 농도를 체크하는 테스트용 프로토타입 REALSTP_DUST이다.

그림 2의 좌측 스캐너는 Raspberry_pi zero W에 탑재되어있는 Wi-Fi 모듈만을 활용하여 AP 모니터링으로 수집되는 MAC 주소값을 비교하기 위해 제작한 테스트용 센서이다.

세 종류의 스캐너는 서로 다른 Wi-Fi 통신 모듈이 탑재되어있지만, Wi-Fi Probe Request의 동일한 원리를 이용한다.

우선 주변 스마트폰의 MAC주소를 30초 동안 스캔한 뒤 서버로 보내준다. 서버에서는 이 데이터를 5분 동안 정제하는 과정을 통해 중복되는 Real-MAC address (이하 Real-MAC) 값과 13개의 Wi-Fi 채널에서 탐지하지 못한 MAC 주소값을 보상한다.

이렇게 일정 시간 동안 데이터를 누적하고 정제하는 과정 덕분에 스캐너별로 수집된 MAC 주소의 데이터값에 대한 차이는 없었다.

그러나 외장 안테나를 사용하는 REALSTEP SPOT과 내장 안테나를 사용하는 ZERO W기반의 스캐너와 ESP32 기반의 스캐너간에 계측범위에서의 차이가 발생하였다. 계측범위의 넓이에 따라 탐지되는 스마트폰의 수가 늘어나고, 늘어난 수량만큼 Rand-MAC의 수치가 함께 증가하여 Wi-Fi 스캐너간의 계측 오차가 발생하였다.

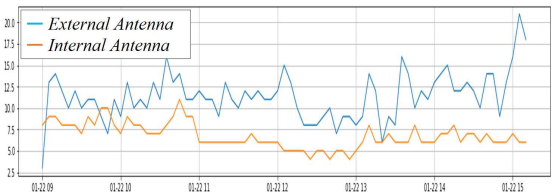

그림 3. 내장/외장 안테나의 데이터 수집량 비교

그림 3의 그래프와 같이 주황색 선(내장 안테나)과 하늘색 선(외장 안테나)의 데이터 수집량이 약 2~2.5배 정도 차이 났다.

해운대구 반송큰시장에 설치되어있는 모델은 REALSTEP SPOT이므로 사전 테스트에서 진행한 ZERO W 기반의 스캐너와 ESP32 기반의 스캐너보다 더 넓은 범위에서 유동인구를 탐지할 수 있다. 무엇보다 Wi-Fi 스캐너의 정확도 오류에 큰 비중을 차지하는 변수 값인 Rand-MAC을 더 넓은 범위에서 보다 많이 추출하기 때문에, 실제 유동인구 데이터와의 비교를 통한 알고리즘의 성능을 효과적으로 테스트할 수 있다.

3.3. Wi-Fi Probe Request 데이터 수집 절차

Wi-Fi 스캐너에서 스마트폰의 MAC 주소(Raw Data)를 수집한 뒤, 유동인구 데이터(Processed Data)로 변환하는 과정은 그림 4와 같다.

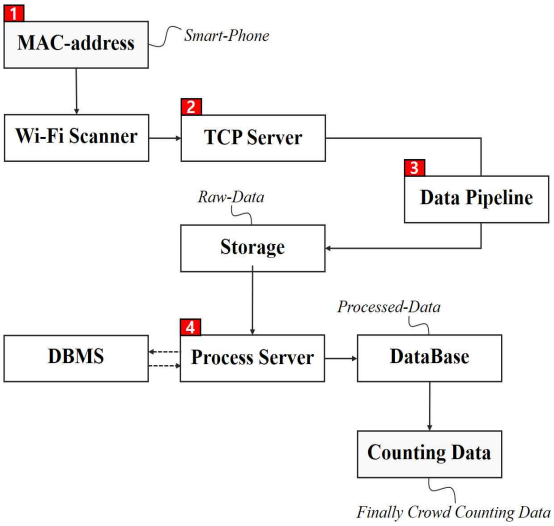

그림 4. Wi-Fi MAC 주소 수집 및 가공 프로세스

첫째, Wi-Fi 스캐너에서 수집된 주변 스마트폰 MAC 주소는 각 스캐너의 Local Memory에 저장되어 있다가 30초 단위로 서버에 Flush한다. 이때 전송하는 프로토콜은 Binary format으로 TCP Server에 저장된다.

둘째, Python Twisted로 만들어진 서버 프로그램으로 위 Binary format을 파싱하고 Data Pipeline에 tsv format으로 전송한다.

셋째, Data Pipeline에서는 여러 개의 daemon으로 돌아가는 TCP Server로부터 tsv format을 네트워크를 통해 입력받아 1MB가 되면 Storage로 flush하는 기능을 수행한다. 즉, Wi-Fi 스캐너의 설치 수량이 증가하여도 Data Pipeline을 통해 여러 곳에서 각각 수신하는 것이 아니라 하나의 서버로부터 데이터를 받은 것처럼 Storage에 하나의 파일을 생성하여 운영상 성능을 높인다.

넷째, Storage에 파일이 생성되면 자동으로 파일 생성 event를 DBMS에 전송하며 함수가 자동으로 실행된다. tsv 형태로 되어있는 파일을 읽어 최근 1시간 동안의 유동인구, 시간별 유동인구 등의 결과물로 가공되어 DataBase에 저장되며, 최종적으로 가공된 유동인구 데이터는 SQL문을 통해 확인한다.

3.4. OUI 및 IEEE 규칙에 따른 Rand-MAC 분류

일련의 과정을 거쳐 Wi-Fi 스캐너에서 수집된 로우 데이터(MAC 주소)를 유동인구 데이터로 가공하면 주변의 실제 존재하는 스마트폰의 개수보다 유동인구 데이터가 더 많이 집계되는 양상이 확인되는데, 이것은 하나의 스마트폰 기기에서 Real-MAC뿐만 아니라 Rand-MAC 중복되어 검출되기 때문이다.

과거에는 스마트폰이 Probe Request를 보낼 때만 Rand-MAC을 사용하고 공유기(AP)와 연결될 때는 Real-MAC을 사용하였지만, 현재는 개인정보 보호 강화조치로 AP에 연결될 때도 Rand-MAC을 사용한다. 따라서 정확한 유동인구 계측 값을 위해서는 Real-MAC과 Rand-MAC에 대한 분류가 반드시 필요하다.

본 연구에서는 Rand-MAC 분류를 위해 가장 간단하게 사용할 수 있는 2가지 방식을 사용하였다.

가. Wi-Fi Probe Request로 포착된 스마트폰의 MAC은 총 48비트 중 첫 24비트로 구성된 OUI(Organizational Unique Identifier) 제조회사 식별코드를 통해 Vendor(제조사) 데이터를 추출하여 존재하지 않는 번호는 Rand-MAC으로 분류한다. OUI MAC 주소는 제조사별로 고유하며, 다른 회사에서 사용하는 MAC 주소와는 중복되지 않기 때문에 Rand-MAC은 IEEE 데이터베이스에 존재하지않는 점을 활용하는 것 이다.[11]

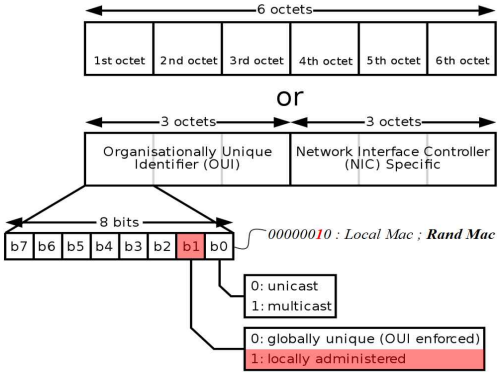

그림 5. Mac 주소의 구조적인 요소 (IEEE)

나. 그림 5에서 붉은색으로 음영 처리한 b1비트는 MAC 주소가 전역(0:globally) 또는 로컬(1:locally)로 관리되는지 결정한다. 이때 IEEE의 규칙에 의해 Rand-MAC으로 간주 되려면 MAC 주소가 로컬에서 관리되어야 한다. 따라서 첫 번째 옥텟(8bit)이 두 번째로 작은 비트(NN10)로 설정된 경우, 로컬에서 관리되는 Rand-MAC으로 간주한다.[12]

표 1. Local MAC Address quadrants [13]

Second Hex Digit of MAC address	
2 (0010)	Administratively Assigned Identified (AAI)
6 (0110)	Reserved
A (1010)	Extendel Local Identifier (ELI)
E (1110)	Standard Assigned Identifier (SAI)

위 표 1의 붉은색 숫자 1을 확인하면 첫 옥텟의 마지막에서 2번째 비트가 1로 표기되는 16진법 수는 2, 6, A, E임을 알 수 있다.

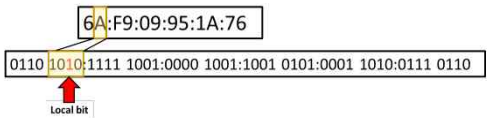

그림 6. Local Bit Within MAC [13]

따라서 그림 6의 확대 표시된 Local bit와 같이 16진법으로 표현된 6개의 옥텟 중에 첫 번째 옥텟이 2, 6, A 또는 E로 끝나는지를 확인함으로써 Rand-MAC을 추출한다. 예를 들어 MAC주소 6A:F9:09:95:1A:76는 첫 번째 옥텟은 6A이고 A(16진법):1010(2진법)로 끝나기 때문에 Rand-MAC으로 분류 된다.

3.5. 원본 데이터와 필터링 데이터 비교 분석

Wi-Fi 스캐너의 원본 데이터와 Rand-MAC을 제거한 필터링 데이터간의 비교분석을 위해 부산광역시에 위치한 지하철 해운대역에서 30분간 테스트를 진행하였다. 그 결과는 다음 그림 7과 같다.

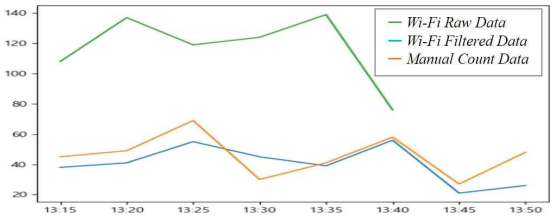

그림 7. Wi-Fi 스캐너 수집/가공 데이터 비교

그림 7의 녹색으로 표기된 Rand-MAC이 필터링 되지 않은 원본 데이터 수치는, 주황색으로 표기된 실측 데이터의 수치와 큰 차이를 보인다. 또한 파란색으로 표기된 Rand-MAC을 일괄 제거한 필터링 데이터의 수치도 실측 데이터 수치와의 차이가 존재한다. 이에 따라 유동인구 추정에 대한 정밀도 개선을 위해서는 Rand-Mac을 일괄 제거하는 것이 아닌, Rand-MAC 수치를 몇 개의 Real-MAC이 만들어냈는지 추정하는 알고리즘이 필요하다.

3.6. Repeat and Non-repeat Rand-MAC 분류

공공와이파이 등의 특정 SSID에 연결되어있는 Repeat Rand-MAC은 Wi-Fi 스캐너가 감지하는 동안 하나의 Rand-MAC으로 감지된다. 따라서 6회 이상 동일한 Rand-MAC 검출 시 Repeat Rand-MAC으로 간주하고 1개의 유동인구 수치로 변환한다.

그러나 Non-repeat Rand-MAC은 한 번의 Probe Request에서 1~3회 사용된 뒤 또 다시 새로운 Rand-MAC값으로 변환된다. 따라서 Wi-Fi 스캐너에 감지된 Rand-MAC의 수치가 몇 개의 Real-Mac에서 파생되었는지 확인을 하고자 다음과 같은 테스트를 진행하였다.

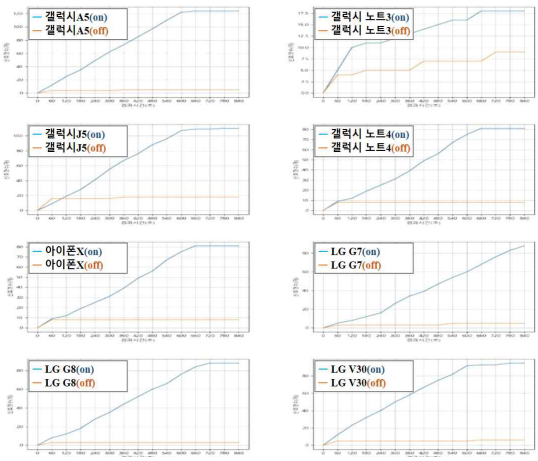

그림 8. 스크린 on/off에 따른 Rand-Mac 비교

그림 8은 유동인구에 카운팅 되는 사람의 스마트폰 화면은 항상 On 상태가 아닐 것이라는 상황을 가정하여 화면이 On 되어있을 때와 Off 되어있을 때의 차이를 구분하기 위한 테스트 결과이다.

스마트폰 화면이 On 상태일 경우 시간에 따라 일정한 신호를 계속 보내는 반면, 화면이 Off 된 경우에는 처음 1분 이외의 구간에서는 Rand-MAC의 개수가 다소 일정한 것으로 나타났다.

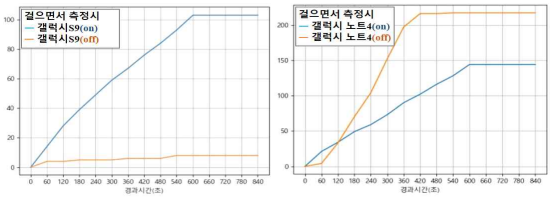

그림 9. 걷고 있을 때의 Rand-Mac 수치 비교

그림 9는 건물에 사람이 없는 새벽 2시경 스마트폰을 소지한 실험자가 실험실 밖의 복도를 걸으면서 측정한 결과이다.

좌측 삼성 갤럭시 S9 기기의 그래프를 확인하면, 화면이 OFF 되어있을 때의 Rand-MAC 수치가 화면 ON 상태에서보다 더 낮게 측정되었다. 즉, Rand-Mac의 수치 변화는 움직이고 있는 상황에서보다는 스마트폰 화면의 On/Off 여부에 영향을 더 많이 받는 것으로 나타났다.

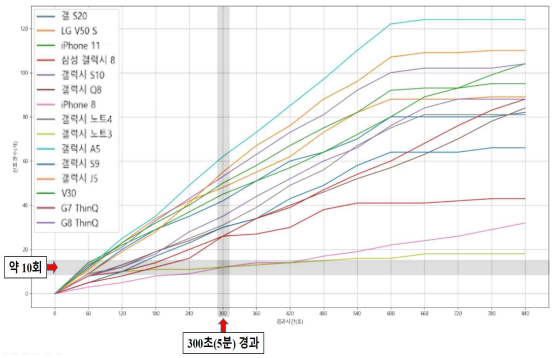

그림 10. 스마트폰 기종별 Rand-MAC 발생추이

그림 10은 외부전파가 차단되는 테스트용 박스에 Wi-Fi 스캐너와 스마트폰 15종을 각각 하나씩 보관하여, 스마트폰마다 1개의 Real-MAC이 몇 개의 Rand-MAC을 발생시키는지 확인한 결과이다.

스마트폰 기종별 반복 실험을 진행한 결과, 똑같은 스마트폰이더라도 Rand-MAC을 보내는 패턴이 일정하지 않았다. 즉, MAC 주소의 무작위화는 디바이스의 기종이나 펌웨어의 버전 그리고 시간의 흐름과 무관한 비선형 관계를 지닌다.

그림 8, 9, 10 에서의 테스트를 통해 확인한 특징을 기반으로, 본 연구에서는 Wi-Fi 스캐너에 감지된 Rand-MAC의 수치가 몇 개의 Real-Mac에서 파생되었는지 적용하는 기준점을 Wi-Fi 스캐너가 데이터를 정제하는 5분(300초)을 Y축 기준점으로 X축과 만나는 최소 지점인 10회로 정하였다. 즉, 필터링 된 Rand-MAC 10회당 1개의 Real-MAC이 존재한다고 가정하여 실험에 임한다.

3.7. 실험 환경에서의 Rand-MAC 분류 절차

앞서 열거한 내용을 토대로 반송큰시장의 CCTV와 비교하기 위한 Wi-Fi 스캐너 데이터는 다음 그림 11과 같은 프로세스로 수집되어 가공된다.

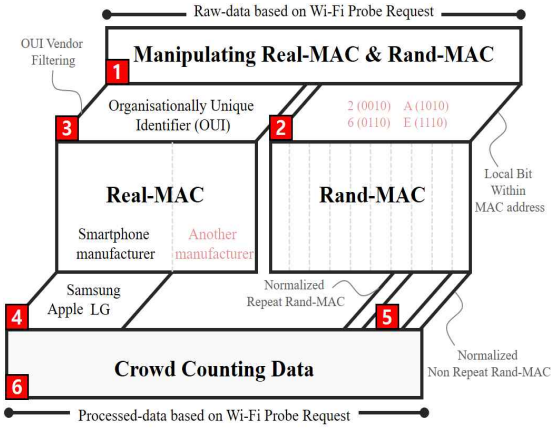

그림 11. Rand-MAC 분류 프로세스

첫째, Wi-Fi 스캐너의 범위에 존재하는 스마트폰 디바이스가 Wi-Fi Probe Request를 진행함에 따라 추출되는 MAC 주소가 Raw-Data로 수집된다.

둘째, 수집된 MAC 주소값이 로컬에서 관리되는 Rand-MAC인지를 확인하기 위해 첫 번째 옥텟이 두 번째로 작은 비트인 2, 6, A, E로 끝나는 MAC 주소값을 Rand-MAC으로 1차 분류한다.

셋째, Rand-MAC으로 분류되지 않은 MAC주소값 중 IEEE의 제조사 번호 OUI에 등록되어있지 않은 MAC 주소값을 제거하고 남겨진 MAC 주소값을 Real-MAC으로 1차 분류한다.

넷째, 1차 분류된 Real-MAC 값에서 스마트폰 제조사인 Samsung, LG, Apple 3가지 기업의 OUI로 2차 분류한다. (본 논문에서 메인 실험이 이루어지는 장소는 부산광역시 해운대구 반송큰시장으로 3개 제조사 이외의 스마트폰 점유율은 현저히 낮다.)

다섯째, 1차 분류된 Rand-MAC 값에서 실제 Real-MAC 개수를 다시 필터링하기 위해, Repeat Rand-MAC이 6회 이상 검출 시에 Real-MAC 1개로, Non Repeat Rand-Mac이 10회 검출 시에 Real-mac 1개로 처리하여 분류한다. (Normalized Rand-MAC)

여섯째, 2차 분류된 Real-MAC 값과 Normalized Rand-Mac 값을 합산하여 유동인구 데이터로써 가공하여 사용한다.

Ⅳ. 실 험

4.1. Wi-Fi 스캐너와 CCTV 데이터의 비교

Rand-MAC을 분류하여 추출한 Wi-Fi 스캐너의 유동인구 데이터와 CCTV에서 실제 집계된 유동인구 데이터간의 비교를 위한 부산 반송큰시장에서의 실험은 오전 9시부터 오후 6시 59분까지 총 9시간 동안 진행되었다.

그림 12. 반송큰시장 CCTV 영상 USB 백업

CCTV 유동인구 데이터 계측을 위해 실험 참여자 4명은 반송큰시장 상인회 사무실에 있는 CCTV 영상의 화면 녹화(그림 12)와 동시에 모니터 화면에서 사람들이 지나갈 때마다 손으로 직접 계측하는 수동 인원 계수기를 활용하여 카운팅 하였다. 이와 동시에 Wi-Fi MAC 주소 기반 유동인구 데이터가 반송큰시장에 구축되어있는 Wi-Fi 스캐너를 통해 자동으로 데이터베이스에 잘 쌓이고 있음을 확인하였다.

Wi-Fi 스캐너 1 설치 위치의 CCTV 1 유동인구 데이터 집계는 CCTV 영상화면 A(그림 13)를 통해 확인하였으며, 해당 영역의 시간당 유동인구 데이터는 다음 표 2와 같다.

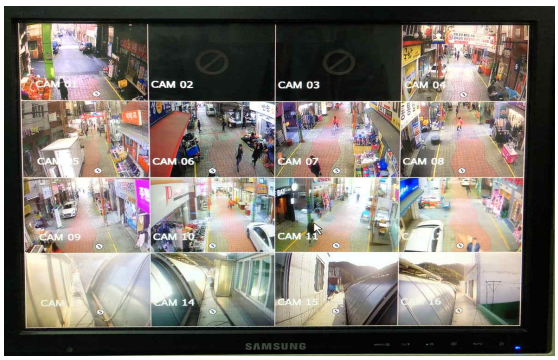

그림 13. 반송큰시장 CCTV 영상화면 A

표 2. Wi-Fi Scanner 1 vs CCTV 1 Count

Time	09	10	11	12	01	02	03	04	05	06
Wi-Fi 1	435	531	523	530	537	569	601	621	489	634
CCTV 1	310	223	438	381	408	330	445	385	359	474

Wi-Fi 스캐너 2가 설치되어있는 CCTV 2의 유동인구 데이터 집계는 CCTV 영상화면 B(그림 14)를 통해 확인하였으며, 해당 영역의 시간당 유동인구 데이터는 다음 표 3과 같다.

그림 14. 반송큰시장 CCTV 영상화면 B

표 3. Wi-Fi Scanner 2 vs CCTV 2 Count

Time	09	10	11	12	01	02	03	04	05	06
Wi-Fi 2	297	223	367	489	401	437	398	482	512	433
CCTV 2	147	177	244	250	227	298	354	220	281	419

Wi-Fi 스캐너 3과 4가 설치되어있는 CCTV 3과 CCTV 4의 유동인구 데이터 집계는 CCTV 영상화면 C(그림 15)를 통해 확인하였으며, 해당 영역의 시간당 유동인구 데이터는 각각 다음 표 4, 표 5와 같다.

그림 15. 반송큰시장 CCTV 영상화면 C

표 4. Wi-Fi Scanner 3 vs CCTV 3 Count

Time	09	10	11	12	01	02	03	04	05	06
Wi-Fi 3	428	642	1031	1272	1240	1061	1250	1363	1392	1574
CCTV 3	255	394	842	1069	817	569	1039	967	1149	890

표 5. Wi-Fi Scanner 4 vs CCTV 4 Count

Time	09	10	11	12	01	02	03	04	05	06
Wi-Fi 4	529	448	752	928	905	774	990	872	1016	1224
CCTV 4	241	337	405	581	624	522	707	799	610	1002

아래 그림 16의 실선은 9시부터 18시까지의 Wi-Fi 스캐너 4대의 데이터 값을 1시간 단위로 시각화한 값이다. 유동인구의 흐름은 Wi-Fi 3, Wi-Fi 4, Wi-Fi 1, Wi-Fi 2 순으로 밀집도가 높은 것으로 확인되었다.

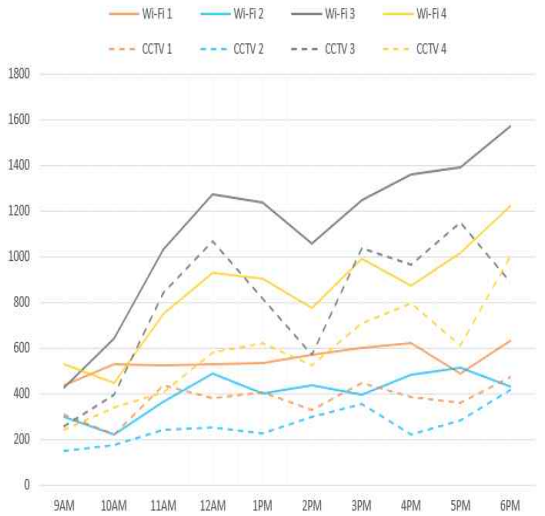

그림 16. Comparison of Wi-Fi and CCTV Data

그림 16의 점선은 CCTV 화면을 통한 수동 집계 데이터로써 CCTV 1,2,3,4간의 유동인구의 밀집도 순서는 각각 Wi-Fi 스캐너 1,2,3,4와 순서가 같았다. 그러나 집계 수치에서 차이가 발생하였는데, 특히 CCTV 3과 Wi-Fi 스캐너 3의 데이터 차이가 높았다. 이러한 차이는 집계 위치가 시장의 메인 입구로써 하루 동안 실제 유동인구가 가장 높은 지역이며 입구 앞에 큰 도로가 위치하기 때문에 변수값이 큰 것으로 예상하였다.

4.2. 데이터 오차에 대한 가설 설정과 확인

CCTV 3과 같은 특수상황을 제외하고 전반적으로 나타나는 집계 차이는, CCTV 기반의 수동 카운팅과 Wi-Fi 스캐너 기반의 계측방식 차이로써 다음과 같은 4가지 변수를 가정하였다.

가. CCTV 수동 계측으로는 통로가 아닌 상점 안에 있는 가게의 주인을 카운팅하지 못한다.

나. CCTV 화면의 사각지대에 있는 일부 고객이나 자동차를 타고 지나가는 사람들의 수치를 카운팅 하지 못한다.

다. Wi-Fi 스캐너의 경우 벽과 같은 장애물이나 시야에서 보이지 않는 사각지대에 상관없이 계측한다.

라. Wi-Fi 스캐너의 넓은 센싱 범위로 인하여, CCTV처럼 화면 속 특정 공간만을 계측한 것이 아니라 시장 밖에 위치한 사람이나 인근 도로의 운전자가 계측되었다.

CCTV 수동 계측에서 도출된 변수인 (가)와 (나)의 변수는 반송큰시장의 각 매장별로 실내에서 근무 중인 상인의 수를 확인하고, CCTV의 사각지대에는 별도의 웹캠을 설치하여 측정되는 보완값을 CCTV 수동 계측값에 추가함으로써 극복할 수 있다. 그러나 (가)와 (나)의 변수는 CCTV를 이용하는 수동 계측방식의 정확도를 향상시키기 위한 조치이므로 Wi-Fi Probe Request를 기반으로 유동인구를 추정하는 본 연구에서는 적용하지 않는다.

Wi-Fi 스캐너에서 도출된 변수인 (다)와 (라)의 커버리지 변수를 해결하기 위해서는 다음과 같은 두 가지 필터를 활용한다.

4.3. RSSI 거리 필터 적용

Wi-Fi 스캐너의 raw-data에 수집되는 항목인 범위 내 스마트폰 RSSI(신호세기)값이 스캐너와 스마트폰 간의 위치적인 거리값과 상관관계가 있다고 가정했을 때, 강한 RSSI값이 나타나는 스마트폰은 스캐너와 가까운 곳에 위치하고 약한 RSSI값의 스마트폰은 멀리 있는 곳에 있다는 것을 의미한다. 이러한 상관관계를 이용하여 임의의 RSSI값을 설정하고 해당 수치보다 약한 RSSI 값은 제외하면 스캐너와 가까운 스마트폰만을 집계할 수 있다.

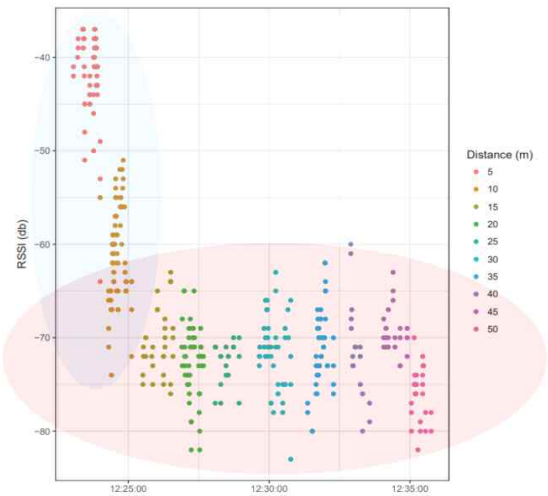

그림 17. RSSI changes as the distance [14]

그림 17의 그래프 추이를 확인하면 광활지에서 Wi-Fi Probe Request가 진행될 때, 스마트폰으로부터 받는 RSSI값은 스캐너로부터 50m 떨어진 곳에서도 도달하며 [14], [15] RSSI 신호값이 약해질수록 거리가 증가하는 상관관계가 있다. 그러나 특정 RSSI(-40)보다 신호가 약해지는 경우에는 스마트폰의 종류나 성능에 따라서 RSSI와 거리값의 상관관계가 급격히 떨어지는 경향을 보인다.

이 현상을 구분 짓기 위해 그림 17의 좌상단에 푸른색 음영지역으로 5m, 10m의 거리에 위치한 스마트폰의 신호세기값 (RSSI -40 ~ -70)을 그룹

화 해두었으며, 하단의 붉은색 음영지역에는 10m~50m 거리에 위치한 스마트폰의 신호세기값(RSSI -60 ~ -80)을 그룹화하여 표시하였다. 푸른색 음영지역과 붉은색 음영지역의 교차지점을 확인하면 RSSI -60까지가 RSSI와 거리의 상관관계가 유효한 지표임을 알 수 있다. 따라서, RSSI -60보다 약한 신호세기의 MAC 주소를 제외하면 스캐너 설치 반경 10m 이내에 위치한 스마트폰만을 유동인구 수치로 변환할 수 있다.

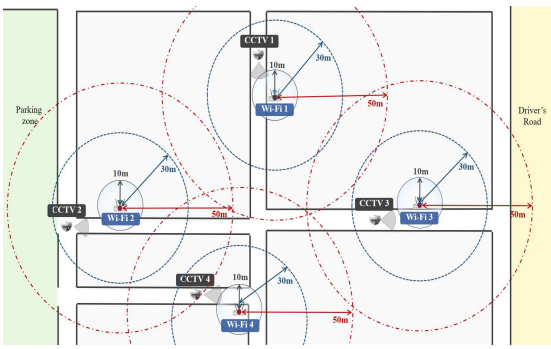

그림 18. Estimated signal distance from Wi-Fi

반송큰시장 상점 앞 메인통로의 폭은 약 6m이다. 따라서 RSSI값이 -60보다 약한 신호를 일괄 제거함으로써 상점 앞을 지나가는 유동인구만 계측하는 RSSI 거리 필터를 적용한다.

4.4. 속도, 거리에 의한 Time 필터 적용

유동인구의 사전적 의미는 일정한 기간 동안 한 지역을 오가는 사람의 수[16]로 정의되어 있지만, 데이터의 사용 주체에 따라 도로변의 자동차나 오토바이 운전자들은 일반적으로 노이즈 데이터로 처리한다. 더욱이 테스트가 진행되는 반송큰시장의 경우 시장 외곽에 일반도로가 형성되어 있기 때문에 거리필터에서 미처 필터링 되지 못한 차량 데이터를 처리하기 위해 시간 필터링을 사용한다.

차량에 탑승하고 있는 운전자는 일반보행 속도보다 빠르기 때문에 하나의 스캐너의 범위에서 빠르게 사라지게 된다. 또한 A스캐너의 탐지범위에서 B스캐너의 탐지범위로의 이동되는 속도가 빠름을 알 수 있다. 이러한 특징을 활용하여 짧게 사라지는 데이터와 둘 이상의 Wi-Fi 스캐너에서 t (t=d/v)에 캡처된 데이터는 차량 탑승자로 간주하여 유동인구 계측값에서 제외하는 Time 필터를 적용한다.

4.5. 필터링된 데이터와의 비교 및 상관 계수

Wi-Fi 스캐너에서 도출된 커버리지 변수 (다)와 (라)를 해결하고자 RSSI 거리 필터와 Time 필터를 적용한 최종 결과값은 각각 다음의 그림 19 및 표 6, 7, 8, 9와 같다.

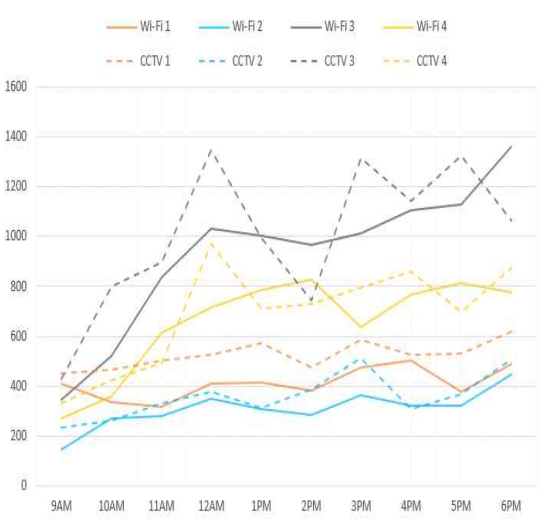

그림 19. Comparison of Processed-data

표 6. Filtered Wi-Fi 1 vs CCTV 1 Count

Time	09	10	11	12	01	02	03	04	05	06
Wi-Fi 1	412	334	319	408	413	384	474	501	376	488
CCTV 1	453	466	501	524	571	473	588	528	532	617

표 7. Filtered Wi-Fi 2 vs CCTV 2 Count

Time	09	10	11	12	01	02	03	04	05	06
Wi-Fi 2	148	270	281	349	307	284	365	321	324	446
CCTV 2	234	264	331	377	314	385	511	307	368	506

표 8. Filtered Wi-Fi 3 vs CCTV 3 Count

Time	09	10	11	12	01	02	03	04	05	06
Wi-Fi 3	346	520	835	1030	1004	965	1012	1104	1127	1358
CCTV 3	430	799	897	1344	992	744	1314	1142	1324	1065

표 9. Filtered Wi-Fi 4 vs CCTV 4 Count

Time	09	10	11	12	01	02	03	04	05	06
Wi-Fi 4	271	360	615	714	787	828	635	765	815	774
CCTV 4	330	426	494	970	713	731	796	858	699	872

RSSI 거리 필터와 Time 필터를 통해 가공된 유동인구 데이터와 CCTV 기반 수동계측 데이터의 상관계수가 0.75로 나타남에 따라, 두 변수 값이 어느 정도 유사한 상관 관계에 있음을 확인하였다.

$$r_{xy} = \frac{\sum_{i}^{n}(X - LSUBi\overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i}^{n}(Y_i - \overline{Y})^2}}$$

$P(Wi\text{-}Fi1,\ CCTV1) = 0.64$
$P(Wi\text{-}Fi2,\ CCTV2) = 0.84$
$P(Wi\text{-}Fi3,\ CCTV3) = 0.75$
$P(Wi\text{-}Fi4,\ CCTV4) = 0.80$
$Average\ correlation\ of\ all\ P(Wi\text{-}Fi,\ CCTV) = 0.75$

그러나 변수 극복을 위한 데이터 필터링을 진행하였음에도 불구하고 몇몇 특정 시간에서는 여전히 Wi-Fi 스캐너의 가공된 데이터가 CCTV 기반의 수동 계측값 보다 높거나 낮은 영역이 존재한다. 이러한 현상은 다음 4가지 변수로 발생 된다.

첫째, 계측 공간에 진입한 스마트폰의 사용자가 Wi-Fi 모듈을 완전히 OFF하거나 스마트폰의 전원을 OFF하여 Probe Request 탐지가 되지 않는 경우

둘째, Wi-Fi 스캐너 반경 내에 위치한 스마트폰의 기종에 따라 Probe Request 탐지하는 간격의 차이가 길거나 사용자의 걸음이 빨라서 MAC 주소를 확인하기 전에 스캐너 반경을 먼저 이탈한 경우

셋째, Wi-Fi 스캐너의 반경 내에 위치한 스마트폰이 active 상태가 아닌 sleep mode로 진입되어 있어서 Probe Request 탐지하지 않고 있는 경우

넷째, 스마트폰마다 직전과 다른 Rand-MAC으로 변환하는 시간과 횟수가 다양하지만 일괄적으로 Rand-MAC n개당 1개의 Real-MAC으로 처리하는 과정 중에 실제 위치하는 스마트폰보다 많거나 적은 값이 적용된 경우

이와 같은 변수를 해결하기 위해 Wi-Fi 모듈의 감지 감도의 정밀성과 Rand-MAC을 만들어내는 Real-MAC의 실제량에 대한 정밀한 알고리즘이 추가적으로 연구된다면 수동 계측 방식의 유동인구 데이터와 Wi-Fi Probe Request 기반의 유동인구 데이터의 계측값이 상당히 유사해질 것으로 기대된다.

V. 결 론

본 논문에서는 Rand-MAC 값이 필터링 된 Wi-Fi Probe Request 기반의 유동인구를 추정하는 방법에 대해 서술하고, 이후 전통시장에서의 실험을 통해 CCTV 화면을 토대로 수동 계측한 데이터와의 상관 계수를 확인하였다.

실험 결과 Wi-Fi 스캐너와 CCTV 수동 계측 데이터 간의 상관 계수는 (Pearson Correlation) 평균 0.75로 두 데이터가 어느 정도 유사하다는 결과가 나왔다.

이러한 결과를 기반으로, Wi-Fi를 활용한 유동인구 계측 시스템을 고비용의 지능형 CCTV 시스템 등을 대체하여 활용할 수 있다. 그러나 정밀한 수치에서는 두 데이터 간의 차이가 분명 존재하기 때문에, 단 1명의 오차도 허용되지 않는 보안/안전 목적의 Wi-Fi 기반 유동인구 데이터 사용에 있어서는 더욱 정밀한 알고리즘 연구가 필요하다.

개개인의 스마트폰의 MAC 주소값을 수집하여 분석하는 Wi-Fi Probe Request의 특징상 여러 목적을 위해 악용될 수 있는 여지가 있다. 따라서 수집되는 MAC 주소에 대한 비식별처리를 통해, 특정 구역에서 사람들의 존재를 계측하는 용도로만 사용되어야 한다. 또한, 앞으로의 Wi-Fi 기반 유동인구 계측에 대한 연구 방향에 있어서는 데이터의 신뢰도와 정확성은 끌어올리되, 고유 MAC 주소를 분석하여 개개인을 추적하는 연구 방향은 지양함으로써, 시민의 개인정보보호에 앞장서는 성숙한 연구와 사업화가 진행되어야 한다.

Acknowledgement

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음
(과제번호 21TBIP-C160515-01).

This work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 21TBIP-C160515-01).

References

[1] J. Kim, K. Choi, B. Choi, and S. Ko, “Real-time vision-based people counting system for the security door,” … *Tech. Conf. Circuits Syst.* …, vol. 3672, 2002.

[2] [A. Rao, J. Gubbi, S. Marusic, P. Stanley, and M. Palaniswami, “Crowd density estimation based on optical flow and hierarchical clustering,” 2013 *Int. Conf. Adv. Comput. Commun. Informatics,* pp. 494–499, 2013.

[3] R. Melfi, B. Rosenblum, B. Nordman, and K. Christensen, ‘‘Measuring building occupancy using existing network infrastructure,’’ in *Proc. Int. Green Comput. Conf. Workshops,* pp. 1–8, Jul. 2011, .

[4] Z. Wang and L. Wang, ‘‘Occupancy pattern based intelligent control for improving energy efficiency in buildings,’’ in *Proc. IEEE Int. Conf. Automat. Sci. Eng. (CASE),* Aug. 2012, pp. 804–809.

[5] B. Sun, P. B. Luh, Q.-S. Jia, Z. Jiang, F. Wang, and C. Song, ‘‘Building energy management: Integrated control of active and passive heating, cooling, lighting, shading, and ventilation systems,’’ *IEEE Trans. Autom. Sci. Eng.,* vol. 10, no. 3, pp. 588–602, Jul. 2013.

[6] W. Xi, J. Zhao, X.-Y. Li, K. Zhao, S. Tang, X. Liu, and Z.Jiang, “Electronic frog eye: Counting crowd using WiFi,” *IEEE INFOCOM 2014 - IEEE Conf. Comput. Commun.,* pp. 361–369,

2014.
[7] D. Aubert, ‘‘Passengers queue length measurement,’’ in *Proc. 10th Int. Conf. Image Anal. Process.,* pp. 1132−1135, Sep. 1999, .
[8] U. Stilla, E. Michaelsen, U. Soergel, S. Hinz, and H. J. Ender, ‘‘Airborne monitoring of vehicle activity in *urban areas,’’ Int. Arch. Photogramm. Remote Sens.,* vol. 35, no. B3, pp. 973−979, 2004.
[9] R. H. Dodier, G. P. Henze, D. K. Tiller, and X. Guo, ‘‘Building occupancy detection through sensor belief networks,’’ Energy Buildings, vol. 38, no. 9, pp. 1033−1043, Sep. 2006.
[10] Y. Wang, J. Yang, Y. Chen, H. Liu, M. Gruteser, and R. P. Martin, ‘‘Tracking human queues using single-point signal monitoring,’’ in *Proc. 12th Annu. Int. Conf. Mobile Syst., Appl., Services,*pp. 42−54, 2014.
[11] IEEE-STANDARDS ASSOCIATION, “Guidelines for use of the 24-bit Organizationally Unique Identifiers (OUI)” [PDF]. PP. 10-15, Abailabe : https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/tutorials/eui.pdf
[12] IEEE-STANDARDS ASSOCIATION, “Standard Group MAC Addresses: A Tutorial Guide” [PDF], PP. 1-2, Available : https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/tutorials/macgrp.pdf
[13] Carol Ansley, CommScope, Inc, MAC Randomization in Mobile Devices, *A Technical Paper prepared for SCTE•ISBE*, PP. 6-7, 2019.
[14] L. Gao, S. Tang, Y. Bai, A New Approach for Wi-Fi-Based People Localization in a Long Narrow Space, Wireless Communications and Mobile Computing, Hindawi, pp. 4-7, 2019.
[15] Bullmann M, Fetzer T, Ebner F, Ebner M, Deinzer F, Grzegorzek M. Comparison of 2.4 GHz WiFi FTM- and RSSI-Based Indoor Positioning Methods in Realistic Scenarios. Sensors. PP. 15-18, 2020; 20(16):4515. https://doi.org/10.3390/s20164515
[16] Research Institute of Korean Studies Korea University, Korea Language Dictionary of Korea University, Seoul, South Korea, 2009.