10.20378/irb-97611

Otto-Friedrich-Universität Bamberg

Verbundvorhaben „CrowdAnym“

Eine Vorstudie zu Möglichkeiten anonymer Datenerfassung als Grundlage datengetriebener Besuchslenkung in der Bamberger Altstadt

Gefördert über die Innovationsinitiative „mFUND“ des BMDV

(Bundesministeriums für Digitales und Verkehr)

Kurzdokumentation des Mobilithek Datensatzes

Leonie Ackermann, Christoph Baum, Aleksandr Litvin und Daniela Nicklas

Lehrstuhl für Informatik, insbesondere Mobile Softwaresysteme/Mobilität

Otto-Friedrich-Universität Bamberg

Inhaltsverzeichnis

1. Überblick

In der bei Tourismus und Bevölkerung gleichermaßen beliebten Bamberger Altstadt kommt es immer wieder zu örtlicher Überlastung. Ein daten- und sensorbasiertes System soll daher zukünftig smarte Empfehlungen geben und die Situation für alle Beteiligten verbessern. In einer Vorstudie im Rahmen des Forschungsprojekts CrowdAnym wurde mit einer Testinstallation im Feld untersucht, ob die Datenqualität auch bei starker Anonymisierung für solche Vorhaben ausreicht. Das implementierte System verwendet Wi-Fi-Sensoren als passive, unaufdringliche und kostengünstige Option zur Einschätzung der Besucherfrequenz.

Für die Testinstallation wurden im Stadtgebiet Bamberg insgesamt neun Sensoren (Wi-Fi Tracker) an touristisch relevanten Orten installiert. Die Sensoren messen Wi-Fi Probe Requests von verschiedenen Endgeräten, wie z.B. Smartphones, Smartwatches oder Laptops. Die Sensordaten enthalten die anonymisierte MAC-Adresse (Salted Hashing, mit täglich wechselndem Salt), einen Zeitstempel, die Sensor-ID und die Signalstärke (Received Signal Strength Indicator).

Der Datensatz, der in der Mobilithek zur Verfügung gestellt wird, enthält die Anzahl der erfassten Geräte aggregiert nach Zone (bzw. Sensor), Zeitfenster und Signalstärke.

2. Sensoren und Standorte

Die folgende Tabelle gibt einen Überblick welche SensorID („zone“) zu welchem Standort gehört. Abbildung 1 zeigt die Verteilung der Sensoren in der Bamberger Innenstadt.

SensorID	Ort
bz2452	Grüner Markt/Gabelmann
bz2453	Sandstraße
bz2454	Mußstraße
bz2457	Domkranz
bz2458	Touristeninformation
bz2460	Altes Rathaus
bz2462	Neues Rathaus Ost
bz2463	Neues Rathaus West
bz2464	Maxplatz

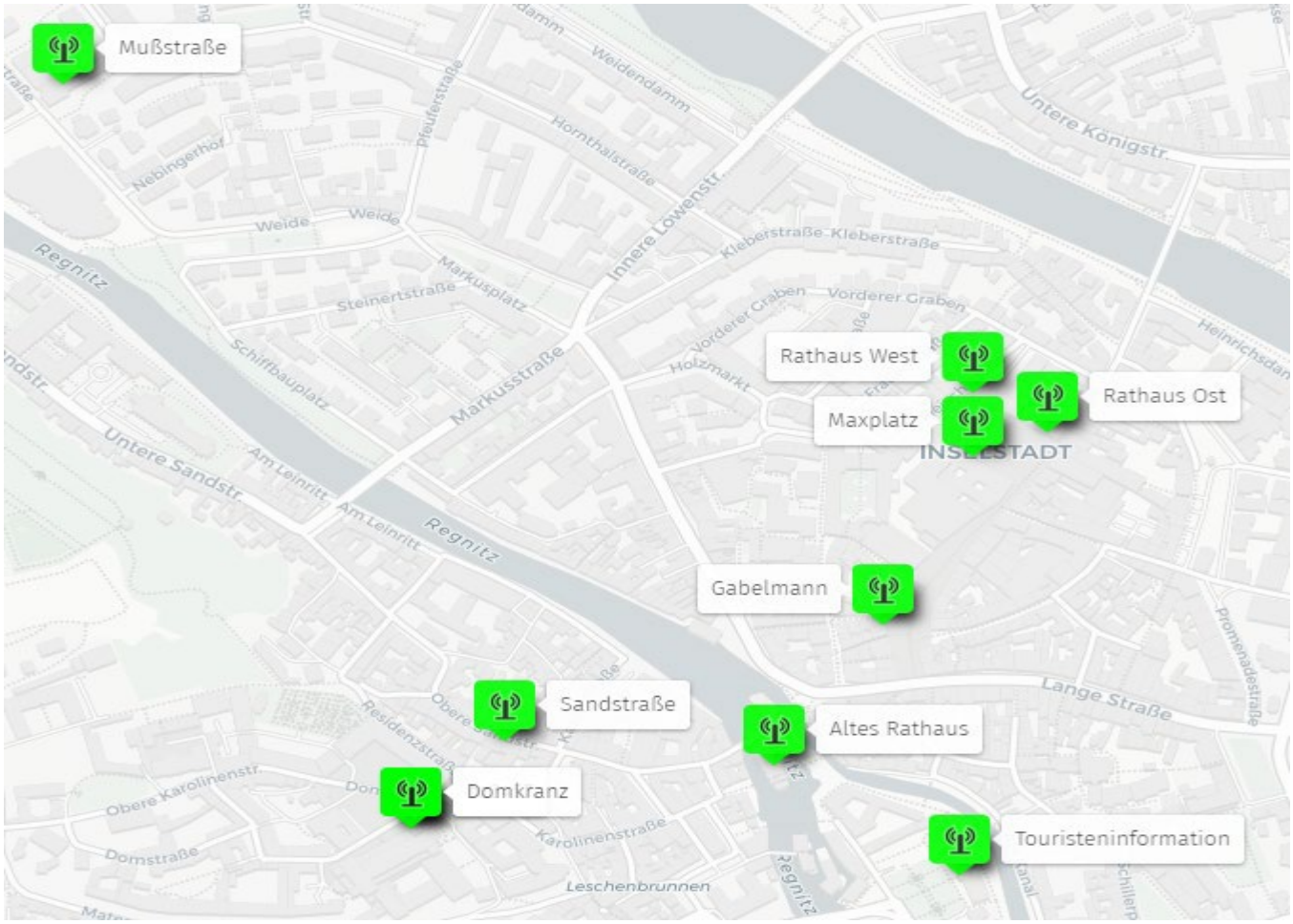

Abbildung 1: Standorte der Sensoren

3. Datenstruktur und Inhalt

Die Anzahl der eindeutigen MAC-Adressen wird als Integer dargestellt und aggregiert nach Zone, Zeitfenster und Entfernung. Der Datensatz enthält insgesamt 3 Tabellen im csv-Format mit Zeitfenstern von 1 Minute, 10 Minuten und 60 Minuten. Der Datensatz deckt den Zeitraum 10. Juli 2023 bis 20. August 2023 ab.

Beschreibung der Datenattribute:

Attribut	Datentyp	Erklärung
zone	String	Sensor ID (Standort)
rssi_group	String	Entfernung zwischen Gerät und Sensor, kategorisiert in: - weit („wide“): von -81 bis – 100 dBm; - mittel („mid“): von -61 bis -80 dBm; - nah („close“): von 0 bis -60 dBm, basierend auf der Signalstärke des eines Wifi Probe Requests (RSSI - Received Signal Strength Indicator)
epocutc	DateTime	Zeitstempel in UTC (Zeitzone CET) im Format JJJJ-MM-TT SS:MM:ss

4. Anonymisierung und Datenschutz

Wi-Fi Probe Requests sind anfällig für Angriffe wie Fingerprinting, können persönliche Informationen preisgeben und ermöglichen die Verfolgung von Gerätebesitzenden. Da es sich bei MAC-Adressen (auch virtuellen, dynamischen) außerdem um personenbezogene Daten im Sinne von Art. 4 Nr. 1 DSGVO handelt, müssen diese Daten vor ihrer Speicherung und Veröffentlichung geeigneten Anonymisierungstechniken unterzogen werden.

Im Folgenden wird der in CrowdAnym gewählte Ansatz zur privatsphärensensiblen Erfassung von Wi-Fi Probe Request für die Schätzung der Besucherfrequenz sowie zusätzliche Maßnahmen zur Verhinderung der Identifizierung und Verfolgung von Personen erläutert. Hierbei wird ein zweistufiger Ansatz verfolgt: Die Basismaßnahmen greifen direkt bei der Datenerhebung, vor der Aggregation und Datenveröffentlichung werden weitere anwendungsspezifische Anonymisierungsmaßnahmen durchgeführt.

4.1. Basismaßnahmen

Die Sensoren wurden nur an touristisch relevanten Orten installiert. Die meisten Sensorbereiche überschneiden sich nicht, und zwischen den Sensoren gibt es große Lücken. Dies verhindert von vornherein ein umfangreiches Tracking.

Die MAC-Adressen werden von den Sensoren als „Salted Hash“ Wert an die Datenplattform gesendet. Hierfür wird zunächst an die Original MAC-Adresse ein „Salt“ Wert angehängt. Der Salt ist eine zufällige Zeichenfolge, welche sich jeden Tag ändert bzw. täglich auf den Sensoren neu generiert wird.

Im Anschluss wird auf die neue Zeichenkette (MAC-Adresse + Salt) ein SHA-224 Hash Algorithmus angewendet. Das Resultat ist der beschriebene „Salted Hash“ Wert.

Schließlich werden keine Probe Request Frames und Service Set Identifiers (SSIDs) übertragen oder gespeichert. Diese Daten könnten für das Fingerprinting verwendet werden und sensible Informationen preisgeben.

4.2. Anwendungsspezifische Anonymisierungsmaßnahmen

In den nächsten Abschnitten werden die angewandten Filter- und Aggregationsmethoden beschrieben, die den Datensatz für die Analyse der Besucherfrequenz eines Sensorstandorts optimieren und gleichzeitig die Privatsphäre der Gerätebesitzenden zu schützen.

4.2.1. Filterung von statischen MAC-Adressen

Die Sensoren erkennen alle Wi-Fi Probe Requests innerhalb ihrer Reichweite, einschließlich derjenigen, die von Geräten wie Druckern, Laptops und Smart-Home-Geräten gesendet werden. Die Erkennung von Geräten von Anwohnenden oder nahe gelegenen Geschäften und deren Angestellten ist für die Schätzung der Besucherfrequenz nicht relevant. Durch die Analyse der eindeutigen MAC-Adressen und ihres Auftretens innerhalb eines Tages konnte festgestellt werden, dass MAC-Adressen mit einer Häufigkeit von 10 oder mehr eine lange Verweildauer am Sensorstandort haben. Unterhalb dieser Grenze tauchen die Geräte meist nur temporär in den Daten auf. Daher werden MAC-Adressen, die mehr als 10 Mal pro Tag in den Daten auftauchen, aus den Daten entfernt.

4.2.2. Filterung von Zeiträumen mit geringem Datenaufkommen

Ein weiterer anwendungsspezifischer Filter besteht darin, Zeiträume mit geringen Datenmengen zu eliminieren. Solche Zeiträume treten beispielsweise nachts auf, wenn die Straßen weniger frequentiert sind. Um die Privatsphäre der erfassten Personen zu wahren, werden Zeiträume, in denen innerhalb von 10 Minuten weniger als 10 eindeutige MAC-Adressen erfasst werden, aus dem Datensatz vollständig entfernt.

4.2.3. Aggregation nach Zeitfenster und Signalstärke

Außerdem wurden die MAC-Adressen für die Publikation des Datensatzes nach Zone, Zeitfenster und Signalstärke aggregiert. Daher können einzelne Geräte in diesem Datensatz nicht verfolgt werden.

5. Grenzen des Ansatzes

Da die Testinstallation nur Wi-Fi Probe Requests misst und keine genaue Besucherzählung vornimmt, kann auf Basis der Daten nur eine Schätzung der tatsächlichen Besucherfrequenz vorgenommen werden. Eine weitere Herausforderung ist MAC-Randomisierung, eine Funktion zum Schutz der Privatsphäre, die sich in den letzten Jahren immer mehr durchgesetzt hat und von vielen modernen Geräten (z. B. Apple, Android, Windows) verwendet wird. Da statt der tatsächlichen MAC-Adresse Zufallswerte verwendet werden, erkennt der Sensor die Wi-Fi Probe Requests eines Geräts als mehrere Geräte. Außerdem variieren die Zeitabstände in welchen Geräte Wi-Fi Probe Requests senden von Hersteller zu Hersteller, was sich darauf auswirkt, wie häufig ein Gerät im Datensatz auftaucht.

6. Anwendungsbereich

Die Evaluation des Ansatzes hat ergeben, dass trotz der Grenzen des Ansatzes und der Herausforderungen für die Datenqualität, dieser eine effektiv Möglichkeit bietet die Besucherfrequenz an den Sensorstandorten zu schätzen. Dies ermöglicht unter anderem die Vorhersage der Auslastung von bestimmten Sehenswürdigkeiten oder die Echtzeitübermittlung von Besucherfrequenzen an beispielsweise Touristenführer. Dadurch können bei Bedarf alternative Routenvorschläge unterbreitet werden, um eine optimale Verteilung der Besuchenden zu gewährleisten.

7. Zusätzliche Informationen

7.1. Zeitraum & Installationsdaten weiterer Sensoren

Zu Beginn des Erhebungszeitraums waren folgende Wifi Tracker aktiv:

SensorID	Ort
bz2454	Mußstraße
bz2453	Sandstraße
bz2452	Gabelmann
bz2457	Domkranz
bz2458	Touristeninformation

Im Laufe des Erhebungszeitraums kamen weitere Sensoren hinzu:

SensorID	Ort	Installationsdatum
bz2461	Grüner Markt / Maxplatz	03. August 2023
Bz2460	altes Rathaus	10. August 2023
bz2463	Neues Rathaus West	16. August 2023
bz2462	Neues Rathaus Ost	16. August 2023

7.2. Ereignisse

Folgende Tabelle zeigt eine Übersicht über besondere Ereignisse während des Erhebungszeitraums

Datum	Ereignis	Orte
14. Juli – 16. Juli	Bamberg Zaubert	Innenstadt: Grüner Markt, Gabelmann
31. Juli – 11. September	Sommerferien Bayern	alle
04. August – 13. August	Blues und Jazz Festival	Innenstadt: Grüner Markt, Gabelmann

8. Publikation

Leonie Ackermann, Christoph Baum, Syed Ibrahim Khalil, Aleksandr Litvin, and Daniela Nicklas. 2023. Privacy-aware Publication of Wi-Fi Sensor Data for Crowd Monitoring and Tourism Analytics. In Proceedings of the 1st ACM SIGSPATIAL International Workshop on Geo-Privacy and Data Utility for Smart Societies (GeoPrivacy '23). Association for Computing Machinery, New York, NY, USA, 20–23. https://doi.org/10.1145/3615889.3628513