



Manuel Sonder



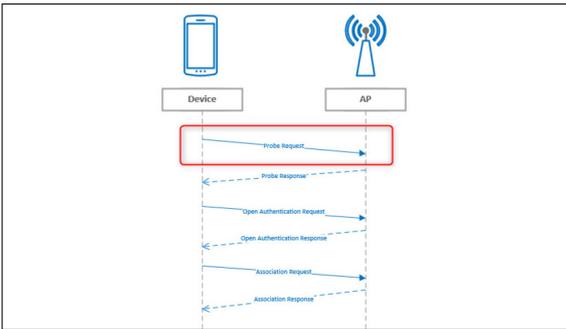
Anojan SHANMUGA NATHAN



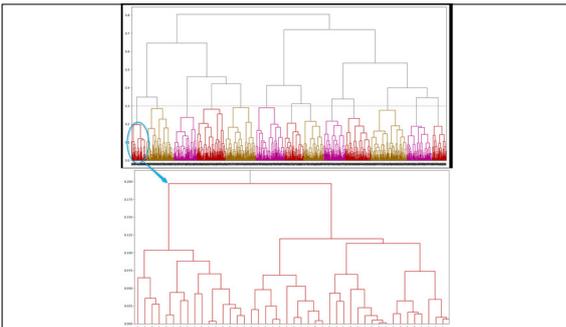
Andreas Fraefel

Studenten	Manuel Sonder, Anojan SHANMUGANATHAN, Andreas Fraefel
Examinator	Prof. Beat Stettler
Themengebiet	Software Engineering - Core Systems
Projektpartner	onway ag, Zürich

Machine Learning für Passenger Tracking



Verbindungsaufbau WiFi Netz
Eigene Darstellung



Beispieldendrogramm einer hierarchischen Clusteranalyse
Eigene Darstellung

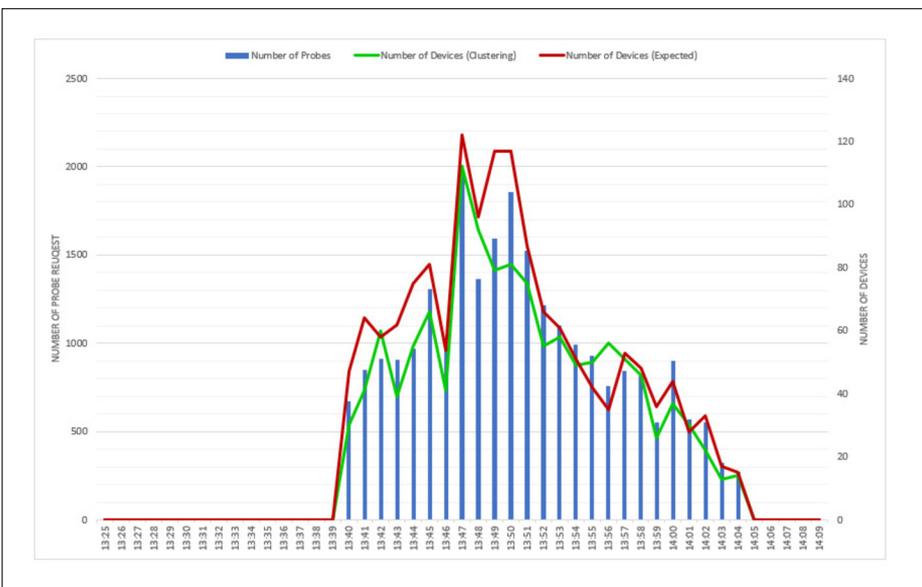
Ausgangslage: Geräte mit eingeschalteter WLAN-Funktion versenden periodisch Probe Requests. Diese Requests enthalten verschiedene Informationen über das Gerät und dessen Eigenschaften (unterstützte Bit-Raten, etc.). Weil Geräte ständig solche Probe Requests senden, wurde die enthaltene MAC Adresse dazu genutzt, um Rückschlüsse auf Personenbewegungen und Aufenthaltsorte zu ziehen. Um dem entgegenzuwirken haben verschiedene Gerätehersteller angefangen die MAC Adresse, in diesen Netzwerkpaketen, zu randomisieren.

Die Firma onway ag betreibt WLAN-Netze in öffentlichen Verkehrsmitteln. Um die Auslastung der Verkehrsmittel effizienter ermitteln zu können, wurde eruiert, ob ein Ansatz mittels Machine Learning erarbeitet werden kann, der trotz randomisierter MAC Adressen, die Geräte erkennt.

Vorgehen: Die von onway ag gelieferten Daten wurden in einem ersten Schritt aufgearbeitet und analysiert. Die Payloads der Probe Requests wurden entpackt und die darin enthaltenen Datenblöcke extrahiert. Anschliessend wurden diverse Machine Learning Techniken auf ihre Eignung untersucht. Schliesslich fiel die Wahl auf eine hierarchische Clusteranalyse. Für das Clustering musste, aufgrund der spezifischen Anforderung, eigens für diese Arbeit, ein Distanzmass entworfen und implementiert werden.

Ergebnis: Für die Aufarbeitung und Klassifizierung der Daten wurde ein Prototyp in Python entwickelt, der die Daten anhand verschiedener Attribute der WLAN-Pakete klassifiziert. Die Gewichtung der Attribute wurde mit einem stochastischen Suchalgorithmus optimiert.

Die Auswertung der Resultate hat ergeben, dass der Algorithmus die Anzahl der Geräte mit einer Genauigkeit von circa 90% abschätzen kann. Allerdings bereiten Probe Requests, welche von Geräten mit gleichen Herstellern verschickt wurden, dem Algorithmus Schwierigkeiten.



Auswertung: Die Grafik zeigt die Anzahl Probe Requests (blau) sowie die erwartete Anzahl an Geräten (rot) und die vom Algorithmus gelieferte Anzahl an Geräten (grün)
Eigene Darstellung