Learning On-Air Hand Gestures From Wi-Fi Signals on Smartphones

Deep Learning Seminar

Dr. Ramviyas Parasuraman

Assistant Professor, UGA Computer Science

Work: Master's thesis (2017) of Mohamed Haseeb at KTH Sweden

Slides courtesy: Mohamed from his thesis presentation.

November 8, 2019

Department of Computer Science
Franklin College of Arts and Sciences
UNIVERSITY OF GEORGIA

Heterogeneous Robotics Lab

Contents

The problem

- Smartphones are becoming more and more essential to humans.
- As of 2016, 3.9 billion smartphone subscriptions (expected to reach 6.8 billion in 2022) out of 7.5 billion mobile phone subscriptions in the world [1].

The problem

- Smartphones are becoming more and more essential to humans.
- As of 2016, 3.9 billion smartphone subscriptions (expected to reach 6.8 billion in 2022) out of 7.5 billion mobile phone subscriptions in the world [1].
- Yet, interaction with smartphones is largely bound to their screens (limited by screen size, battery power and computation capability).

The problem

- Quest for intuitive ways to interact with smartphones; examples: speech recognition, gesture recognition.

The problem

- Quest for intuitive ways to interact with smartphones; examples: speech recognition, gesture recognition.
- Based on the sensing mechanism, gesture recognition systems can be grouped into:

The problem

- Quest for intuitive ways to interact with smartphones; examples: speech recognition, gesture recognition.
- Based on the sensing mechanism, gesture recognition systems can be grouped into:
  - **Camera based** systems [2]. Limited camera field of view, sensitive to lighting conditions and consume high power.

The problem

- Quest for intuitive ways to interact with smartphones; examples: speech recognition, gesture recognition.
- Based on the sensing mechanism, gesture recognition systems can be grouped into:
  - **Camera based** systems [2]. Limited camera field of view, sensitive to lighting conditions and consume high power.
  - **Inertia based** systems [3]. Sensors (e.g. accelerometers, gyroscopes) have to be carried by users.

The problem

- Quest for intuitive ways to interact with smartphones; examples: speech recognition, gesture recognition.
- Based on the sensing mechanism, gesture recognition systems can be grouped into:
  - **Camera based** systems [2]. Limited camera field of view, sensitive to lighting conditions and consume high power.
  - **Inertia based** systems [3]. Sensors (e.g. accelerometers, gyroscopes) have to be carried by users.
  - **Radio Frequency (RF) based** systems.

RF based gesture recognition

Some approaches tries to introduce new HW into the smartphones:

- Google's Soli project [4].
- Specialized gesture recognition radar chip.

RF based gesture recognition

Some approaches tries to introduce new HW into the smartphones:

- Google's Soli project [4].
- Specialized gesture recognition radar chip.

Other approaches leverage the existing phone capabilities:

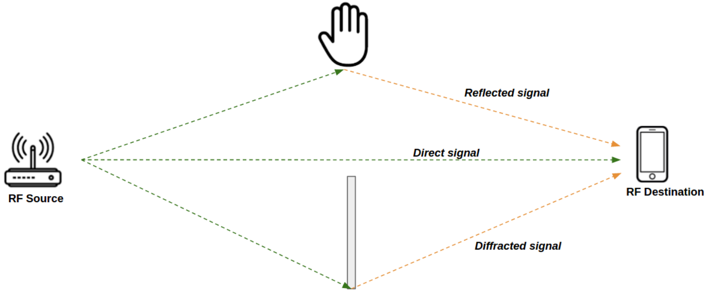
- Sense activity and gestures using FM, GSM/WCDM/LTE or Wi-Fi signals [5], [6], [7] and [8].

RF based gesture recognition

Advantages:

- Require no line of sight between the gesture subject and the smartphone.
- Consume less power.
- Ubiquitous.

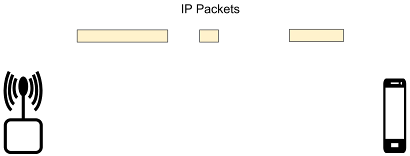
Radio wave propagation



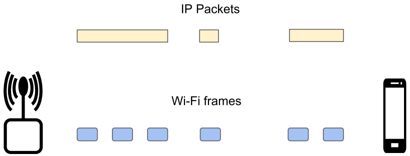
Static or moving objects (e.g. a human hand) impact the signal power at the receiving end.

Wi-Fi RSSI

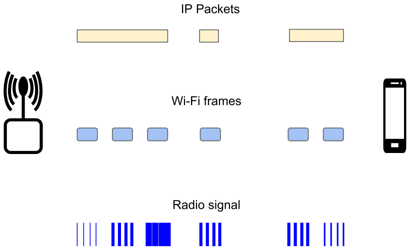
Wi-Fi RSSI



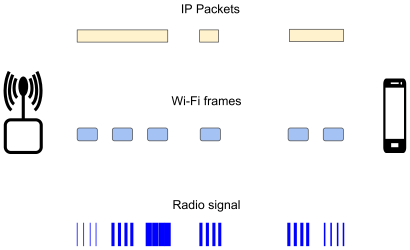
Wi-Fi RSSI



Wi-Fi RSSI

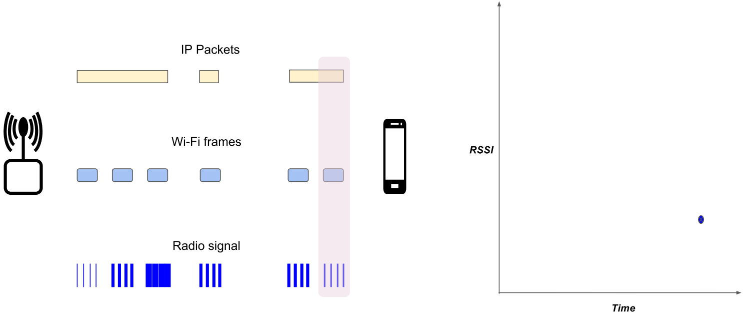


Wi-Fi RSSI



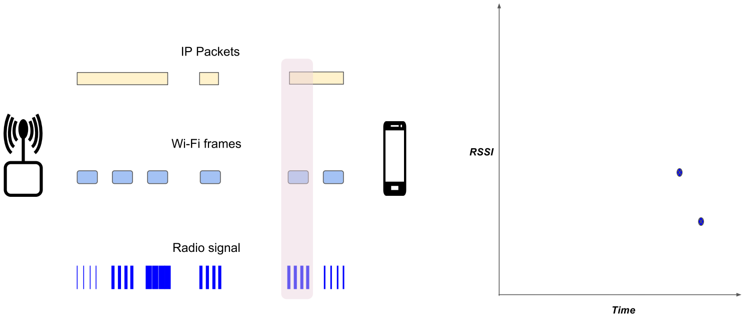
For every received frame, a measurement proportional to the Radio signal strength (aka RSSI) is made

Wi-Fi RSSI



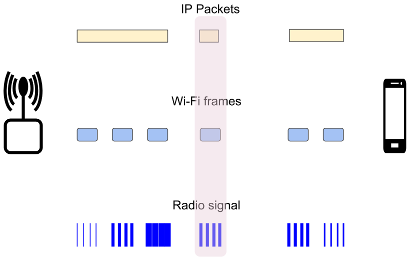
For every received frame, a measurement proportional to the Radio signal strength (aka RSSI) is made

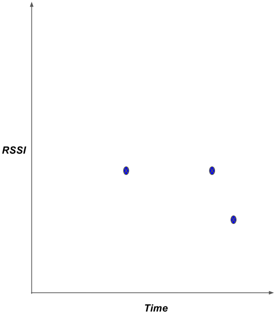
Wi-Fi RSSI



For every received frame, a measurement proportional to the Radio signal strength (aka RSSI) is made

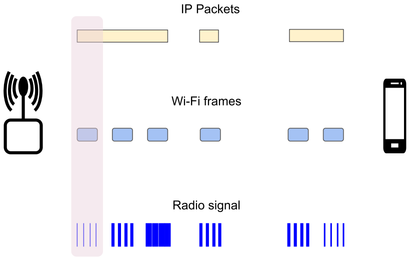
Wi-Fi RSSI

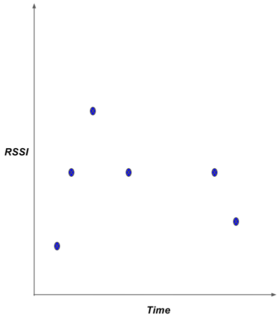




For every received frame, a measurement proportional to the Radio signal strength (aka RSSI) is made

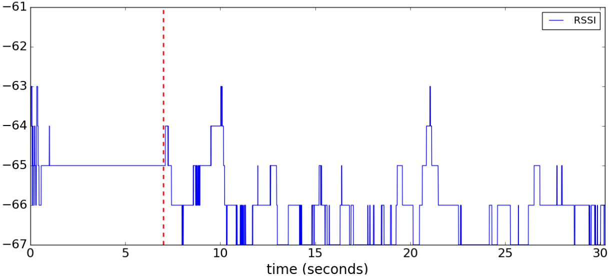
Wi-Fi RSSI



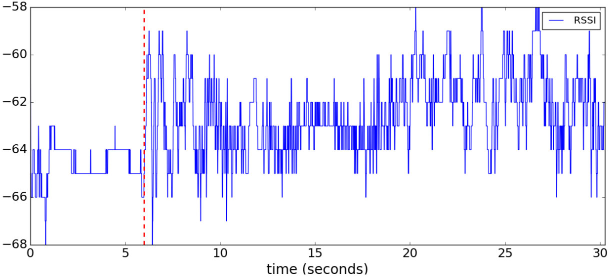


For every received frame, a measurement proportional to the Radio signal strength (aka RSSI) is made

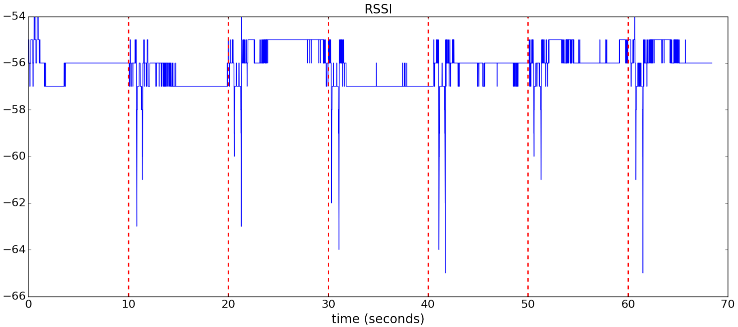
Sample RSSI measurements (typing in a keyboard)



Sample RSSI measurements (walking)

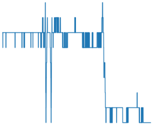


Sample RSSI measurements (performing Swipe gesture)



Hand gesture recognition from Wi-Fi RSSI

Input: RSSI stream

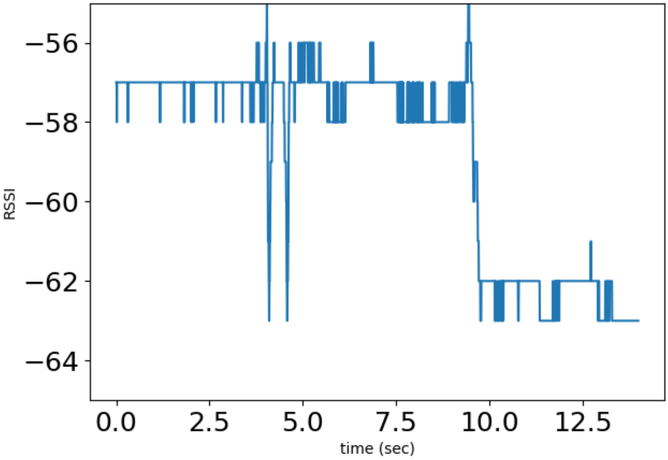


Gesture recognition solution

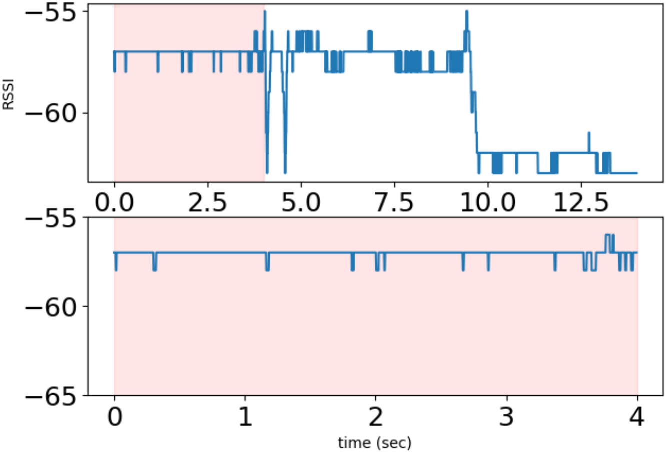
Output: Gesture predictions

Noise, Swipe, Noise, Push, ...

Hand gesture recognition from Wi-Fi RSSI

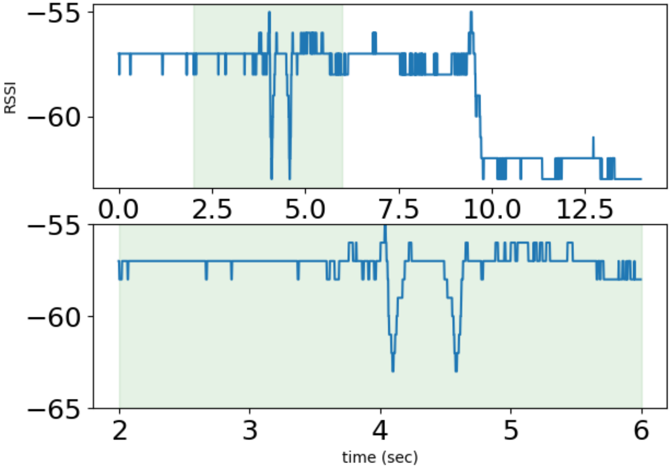


Hand gesture recognition from Wi-Fi RSSI



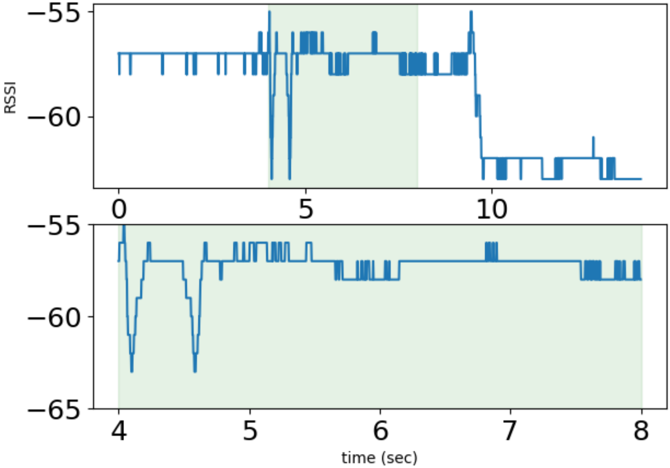
Prediction: no gesture (or Noise)

Hand gesture recognition from Wi-Fi RSSI



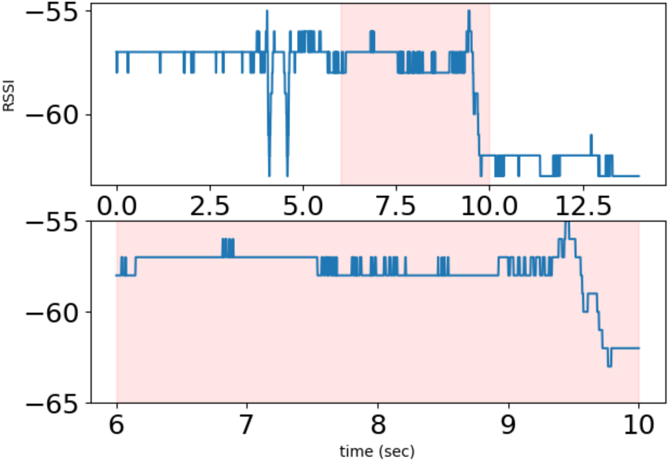
Prediction: Swipe

Hand gesture recognition from Wi-Fi RSS



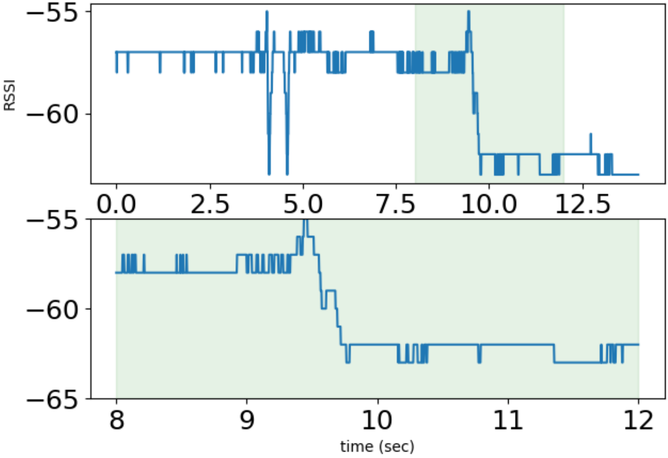
Prediction: Swipe

Hand gesture recognition from Wi-Fi RSSI



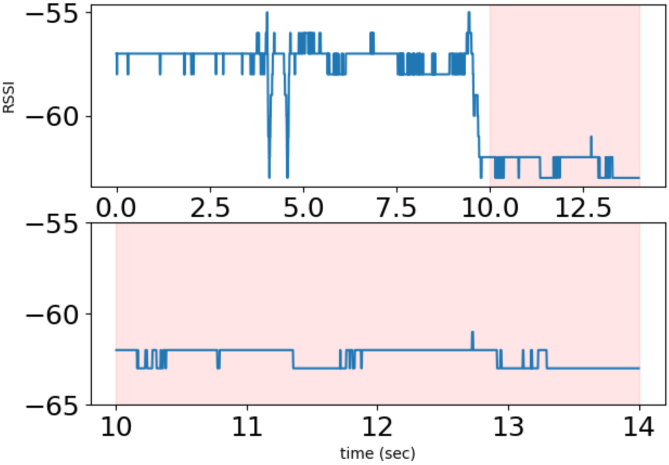
Prediction: Noise

Hand gesture recognition from Wi-Fi RSSI



Prediction: Push

Hand gesture recognition from Wi-Fi RSSI



Prediction: Noise

Challenges

Challenges

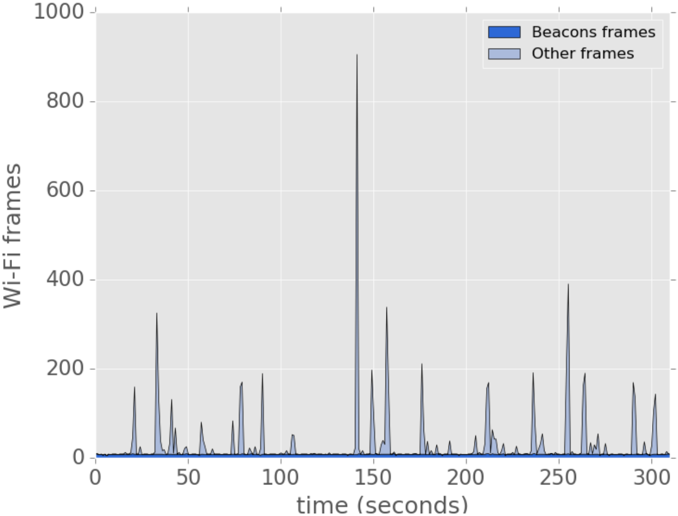
- Hand gestures and other background activities (e.g. walking) have closely similar impacts on the Wi-Fi signal.

Challenges

- Hand gestures and other background activities (e.g. walking) have closely similar impacts on the Wi-Fi signal.
- Wi-Fi RSSI stream is bursty (occurs in short non regular episodes).

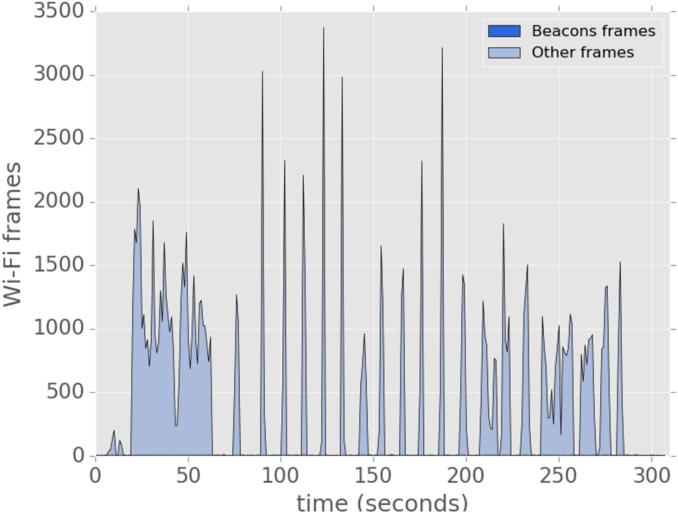
Wi-Fi RSSI stream is bursty

Wi-Fi frames received by a smartphone while **browsing Facebook**.



Wi-Fi RSSI stream is bursty

Wi-Fi frames received by a smartphone while **playing a Youtube video**.



Previous work

Previous work

- Wi-Fi RSSI was used to recognize activities (e.g. walking) on smartphones [9].

Previous work

- Wi-Fi RSSI was used to recognize activities (e.g. walking) on smartphones [9].
- It was also used to recognize moving hand gestures on smartphones [8], [10] and [11].

Previous work

- Wi-Fi RSSI was used to recognize activities (e.g. walking) on smartphones [9].
- It was also used to recognize moving hand gestures on smartphones [8], [10] and [11].
- But, to gain access to enough RSSI samples:
  - The Wi-Fi interface has to operate on the **monitor mode** (which **prevents other applications from using the Wi-Fi interface**).
  - A rooted Android OS was needed, to install a special Wi-Fi firmware.
  - Supported by a limited subset of the Wi-Fi devices.

Objectives

Project objectives

Demonstrate the possibility to recognize **dynamic** hand gestures on smartphones from the Wi-Fi RSSI stream, **without modification**, in a **passive online** setting.

- **dynamic**: involves hand movement.
- **passive**: leverages existing Wi-Fi sources.
- **online**: in realtime on the smartphone.
- **without modification**: without requiring additional HW, or core SW modification.

Proposed solution

Proposed solution

Core ideas:

Proposed solution

Core ideas:

- Induce Wi-Fi traffic between the AP and the smartphone to make enough RSSI measurements.

Proposed solution

Core ideas:

- Induce Wi-Fi traffic between the AP and the smartphone to make enough RSSI measurements.
- Use an LSTM RNN model:

Proposed solution

Core ideas:

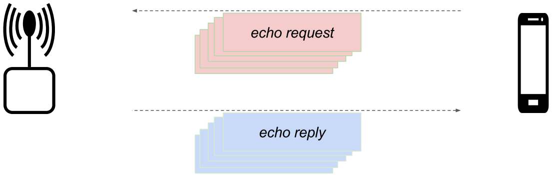
- Induce Wi-Fi traffic between the AP and the smartphone to make enough RSSI measurements.
- Use an LSTM RNN model:
  - Suitable for sequential inputs (e.g. audio and video signals).

Proposed solution

Core ideas:

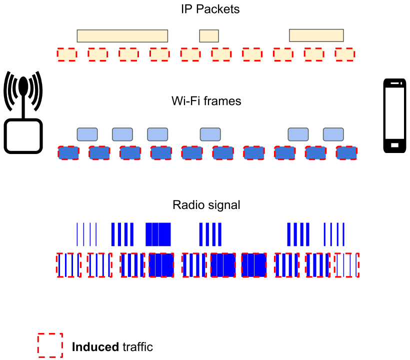
- Induce Wi-Fi traffic between the AP and the smartphone to make enough RSSI measurements.
- Use an LSTM RNN model:
  - Suitable for sequential inputs (e.g. audio and video signals).
- Suitable preprocessing of the input Wi-Fi RSSI stream.

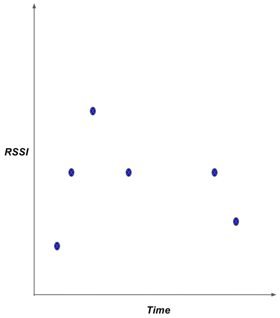
Traffic induction



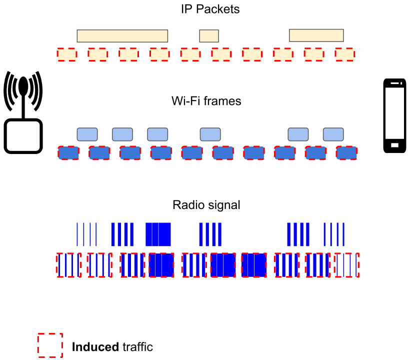
A smartphone continuously sending ICMP ***echo request*** messages to a Wi-Fi Access Point (and receiving the corresponding ***echo reply*** messages)

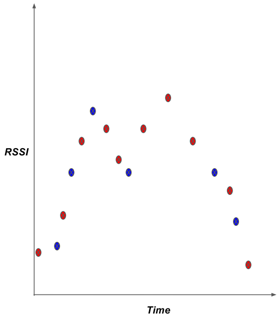
Traffic induction



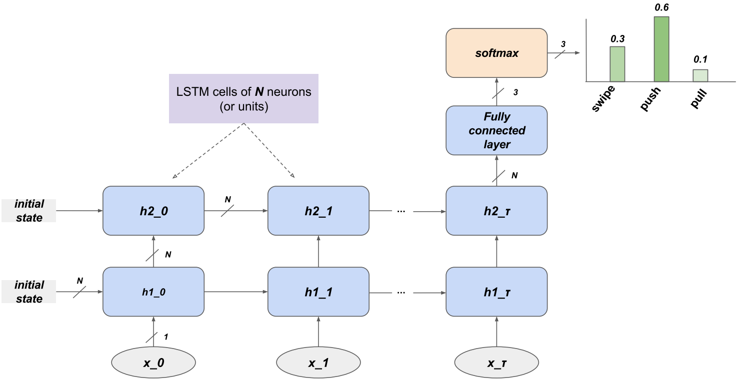


Traffic induction



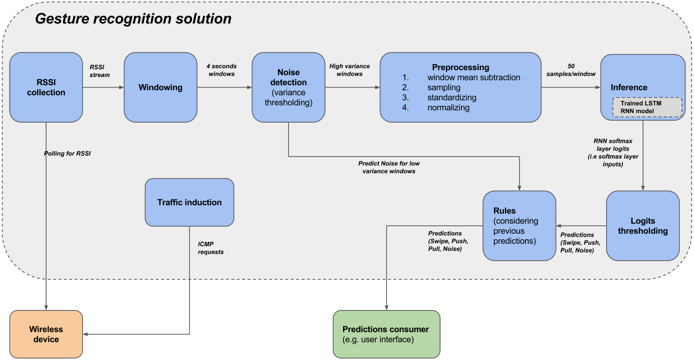


LSTM RNN model



- $N = 200$ neurons/LSTM cell
- $T = 50$ (RNN time steps)

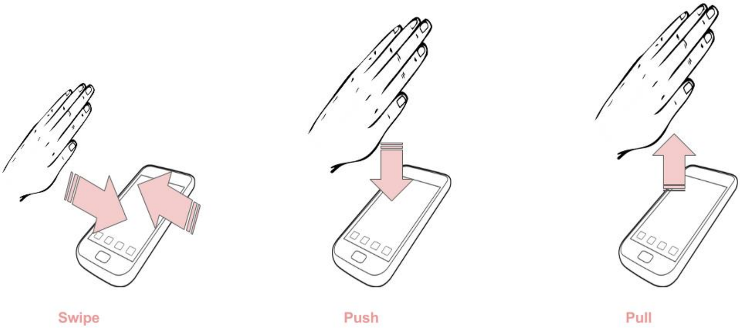
Recognition system diagram



Experiments

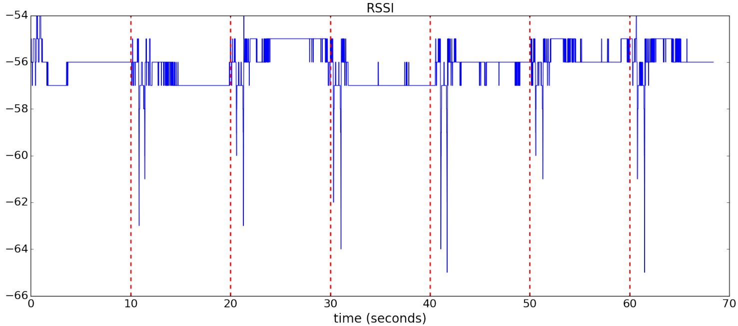
Collected dataset

Performed hand gestures:



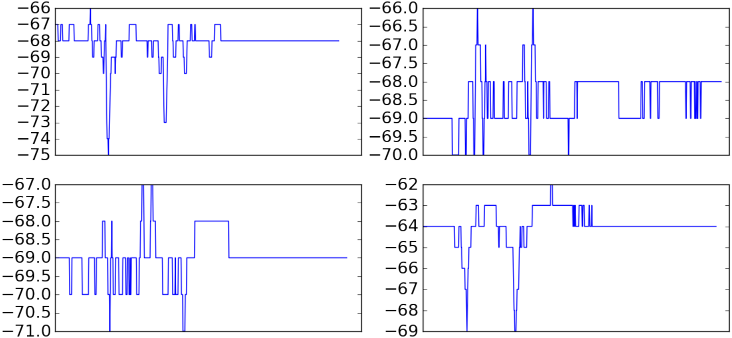
Collected dataset

Swipe samples collection session:



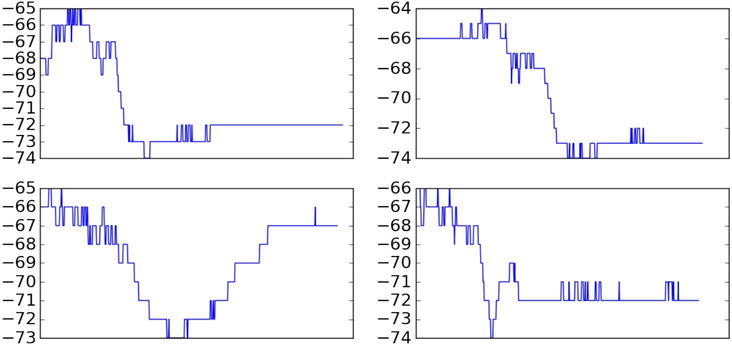
Collected dataset

Sample Swipe gestures:



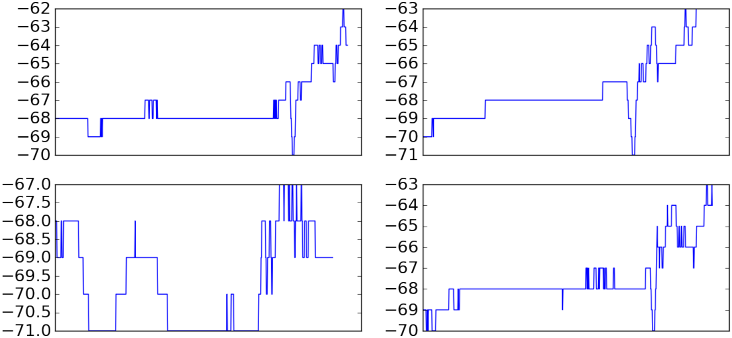
Collected dataset

Sample Push gestures:



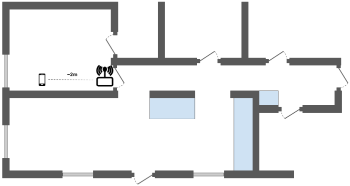
Collected dataset

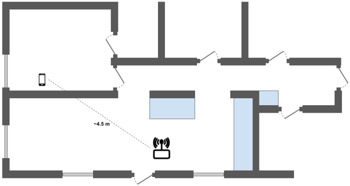
Sample Pull gestures:



Collected dataset

Spatial setup:





Collected dataset

- Summary of the collected dataset:

Dataset	1	2	3	4
Location	room	room	room	two rooms
Induction	√	√		√
Internet		√		
Size	440	432	434	337

- An Android app was developed to collect the dataset.

Offline experiments

Training and Evaluation of the LSTM RNN model was conducted on a Laptop (hence offline), using the collected dataset.

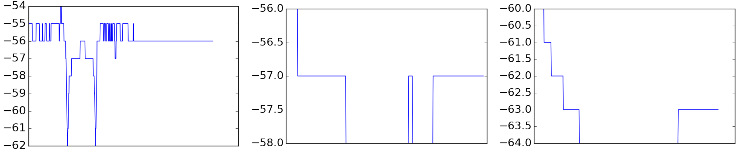
Offline experiments

Traffic induction impact on prediction accuracy.

Dataset	1	2	3	4
Location	room	room	room	two rooms
Induction	$\surd$	$\surd$		$\surd$
Internet		$\surd$		
LSTM accuracy	91%	83%	78%	87%

Offline experiments

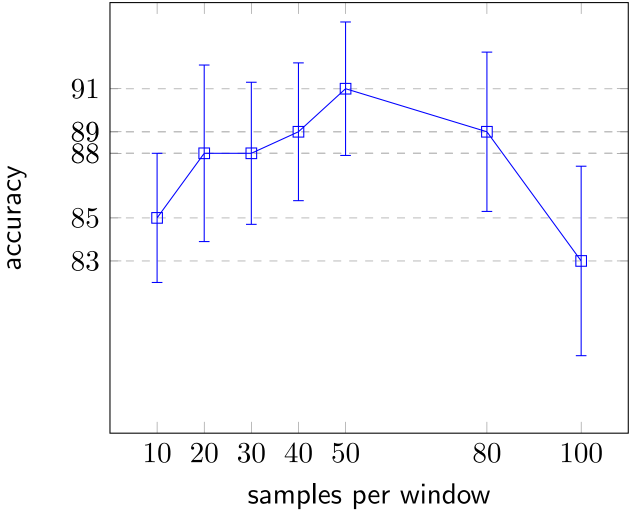
Swipe and Push gestures are hardly distinguishable when induction is OFF.



Left: Swipe gesture with induction ON. Middle and Right: Swipe and Push gestures respectively performed while no traffic is OFF.

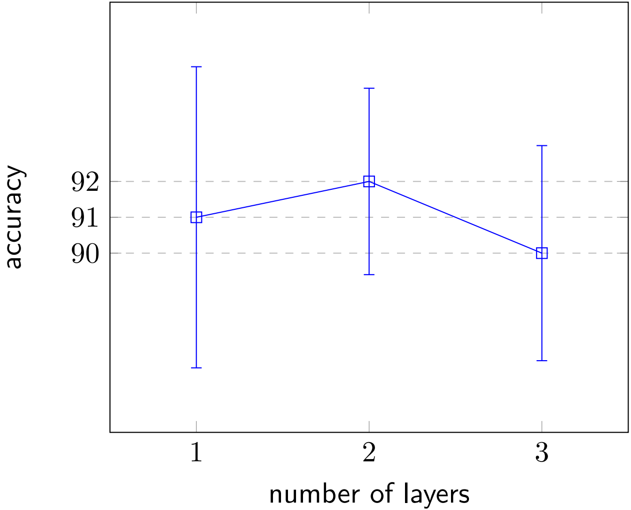
Offline experiments

Prediction accuracy as a function of the number of samples per prediction window.



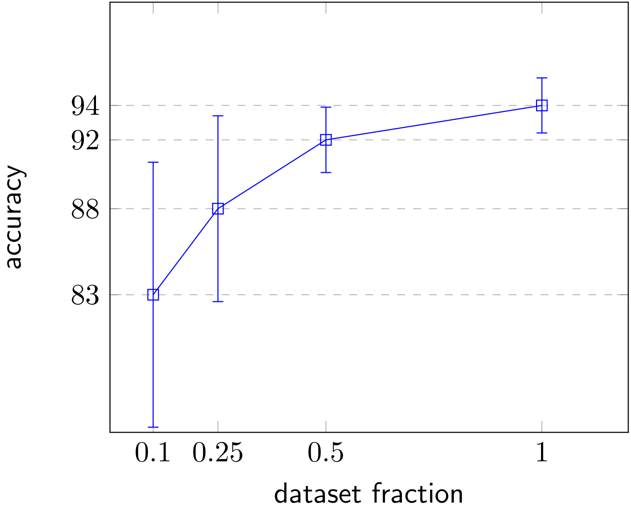
Offline experiments

Prediction accuracy as a function of the number of the hidden (LSTM) layers.



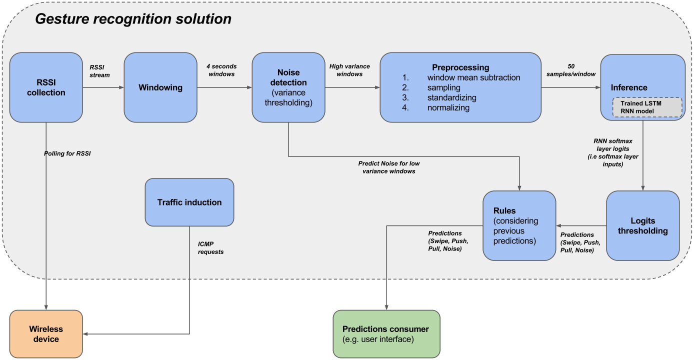
Offline experiments

The LSTM RNN model accuracy when trained with fractions of (Dataset1 + Dataset2 + Dataset4).

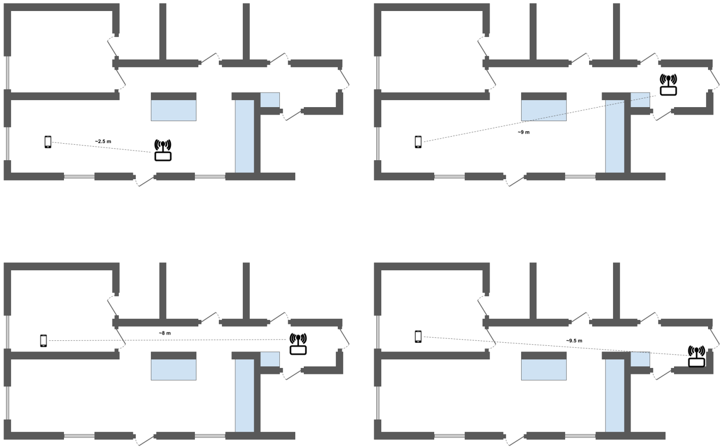


Online experiments

Evaluating the **full solution implementation (an Android app)** on a smartphone.

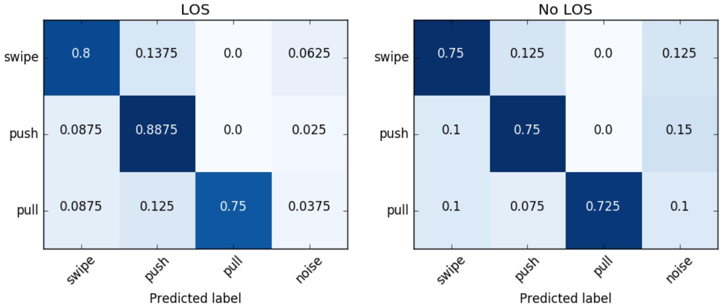


Spatial setup



Online experiments

- Accuracy of Line-of-sight (LOS) experiments is $\sim 81\%$.
- Accuracy of no Line-of-sight (No LOS) experiments is $\sim 74\%$.
- The Overall accuracy is $\sim 78\%$.



Online experiments

When no hand gesture is performed over a period of thirty minutes, the **false positivie rate was** $\sim 8\%$.

Gesture	number of predictions (%)
Noise (correct prediction)	1652 (92.1%)
Swipe (False positive)	61 (3.4%)
Push (False positive)	62 (3.5%)
Swipe (False positive)	18 (1.0%)

Demo

Click here - Takes you to YouTube!

Discussion

Conclusion

- We demonstrated its possible to predict hand gestures on unmodified smartphones from Wi-Fi RSSI.

Conclusion

- We demonstrated its possible to predict hand gestures on unmodified smartphones from Wi-Fi RSSI.
- The recognition accuracy can be improved by collecting more data, and increasing the model size.

Conclusion

- We demonstrated its possible to predict hand gestures on unmodified smartphones from Wi-Fi RSSI.
- The recognition accuracy can be improved by collecting more data, and increasing the model size.
- The recognition accuracy can be improved by sampling RSSI at higher frequency.

Limitations

- Vulnerability to interference from background activities.

Limitations

- Vulnerability to interference from background activities.
- High CPU usage (25%).

Publication

Published in IEEE Sensors (2019).
Citation: **Haseeb MA, Parasuraman R. Wisture: Touch-Less Hand Gesture Classification in Unmodified Smartphones Using Wi-Fi Signals. IEEE Sensors Journal. 2018 Oct 16;19(1):257-67.**

Publication

We open-sourced most of the codes and dataset (data collection and Wisture recognition).
https://github.com/mohaseeb/wisture
https://www.ieee-dataport.org/documents/wi-fi-signal-strength-measurements-smartphone-various-hand-gestures

Future work

Introducing a preamble gesture detection mode:

- Preamble gesture needs to be:

Future work

Introducing a preamble gesture detection mode:

- Preamble gesture needs to be: **hard to confuse with noise**

Future work

Introducing a preamble gesture detection mode:

- Preamble gesture needs to be: **hard to confuse with noise** and **require small power to detect.**

Future work

Introducing a preamble gesture detection mode:

- Preamble gesture needs to be: **hard to confuse with noise** and **require small power to detect.**
- Push gesture is a candidate; easy to recognize without induction.

Future work

Introducing a preamble gesture detection mode:

- Preamble gesture needs to be: **hard to confuse with noise** and **require small power to detect.**
- Push gesture is a candidate; easy to recognize without induction.

Benefits:

Future work

Introducing a preamble gesture detection mode:

- Preamble gesture needs to be: **hard to confuse with noise** and **require small power to detect.**
- Push gesture is a candidate; easy to recognize without induction.

Benefits:

- Increased robustness against interference.

Future work

Introducing a preamble gesture detection mode:

- Preamble gesture needs to be: **hard to confuse with noise** and **require small power to detect.**
- Push gesture is a candidate; easy to recognize without induction.

Benefits:

- Increased robustness against interference.
- Reduced power consumption.

Future work

Native support on Wi-Fi devices for a **cheap** high frequency sampling of Wi-Fi RSS.

Future work

Native support on Wi-Fi devices for a **cheap** high frequency sampling of Wi-Fi RSS.

- By inducing traffic at the Wi-Fi device level, **the OS is bypassed**, which results in a **higher throughput** at a **reduced power level**.

Future work

Native support on Wi-Fi devices for a **cheap** high frequency sampling of Wi-Fi RSS.

- By inducing traffic at the Wi-Fi device level, **the OS is bypassed**, which results in a **higher throughput** at a **reduced power level**.
- Reliable recognition capability at **lower cost**, compared to, for example, introducing a completely new HW like Google's Soli [4].

Questions

References I

Ericsson mobility report.
https://www.ericsson.com/en/mobility-report/latest-mobile-statistics.

[Online; accessed 22-Oct-2019].

Jie Song, Gábor Sörös, Fabrizio Pece, Sean Ryan Fanello, Shahram Izadi, Cem Keskin, and Otmar Hilliges.
In-air gestures around unmodified mobile devices.
In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 319–329, New York, NY, USA, 2014. ACM.

References II

Gabe Cohn, Daniel Morris, Shwetak Patel, and Desney Tan.
Humantenna: Using the body as an antenna for real-time whole-body interaction.
In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1901–1910, New York, NY, USA, 2012. ACM.

Project soli.
https://atap.google.com/soli/, 2016.
[Online; accessed 1-June-2016].

Chen Zhao, Ke-Yu Chen, Md Tanvir Islam Aumi, Shwetak Patel, and Matthew S. Reynolds.
Sideswipe: Detecting in-air gestures around mobile devices using actual gsm signal.
In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 527–534, New York, NY, USA, 2014. ACM.

Stephan Sigg, Shuyu Shi, Felix Buesching, Yusheng Ji, and Lars Wolf.
Leveraging rf-channel fluctuation for activity recognition: Active and passive systems, continuous and rssi-based signal features.
In *Proceedings of International Conference on Advances in Mobile Computing & Multimedia*, MoMM '13, pages 43:43–43:52, New York, NY, USA, 2013. ACM.

References IV

Rajalakshmi Nandakumar, Bryce Kellogg, and Shyamnath Gollakota.
Wi-fi gesture recognition on existing devices.
CoRR, abs/1411.5394, 2014.

S. Sigg, U. Blanke, and G. Tröster.
The telepathic phone: Frictionless activity recognition from wifi-rssi.
In *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*, pages 148–155, 2014.

References V

- Stephan Sigg, Mario Hock, Markus Scholz, Gerhard Tröster, Lars Wolf, Yusheng Ji, and Michael Beigl. *Mobile and Ubiquitous Systems: Computing, Networking, and Services: 10th International Conference, MOBIQUITOUS 2013, Tokyo, Japan, December 2-4, 2013, Revised Selected Papers*, chapter Passive, Device-Free Recognition on Your Mobile Phone: Tools, Features and a Case Study, pages 435–446. Springer International Publishing, Cham, 2014.

- Christoph Rauterberg, Mathias Velten, Stephan Sigg, and Xiaoming Fu. Simply use the force - implementation of rf-based gesture interaction on an android phone. In *IEEE/KuVS NetSys 2015 adjunct proceedings*, 2015.

References VI

C. Rauterberg and X. Fu.
Demo abstract: Use the force, luke: Implementation of rf-based gesture interaction on an android phone.
In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*, pages 190–192, March 2015.

Anthony Bagnall, Aaron Bostrom, James Large, and Jason Lines.
The great time series classification bake off: An experimental evaluation of recently proposed algorithms. extended version.
CoRR, abs/1602.01711, 2016.

Supporting slides

Offline experiments

Prediction accuracies, training and prediction times for different algorithms evaluated using Dataset1.

Algorithm	Accuracy	Sample prediction time (ms)
K-NN DTW	90% (±28)	964.15
FS	85% (±4.6)	0.01
STE	91% (±1.1)	26.86
LTS	93% (±2.3)	9.29
EE	93% (±1.7)	23.09
COTE	94% (±2.4)	178.20
LSTM RNN	91% (±3.1)	7.04

STE, EE and COTE are ensemble methods that are computationally heavy [12].