

Mobility-Traffic Correlations

Tim Bohne

26. August 2019

Bachelor-Seminar: Mobility and Traffic in Computer Networks

① Motivation / Fragestellungen

② FLAMeS-Framework

Phase I: Datensammlung & Preprocessing

Phase II: Analyse der erhobenen Daten

Phase III: Korrelationen und integrierte Modelle

③ Fazit / Ausblick

*„Analyzing Mobility-Traffic Correlations in Large WLAN Traces:
Flutes vs. Cellos“ [1]*

Isolierte Betrachtung

Aktuelle Modelle erfassen nicht das Zusammenspiel von Mobilität und Datenverkehr

Veraltete Daten

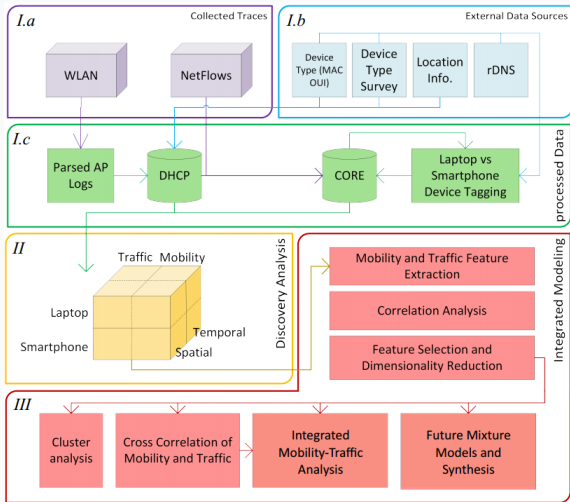
Trace-basierte Modelle verwenden i.d.R. Datensätze aus der Prä-Smartphone-Ära

- Wie unterscheiden sich Mobilitäts- und Datenverkehr-Charakteristiken zwischen unterschiedlichen Gerätetypen, Zeiten und Orten?
- Wie stehen diese Charakteristiken zueinander in Beziehung?
- Sollten neue Modelle entwickelt werden, die diese Unterschiede berücksichtigen?

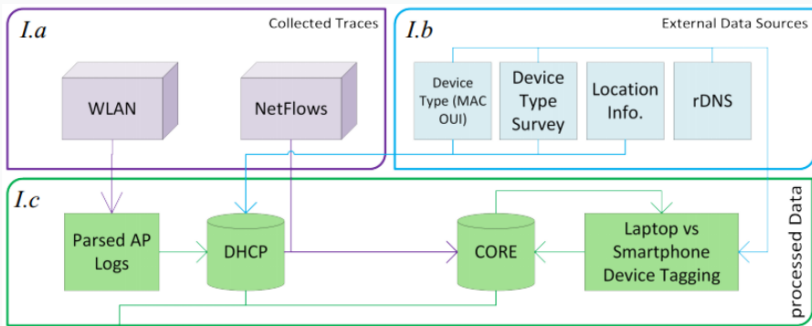
Intuitive Beispiele [3]



Datengetriebene Analysen (30TB von 300K Geräten)



Phase I: Datensammlung & Preprocessing



Phase I [1]

User IP	User MAC	AP name	AP MAC	Lease begin time	Lease end time
10.130.90.3	00:11:22:33:44:55	b422r143-win-1	00:1d:e5:8f:1b:30	1333238737	1333238741

Quelle 1: WLAN-AP-Logs [1]

Start time	Finish time	Duration	Source IP	Destination IP
1334332274.912	1334332276.576	1.664	173.194.37.7	10.15.225.126
Protocol	Source port	Destination port	Packet count	Flow size
TCP	80	60482	157	217708

Quelle 2: NetFlow-Logs [1]

Datenbasis:

- **Quelle 1 + Quelle 2** (*DHCP MAC-to-IP-Mapping*)
- Externe Informationen z.B. durch *rDNS* und *OUI*

Heuristik zur Geräteklassifizierung

- Hersteller mittels *OUI* identifizieren
- Kontakt zu *admob.com* prüfen

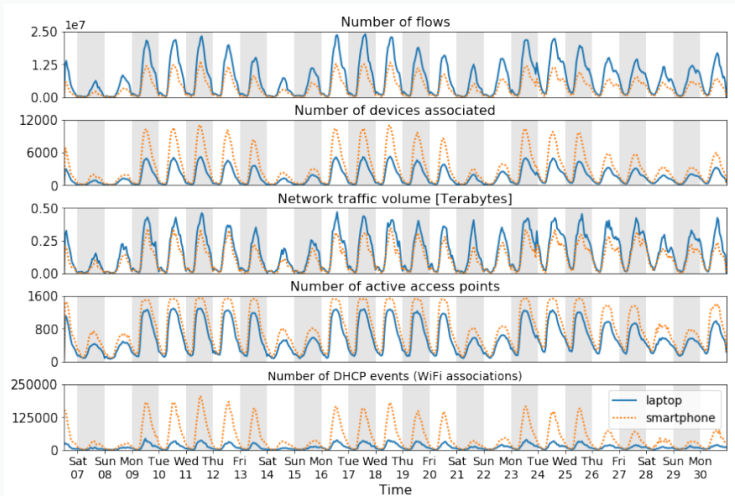
WLAN-Adresse

D0:2B:20:C7:44:84

Wireshark *OUI* Lookup Tool \Rightarrow Apple, Inc.

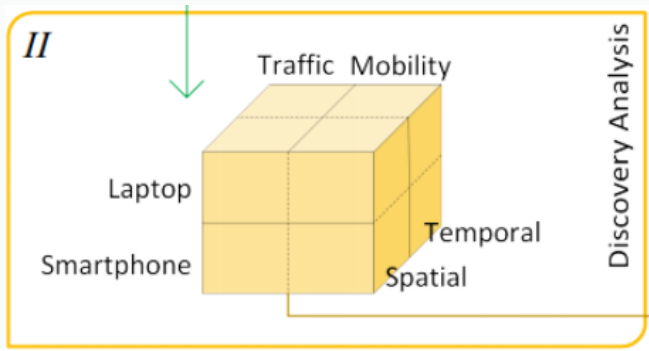
Ergebnis: 86% der Geräte in den AP-Logs und 97% der NetFlow-Traces klassifiziert

Phase I: Datensammlung & Preprocessing



Kombinierte WLAN-AP- und NetFlow-Traces [1]

Phase II: Analyse der erhobenen Daten



Phase II [1]

- WLAN-Sessions \approx Startzeiten von Vorlesungen
- Aktivität von Laptops fällt nach Ende der Vorlesungszeiten
- Abend-Sessions vermehrt in sozialen Einrichtungen / Bib.
- Vorlesungen geben Wochentagen Struktur
- Eingeschränkere Gesamtmobilität bei Laptops
- Smartphones „Always-on-Devices“ \implies leichter zu erfassen
- Laptops besitzen längere Aufenthaltszeiten

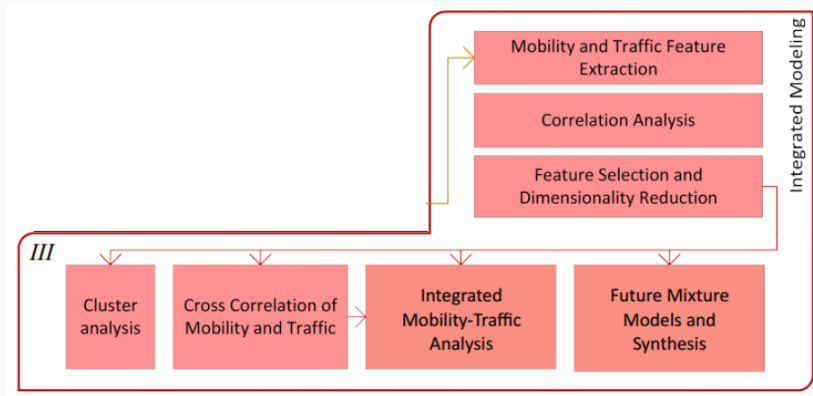
- Smartphone-Flows und Pakete größer
- Laptops verursachen \emptyset **3.7** (#Smartphone-Flows)
 \emptyset **1.6** (#Smartphone-Pakete)
 \implies \emptyset **2.7** (Smartphone-Traffic)
- Wochenenden: Verbleibende Geräte besonders aktiv
- Smartphones mehr extreme Phasen der Inaktivität
- Laptops (**78.5%** *TCP*), Smartphones (**98.2%** *TCP*)
- Großteil der APs an Wochenenden nicht verwendet

	Laptops	Smartphones
AP-Volumen (GB)	< 5	< 3
Datenkonsum (MB)	< 700	< 200
Aktivitätszeit (Std.)	< 3.5	< 1

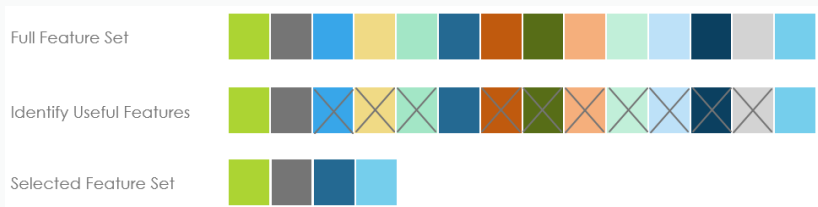
Smartphones:

- Bursty Traffic mit größeren Flows und kleinerer aktiver Dauer
- Für insgesamt deutlich weniger Last verantwortlich
- Besitzen hohe Scores bei Mobilitätsmetriken

Phase III: Korrelationen und integrierte Modelle



Phase III [1]

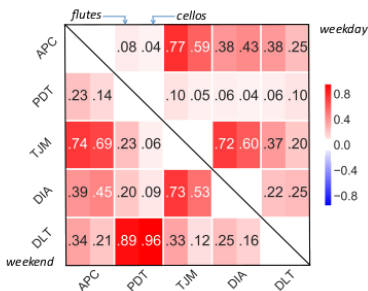


Feature-Selection Beispiel [4]

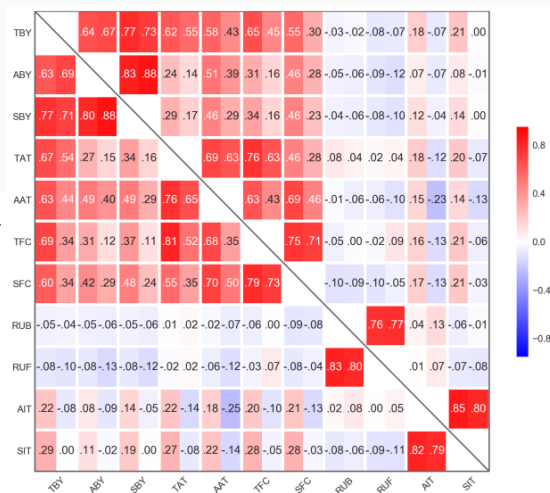
- Correlation Feature Selection (CFS)
- „Pearson-Correlation“-Methode $[-1, 1]$

Phase III: Isolierte Korrelationen

CFS selektiert **5/8** Mobilitäts- und **11/19** Traffic-Features

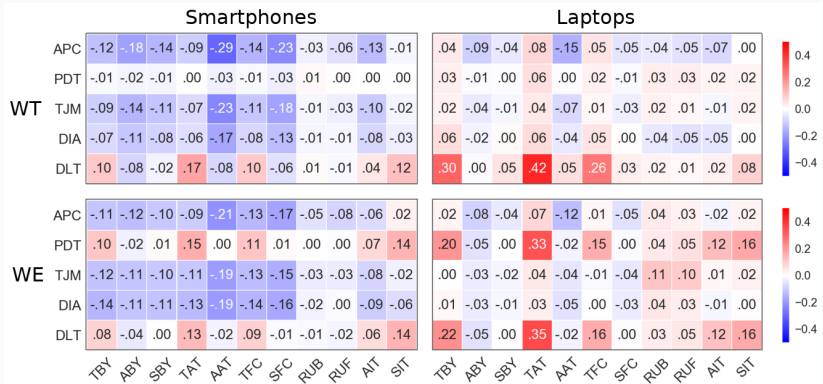


Mobilität [1]



Traffic [1]

Phase III: Korrelationen zwischen Mobilität und Datenverkehr



Korrelationen zwischen Mobilitäts- und Traffic-Features [1]

Unterschiede der Mobilitäts- und Traffic-Charakteristiken zwischen den Gerätetypen signifikant?

- **Supervised Classification (SVM):**
Kombiniert $\approx 81\%$ (Mobilität $\approx 65\%$, Traffic $\approx 79\%$)
- Genauigkeit erhöht sich auf $\approx 86\%$, wenn zwischen Wochenenden und Wochentagen differenziert wird
- **Unsupervised Clustering (k -Means):**
Kombiniert $\approx 81.5\%$ (Mobilität $\approx 60\%$, Traffic $\approx 81.2\%$)

- Generierung erster Traces: GMM mit kombinierten Features
- Vergleich der generierten Samples mit Echtdaten
- Samples bilden Verhalten beider Gerätetypen ab

Vergleich zu isolierten Modellen:

- Realistischer in Bezug auf Traffic-Features
- Keine Verbesserung in Bezug auf Mobilitäts-Features

- Signifikante Unterschiede zwischen den Gerätetypen
- Korrelationen zwischen beiden Dimensionen
- Kombinierte Modelle potenziell realistischer
- Anstoß für Entwicklung zukünftiger kombinierter Modelle
- Konzeption und Validierung konkreter Modelle wird zukünftiger Forschung überlassen

Nutzergruppen

- Mobilität und Datenverkehr nicht nur zwischen Gerätetypen unterschiedlich
- Alter und Geschlecht eines Nutzers berücksichtigen
- Kulturellen Kontext aufgezeichneter Daten berücksichtigen

„Practical Prediction of Human Movements Across Device Types and Spatiotemporal Granularities“ [2]

- Vorhersagbarkeit menschlicher Mobilität
- Bewegung von „sit-to-use“-Geräten besser vorhersagbar
- Signifikante Korrelationen zwischen Vorhersagegenauigkeit, Mobilitäts- und Traffic-Features
- Vorhersagegenauigkeit als Feature in integrierten Modellen
- Realistische Modellierung der Netzwerkaktivität ist aktives Forschungsfeld



B. Alipour, L. Tonetto, A.Y. Ding, R. Ketabi, J. Ott, and A. Helmy.

Flutes vs. cellos: Analyzing mobility-traffic correlations in large wlan traces.

pages 1637–1645, 04 2018.



B. Alipour, L. Tonetto, R. Ketabi, A. Y. Ding, J. Ott, and A. Helmy.

Practical prediction of human movements across device types and spatiotemporal granularities.

03 2019.



Dennis Shinault.

The surprising dangers of cell phone distractions:

Distracted walking.

<https://www.protectiveinsurance.com/blog/blog-post/blog/2016/04/21/the-surprising-dangers-of-cell-phone-distractions-distraction-driving>

Abgerufen am: 21.08.2019.



Mehul Ved.

Feature Selection and Feature Extraction in Machine Learning: An Overview.

<https://medium.com/@mehulved1503/feature-selection-and-feature-extraction-in-machine-learning>

Abgerufen am: 21.08.2019.