



UNIVERSITÀ
DI TRENTO

Estimating the number of people based on Wi-Fi probe request frames

Department of Information Engineering and Computer Science

Supervisors

Fabrizio Granelli

Daniele Miorandi

Student

Samuel Bortolin

Academic year 2019/2020

About this Project

Internship at U-Hopper:

Big Data Analytics

Business Intelligence

Chatbot

IoT solutions

Artificial Intelligence solutions





Problem Statement

Badly handled demand (the number of people who require the service)

can lead to overcrowding and inefficiency of the services

→ Inefficient and bad organized service leads to higher costs

→ Badly managed overcrowding (e.g. in supermarkets) during this global pandemic period due to COVID-19 leads to long queues and new infections



Research Statement

Is it possible to continuously estimate the number of people
based on the Wi-Fi probe request frames?



Thesis Objectives

- Capture and analysis of Wi-Fi probe request frames
- Data extraction, transmission and storage
- Analysis of Wi-Fi probe request frame patterns
- Provide an estimation of the number of customers



State of the Art



→ Analysis of different methods to count/estimate people

Infrared sensors, LSE, treadle switch-based systems, Video methods,
Audio methods, Wi-Fi, Bluetooth, BLE, LTE Radar, RFID approaches

→ Many fields of applicability and different implementations



Why Wi-Fi solution?



High diffusion of Wi-Fi devices,
high accuracy using machine learning,
low-cost implementation,
real-time data transmission,
user privacy ensured.



Standard 802.11 → Management
frames → Probe request frames



Results

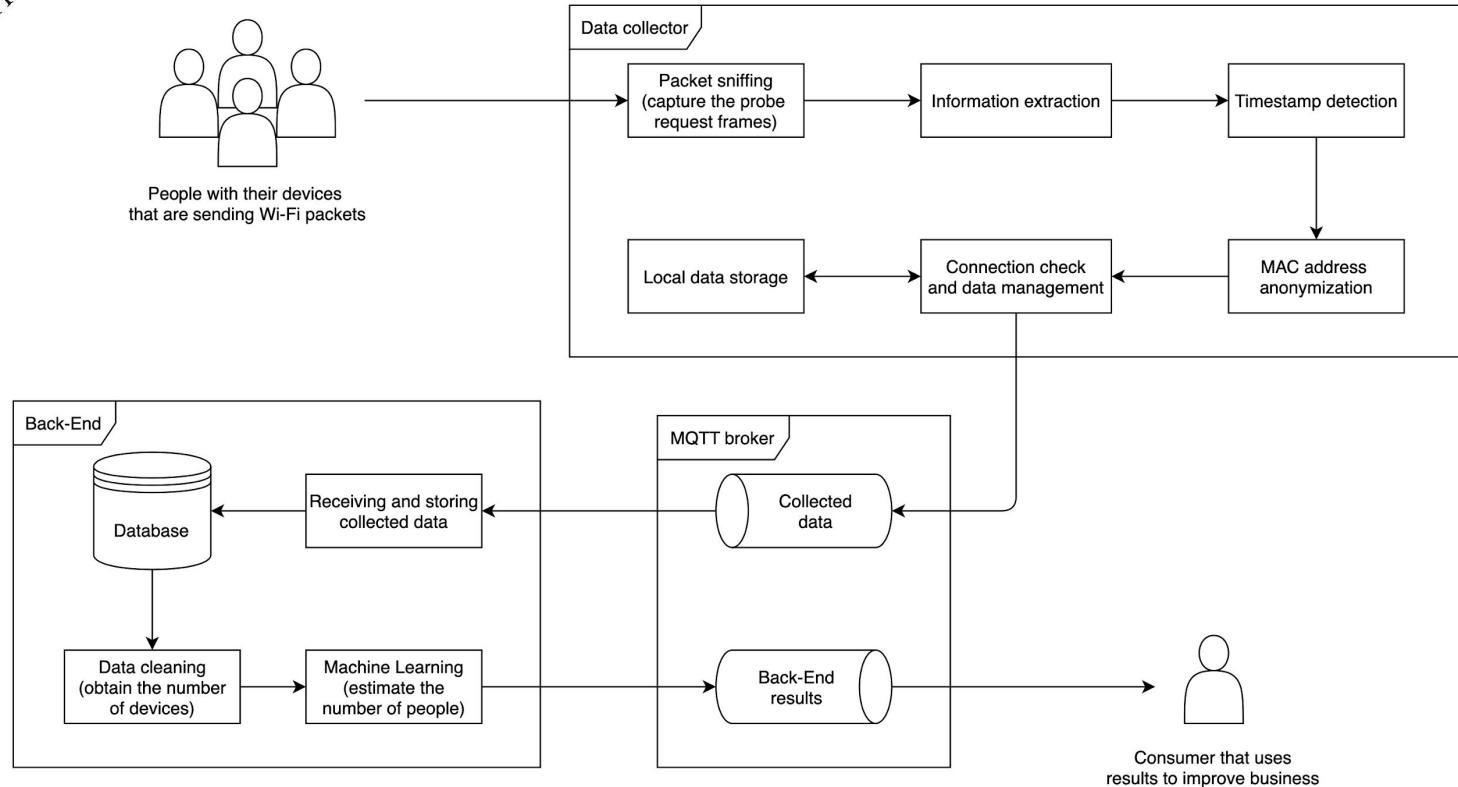


Designed and developed a system
for this problem that could work in
several context

Tested the system in a Cafe and
collected 4 weeks of data and
manually-annotated ground truth



System Architecture





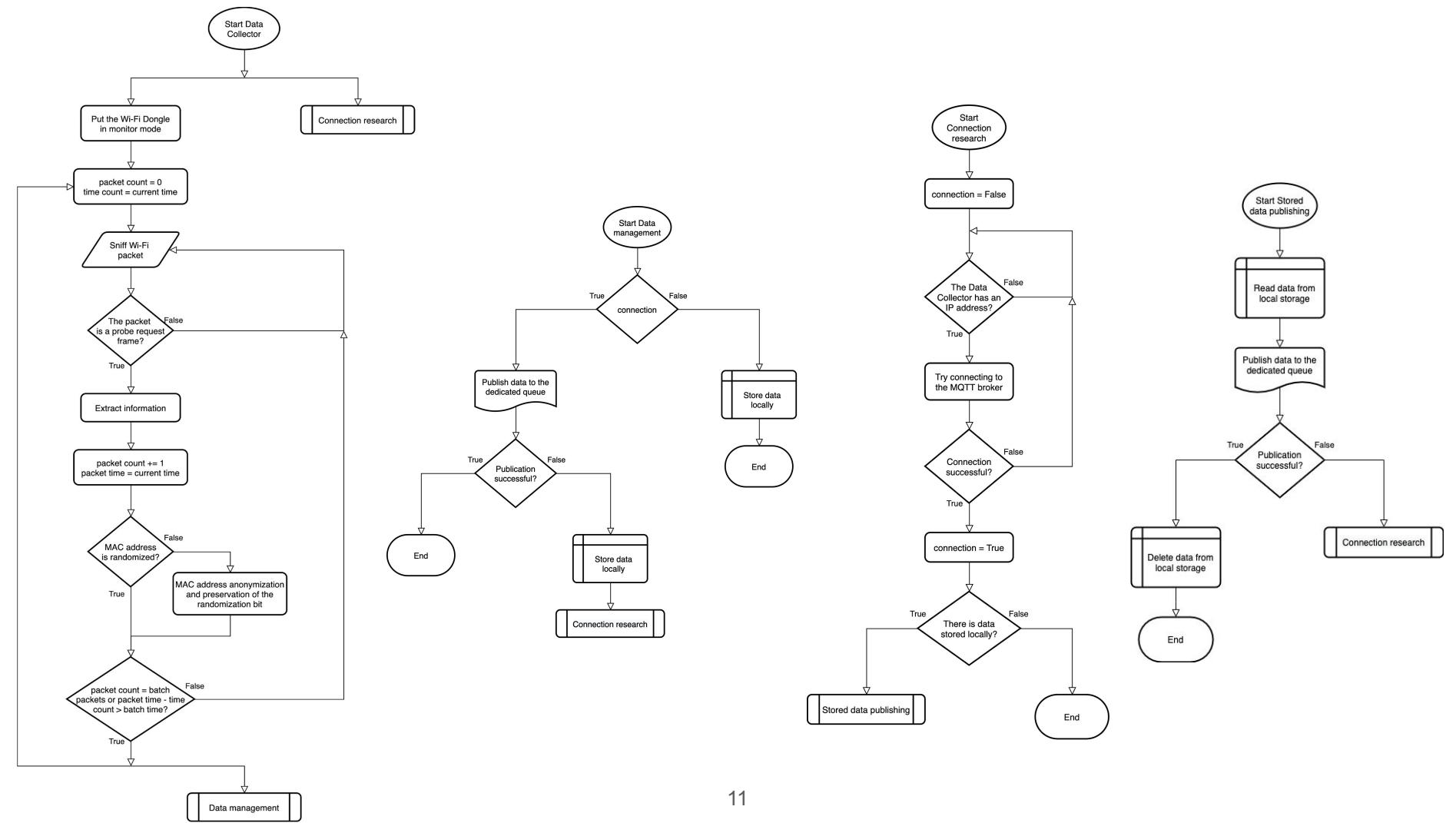
Data Collector Logic

Packet sniffing using Scapy → Information extraction

MAC address anonymization using BLAKE2s

Check connection → Local storage / MQTT transmission

Connect (clean_session = False) and login to the MQTT broker (username e password) → Publish batches to the dedicated queue (QoS = 2)





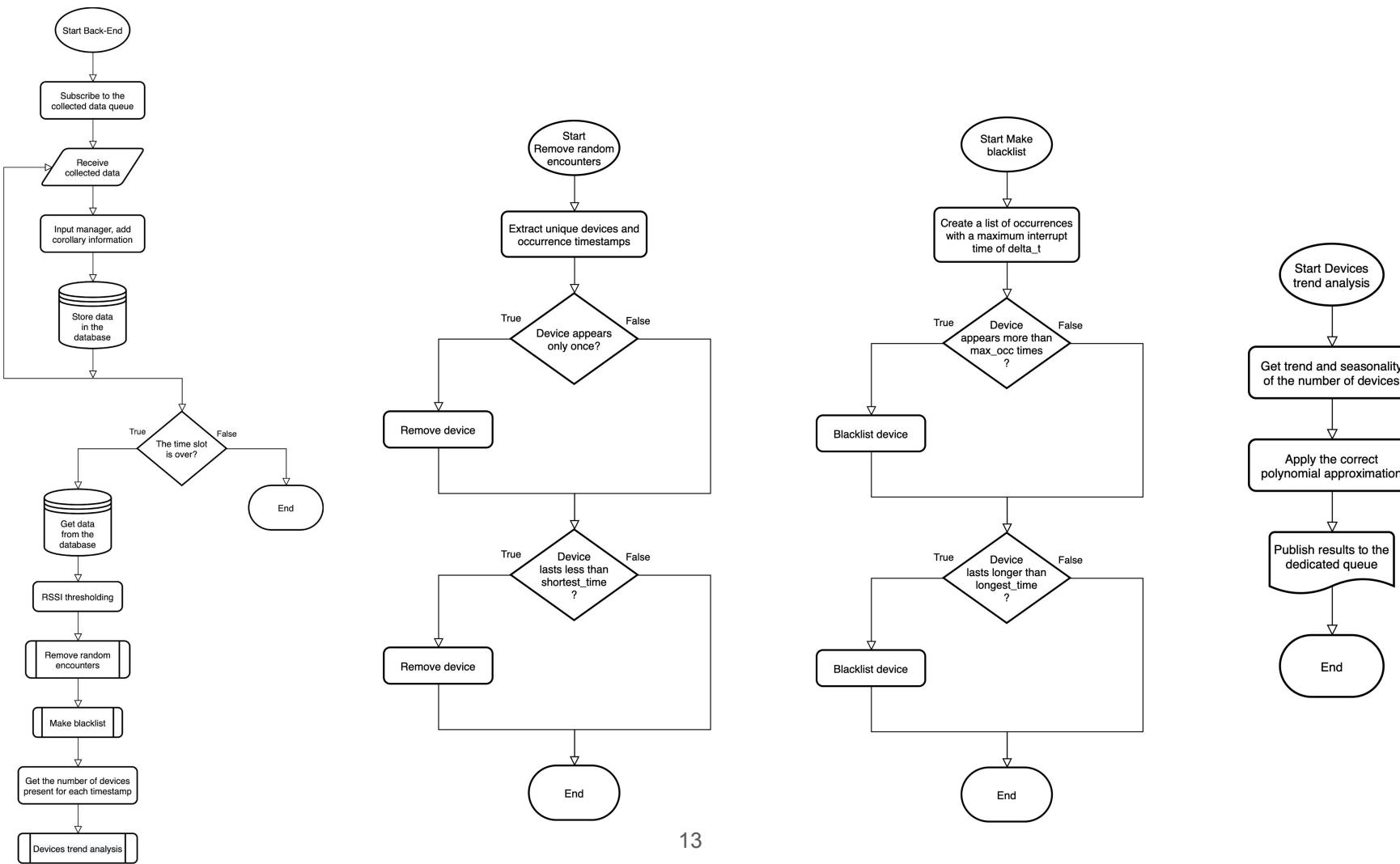
Back-End Logic

Publish to a topic → MQTT broker forwards the data to the subscribers

MQTT receiver in the Back-End → Subscription and storage in a database

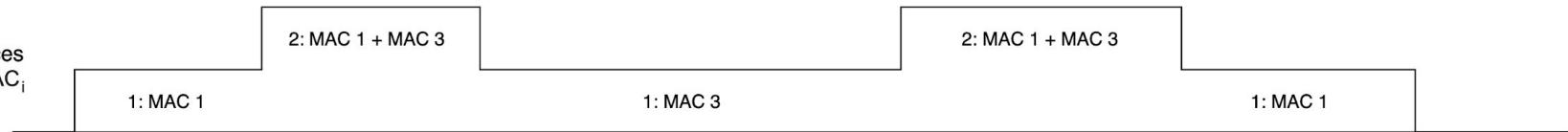
Data cleaning → RSSI threshold, random encounters, blacklist → # devices

Machine Learning algorithm → Fit degree and coefficients of polynomial approximations using trend and seasonality of the # devices → # people

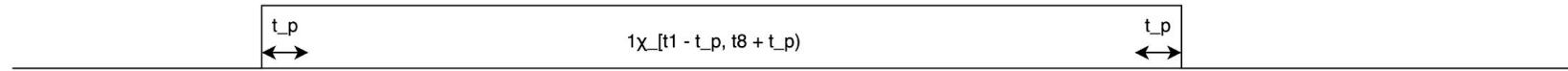


Presence of devices

Presence of devices
 $= \sum_i \text{presence MAC}_i$



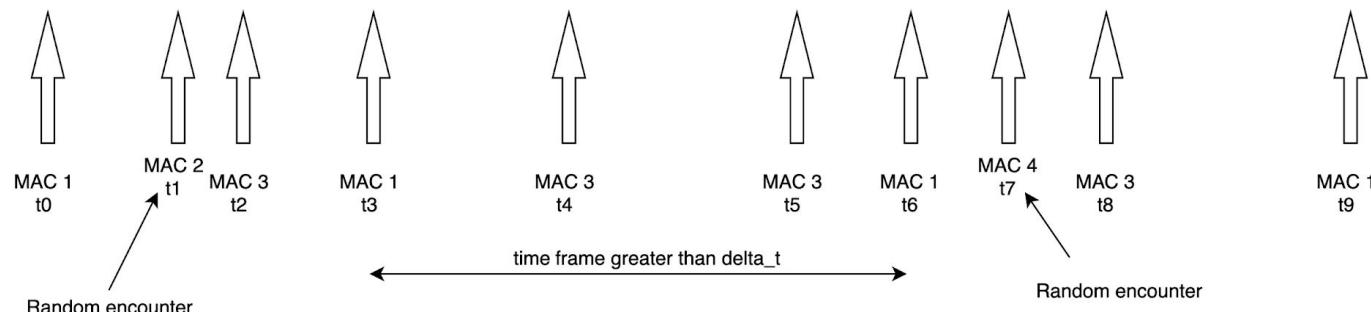
presence MAC_3



presence MAC_1



Probe
revelations





Implementation

Packet sniffing on Raspberry Pi → Eclipse Mosquitto MQTT broker

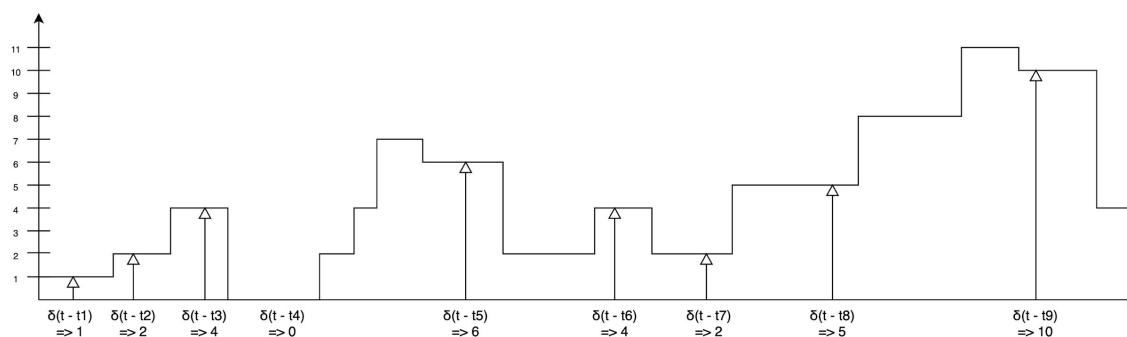
- Eclipse Paho MQTT client as receiver in the Back-End
- Storage in a MongoDB → Data cleaning → # device
- Ground truth (random look) → Machine Learning algorithm → # people



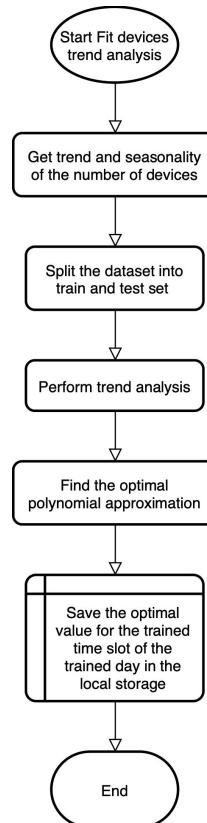
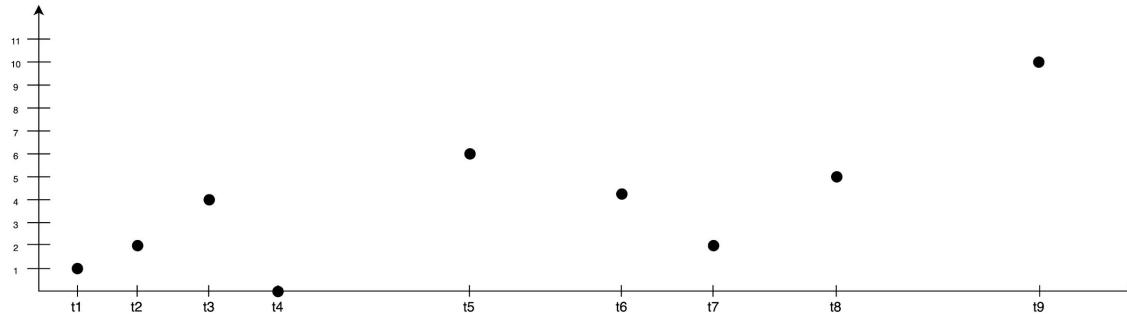
Ground Truth Collection



People present
 $g(t)$



Random look sampling of $g(t)$
=> Ground truth





Feasibility Test at Home

Devices	# probes	Day 1		Day 2		Day 3		Comments		
		RSSI range 1	RSSI range 2	# probes	RSSI range 1	RSSI range 2	# probes	RSSI range 1	RSSI range 2	
Wi-Fi gate	6410	-75 ± -81		4959	-65 ± -77		4819	-65 ± -75		omnipresent, send 2/4 probe every ~ 30 sec, static, ~ 6/7 m away
Smart TV	6	-83 ± -89		/			/			static in the living room, ~ 6/7 m away
PlayStation 4	13660	-71 ± -83		5	-71 ± -75		5	-75 ± -81		static in the living room, ~ 6/7 m away
iMac	294	-83 ± -91		197	-79 ± -87		59	-75 ± -85		static in my bedroom, ~ 10 m away
MacBook	32	-87 ± -91	-61 ± -67	/			/			range 1 → far, not in the kitchen; range 2 → nearby, in the kitchen
Mom's Samsung	38	-75 ± -91		3	-73 ± -79		/			far, not in the kitchen (Wi-Fi usually turned off)
Grandma's Samsung	/			/			39	-71 ± -81	-57 ± -65	range 1 → far, not in the kitchen; range 2 → nearby, in the kitchen
Thomas's Samsung	103	-75 ± -89	-57 ± -69	49	-71 ± -79	-51 ± -69	166	-73 ± -81	-47 ± -69	range 1 → far, not in the kitchen; range 2 → nearby, in the kitchen
Dad's iPhone	94	-75 ± -91	-55 ± -65	1547	-77 ± -83	-49 ± -67	1170	-69 ± -83	-35 ± -67	range 1 → far, not in the kitchen; range 2 → nearby, in the kitchen
Mattia's iPhone	1377	-77 ± -91	-59 ± -75	978	-75 ± -85	-49 ± -65	2051	-73 ± -85	-49 ± -67	range 1 → far, not in the kitchen; range 2 → nearby, in the kitchen
My iPhone	40	-79 ± -91	-65 ± -67	54	-73 ± -85	59	98	-73 ± -83	-53 ± -67	range 1 → far, not in the kitchen; range 2 → nearby, in the kitchen
Printer	/			1	-81		/			static in my bedroom, ~ 10 m away
Other Wi-Fi dongle	/			13084	-19 ± -23	-37 ± -63	12188	-19 ± -23	-37 ± -61	another Wi-Fi dongle, 1/2 probe every 6/7 sec → value swings sometimes
Samsung Galaxy J3	17	-85 ± -91		/			/			non-home device
Samsung Galaxy A20e	/			1	-83		/			non-home device
Randomized MACs	626			568			1190			probes with randomized MAC address, vague values of the RSSI
Total	22698			21445			21785			average of 22000 Wi-Fi probe request frames for ~ 18 hours

3 days of data collection, ~66.000 Wi-Fi PRF → detecting devices in the area



Validation

- Raspberry Pi in a Cafe where I annotate manually the ground truth
- Eclipse Mosquitto Broker MQTT of U-Hopper on their server
- MQTT receiver and MongoDB on U-Hopper server using Docker containers and using a volume for persisting data
- Analyzer/Estimator on my pc to use on the volume + collected ground truth to test accuracy and reliability of the proposed system



Results

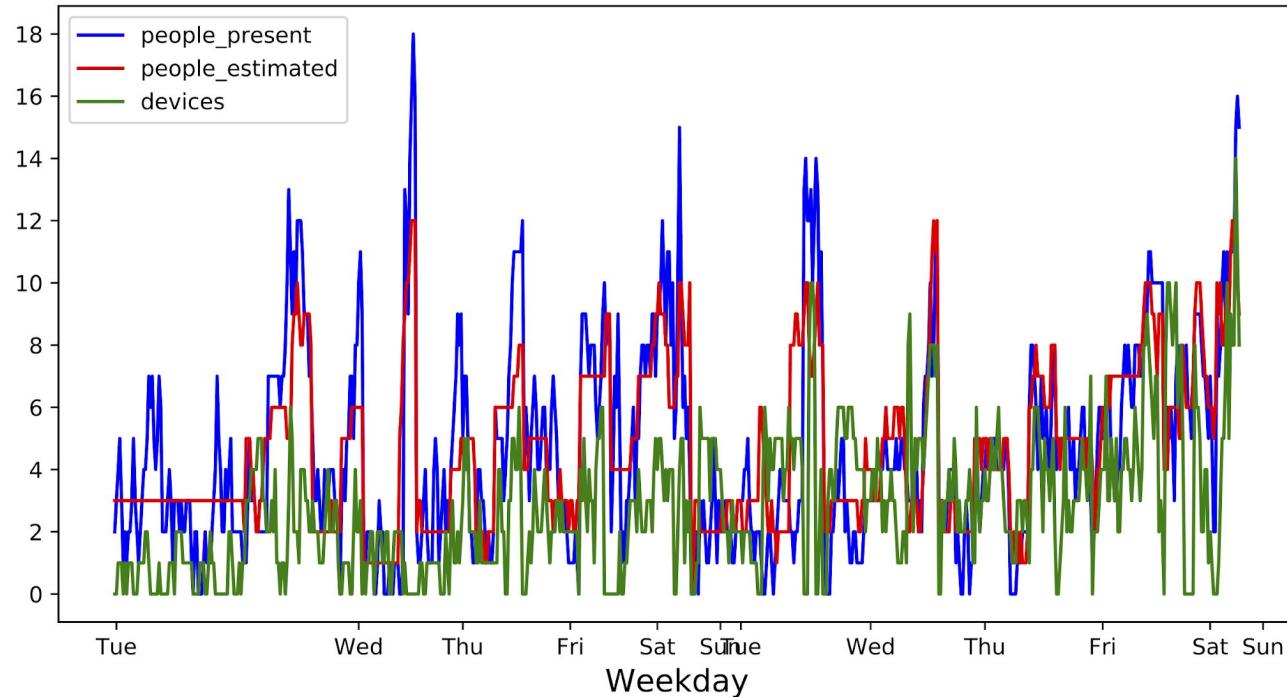


	2020-06-16	2020-06-17	2020-06-18	2020-06-19	2020-06-20	2020-06-21		Total1
Probe captured	24011	19329	22533	23041	19062	11238		119214
Total MACs	1489	873	1281	1307	1447	1187		7584
MACs only registered once	909	504	753	852	1002	634		4654
MACs lasted shorter than 20 seconds	443	280	395	344	342	447		2251
MACs occurred more than 10 times throughout the day	2	3	3	3	2	0		13
MACs lasted longer than 7200 seconds in any of it's occurrences	4	2	5	5	4	3		23
MACs remained	131	84	125	103	97	103		643
Manual annotations	116	61	63	51	37	12		340
							Mean Absolute Error	1.461
							Mean Squared Error	4.039
							Scaled_MAE_trend/count	0.448
							Scaled_MSE_trend/count	0.700

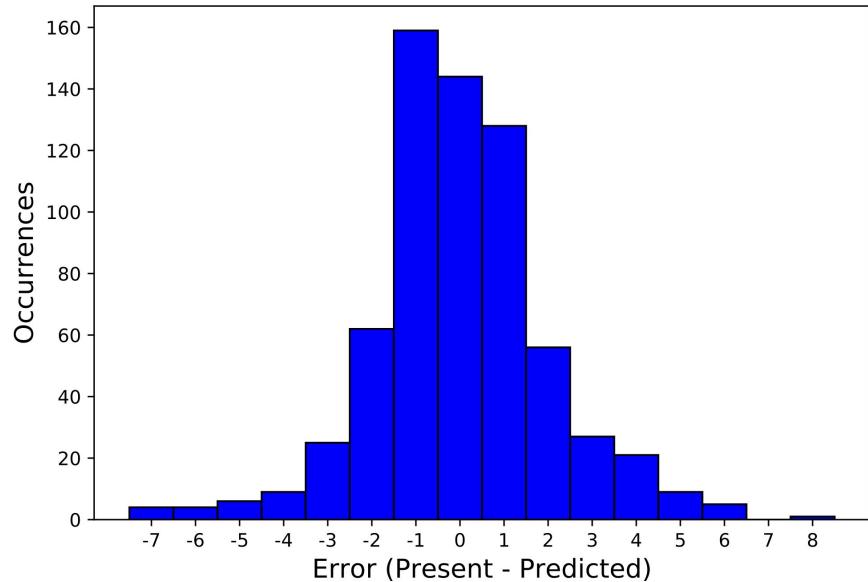
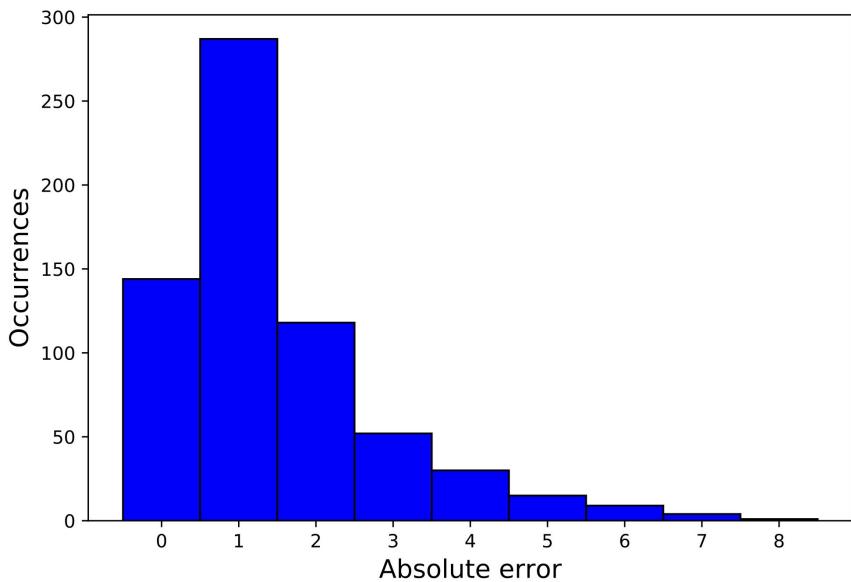
	2020-06-23	2020-06-24	2020-06-25	2020-06-26	2020-06-27	2020-06-28		Total2	Total
45730	56835	58203	61900	28941	19768			271377	390591
1579	2260	2293	2282	1182	715			10311	17895
333	490	509	462	242	106			2142	6796
1060	1568	1543	1577	805	517			7070	9321
0	2	1	2	0	0			5	18
6	5	6	8	3	3			31	54
180	195	234	233	132	89			1063	1706
76	67	69	63	31	14			320	660
							Mean Absolute Error	1.461	1.461
							Mean Squared Error	4.039	4.039
							Scaled_MAE_trend/count	0.448	0.448
							Scaled_MSE_trend/count	0.700	0.700



Results

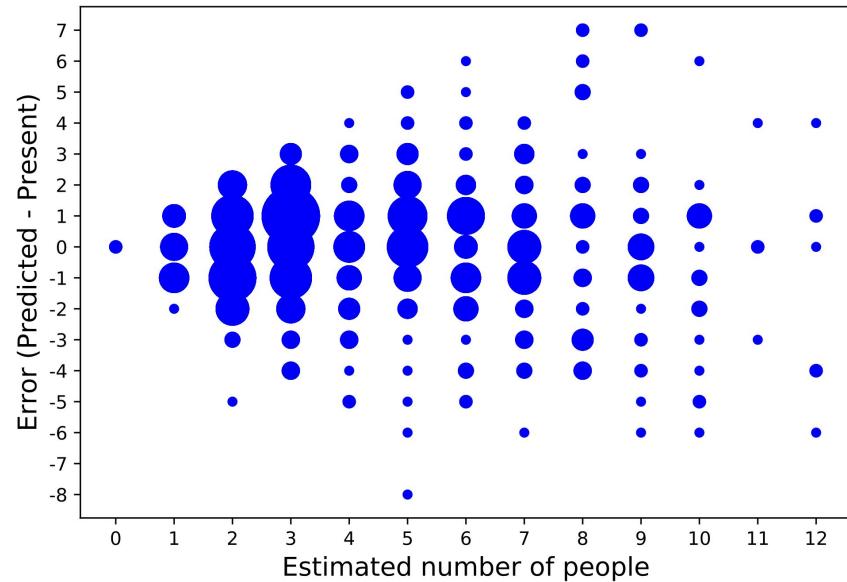
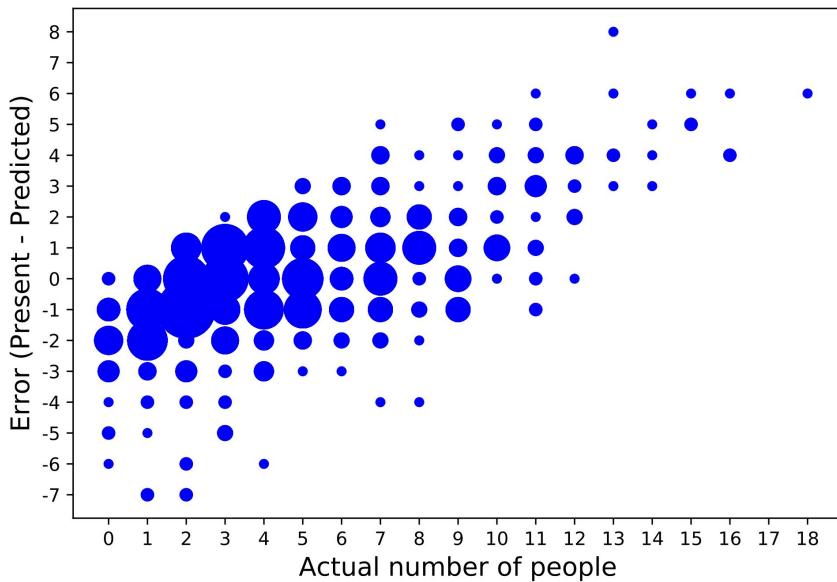


Results





Results





Summary

- It is possible to continuously estimate the number of people based on the Wi-Fi probe request frames
- Designed and developed a system to do that
- Tested the system in a Cafe and collected 4 weeks of data and manually-annotated ground truth



Future Works

- Real-Time integration
- Test the system in different context
- Improve the Machine Learning model