Communication Technologies 2 (CT2)

Machine Learning:

Applications and Algorithms

Wi-Fi Fingerprinting

Doan Duong

Lecture in WS 2018 / 2019
29.11.2018

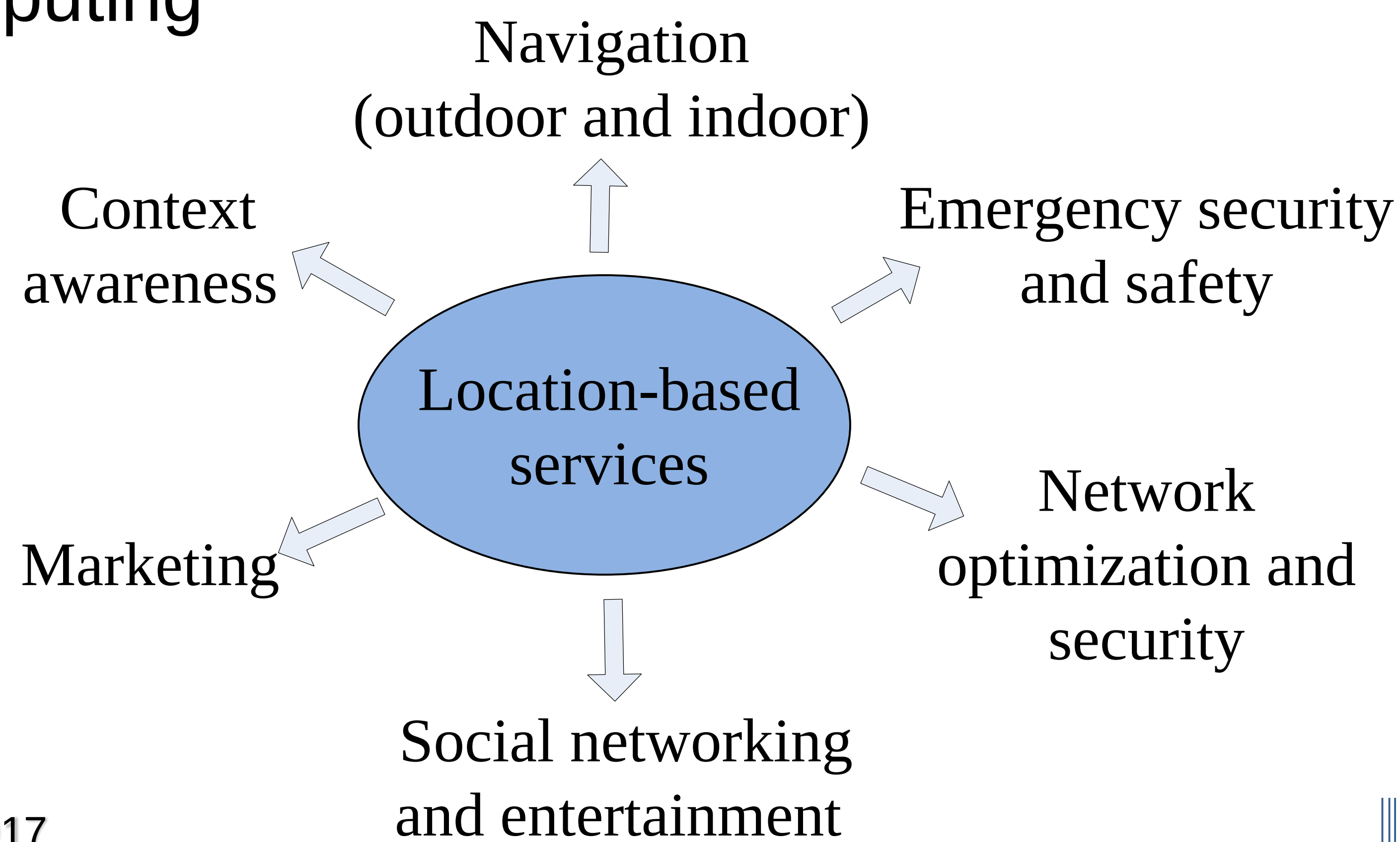
Outline

ComTec

- Location-based services
- Indoor positioning
- Wi-Fi based positioning
- Wi-Fi fingerprinting system
- DCCLA algorithm
- Challenges

Location-based Services

- The location information provided is now one of the most important contexts for context-aware computing



The Global Positioning System (GPS)

ComTec

- Widely used for providing positioning services.
- Require line-of-sight (LoS) transmission with the GPS satellites.
- Poor coverage for indoor environments

Source: http://www.mio.com/

Free-space Path Loss

ComTec

- Assume there are no obstructions between the transmitter and receiver. The signal propagates along a straight line
- Free-space path loss is defined as the path loss of the free-space model:

$$P_L\,(\mathrm{dB}) = 10\log_{10}\frac{P_t}{P_r} = -10\log_{10}\frac{G_l\lambda^2}{(4\pi d)^2}.$$

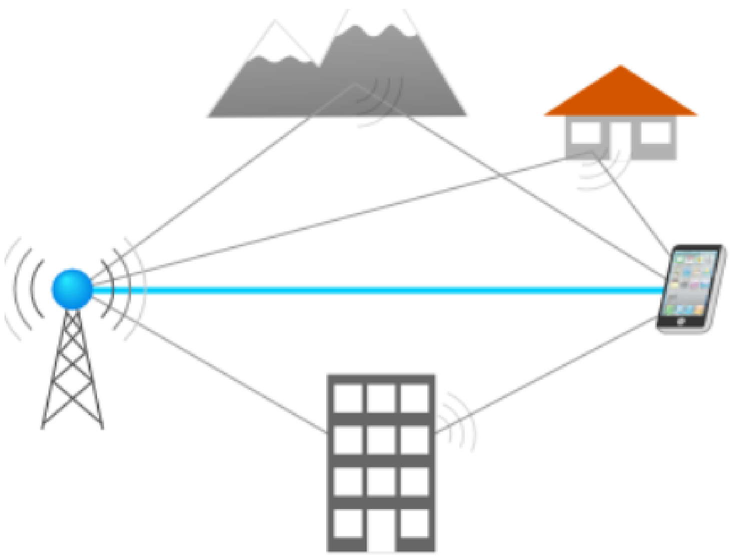
d: distance between the transmit and receive antennas

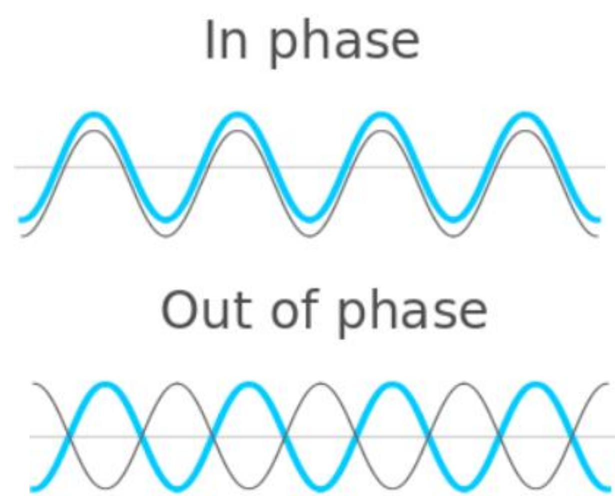
ʎ: the wavelength of the transmitted signal

G_l: the product of the transmit and receive antenna gains

Free-space Path Loss

- Difficult to model the radio propagation in the indoor environment:
  - Multipath, fading
  - non light-of-sight transmission because of obstacles





Source:
http://wiki.yatebts.com/index.php/Radio_Propagation_Concepts

Indoor Positioning

- Different locations can be distinguished by particular features surrounded those locations
  - light, noise, temperature
  - radio signals: Bluetooth, GSM or Wi-Fi
- Good characteristic help to recognize different locations.
- The selected characteristic should be stable over time.

Wi-Fi Positioning System

- Uses the received signal strengths (RSS) from surrounded access points (AP) to locate a mobile node.
- Two methods of using the Wi-Fi RSS to determine a receiver's position
  - lateration methods
  - fingerprinting methods

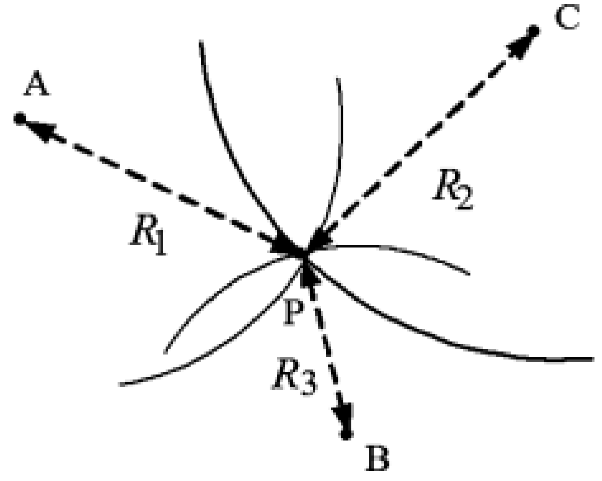
Methods & positioning principle

- Time of Arrival (TOA)
- Time Difference of Arrival (TDOA)
- RSS-Based (Signal Attenuation-Based)
- Roundtrip time of flight of the signal (RTOF)
- Angulation estimation (AOA Estimation)
- Receive Signal Strength (RSS)

Trilateration methods

ComTec

- Time of Arrival (TOA)
  - Measure the propagation time
  - Calculate the distance between the receiver and the transmitter.
  - The clock of transmitters and receivers must be precisely synchronized

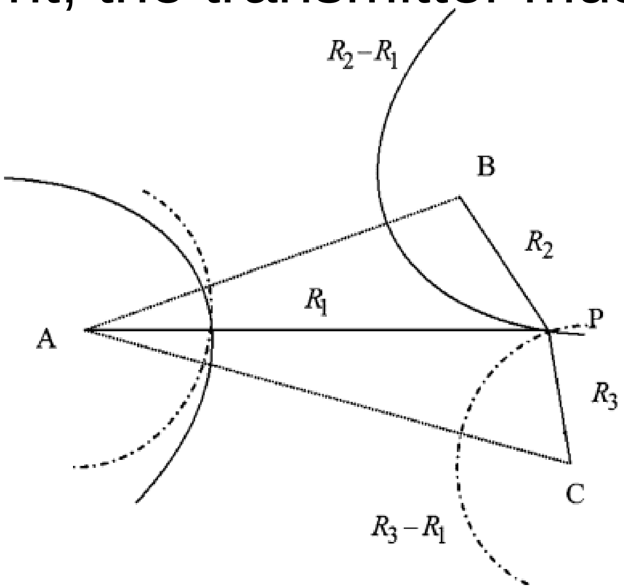


Positioning based on TOA/RTOF measurements.

Source: [2]

Trilateration methods

- Time Difference of Arrival (TDOA)
  - the time difference to send signal to multiple receivers
  - The hyperbolic has a constant range difference between the two measuring units.
  - For each TDOA measurement, the transmitter must lie on a hyperbolic.



Source: [2]

Trilateration methods

ComTec

- RSS-Based (Signal Attenuation-Based)
    - Calculate the attenuation of emitted signal strength
    - Base on the signal path loss model, estimate the travel distance
- Roundtrip time of flight of the signal (RTOF)
    - Measure the propagation time from the transmitter to the receiver and back

Fingerprinting method

ComTec

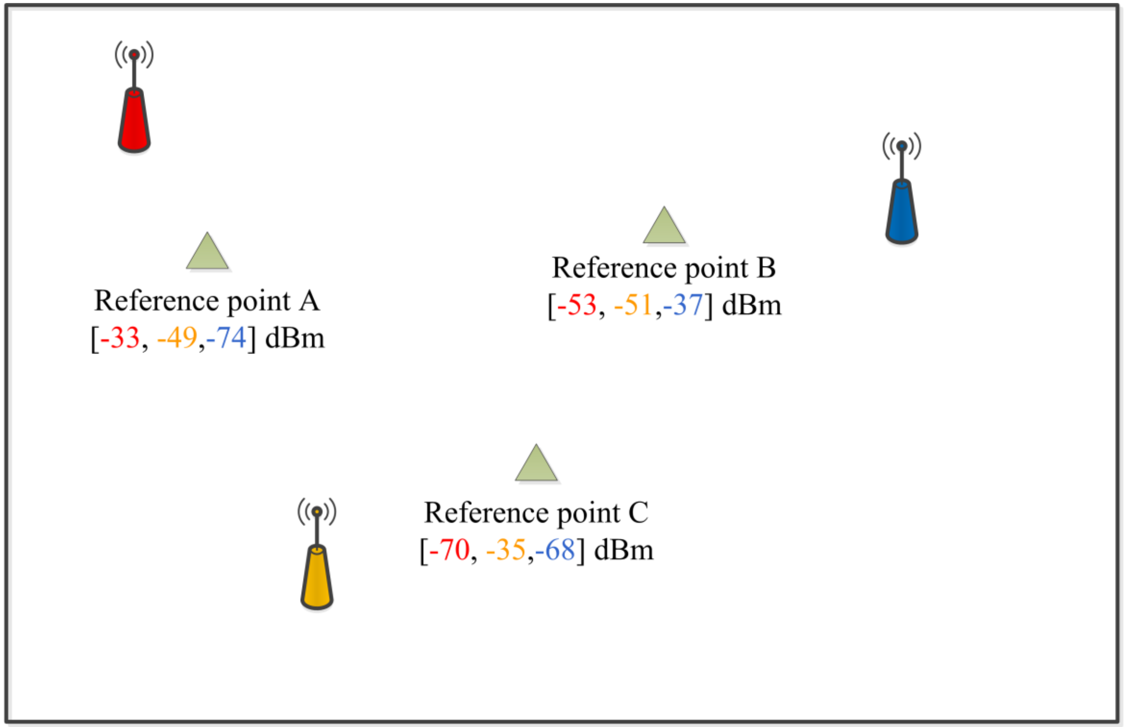
- Human fingerprint is unique
  - Collect human fingerprint and store in a database
  - To identify any person, compare his/her fingerprint with the fingerprinting database.

Source: [1]

Wi-Fi signal

ComTec

- MAC address
- Receive signal strength (RSS)
- Service set identifier (SSID)
- Channel (frequency)



Wi-Fi signal

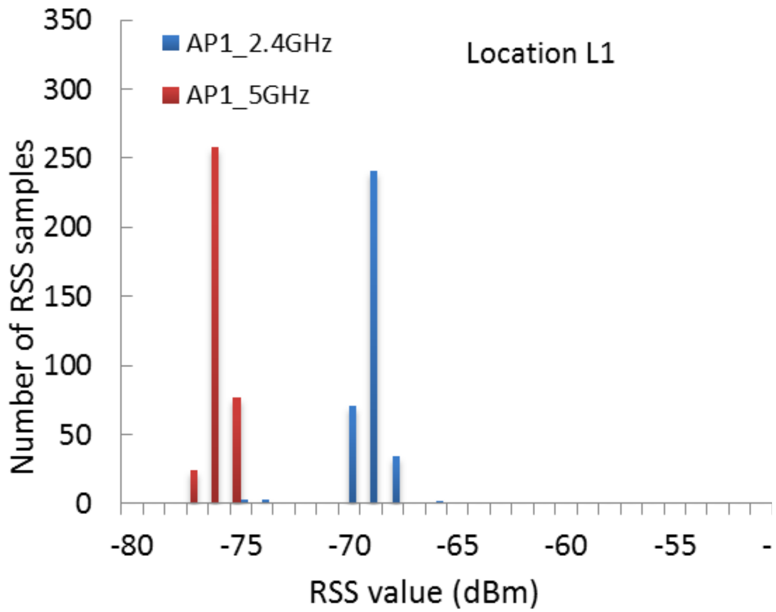
ComTec

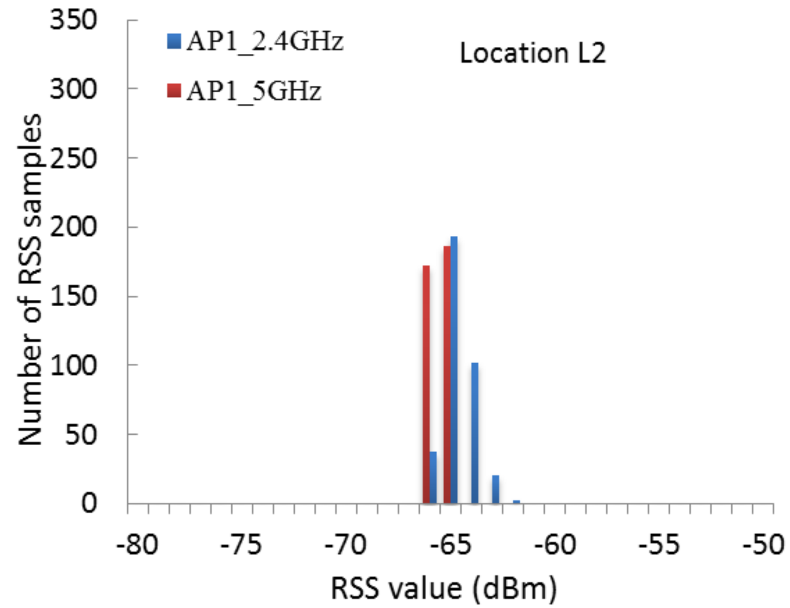
- MAC address
- Receive signal strength (RSS)
- Service set identifier (SSID)
- Channel (frequency)

Source: [1]

Wi-Fi signal

ComTec





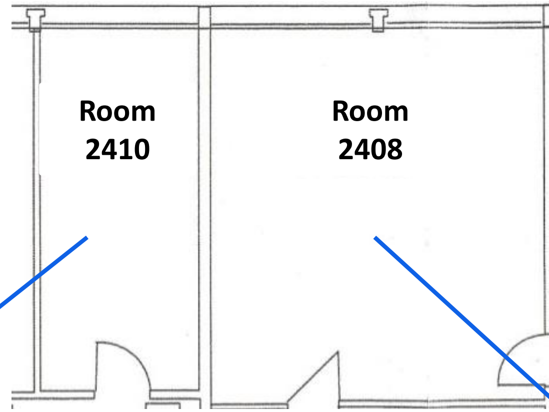
Wi-Fi RSS histogram in 2 adjacent rooms

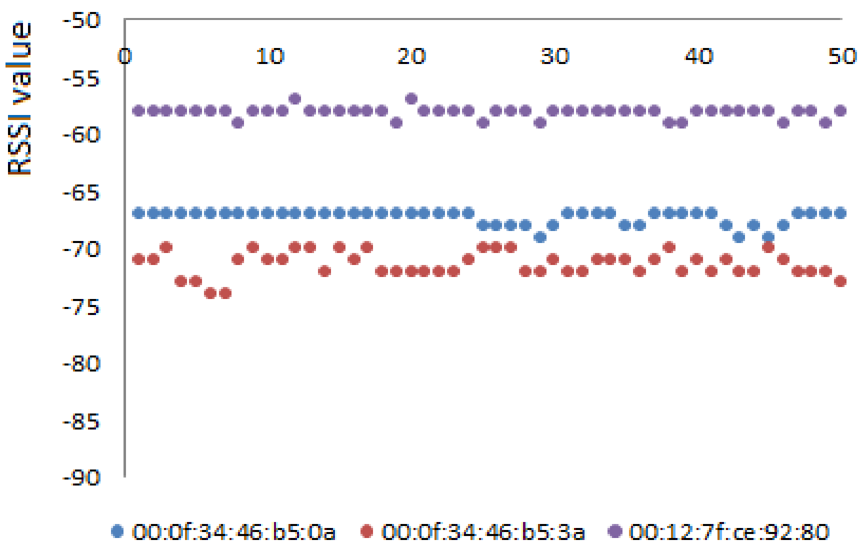
Wi-Fi fingerprinting

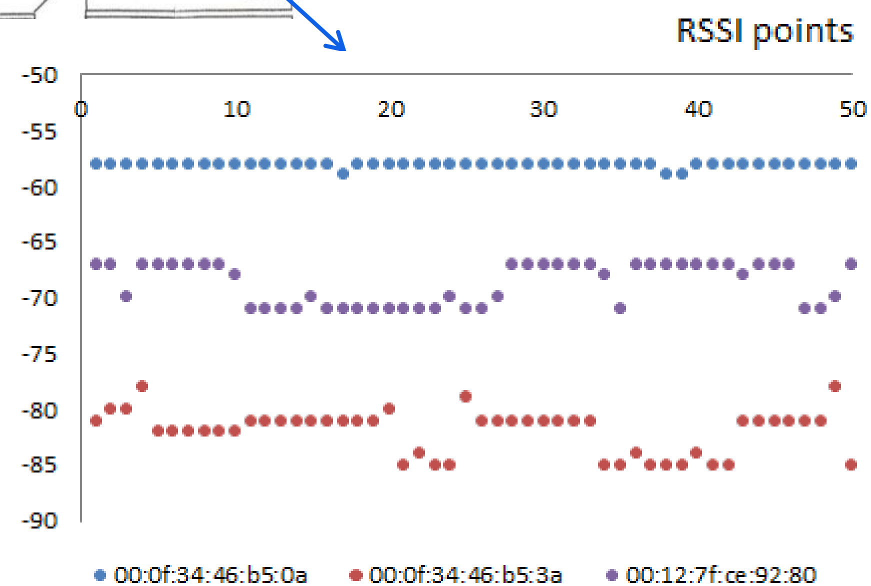
- The fingerprinting technique uses location specific RSSI pattern of neighboring WLAN APs to distinguish different locations
- Not require the dedicated infrastructure installation.
  – Easy to implement
  – Reduce cost

Wi-Fi fingerprinting

What is
Wi-Fi Fingerprint





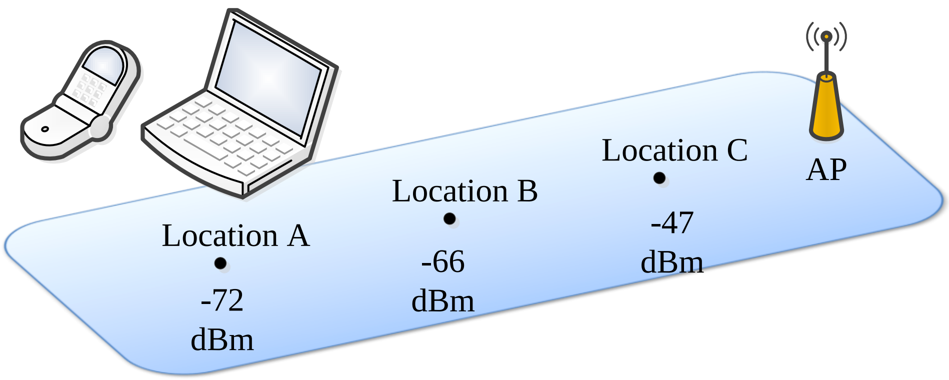


Wi-Fi fingerprint

RSSI: Received Signal Strength Indicator

Concepts

- Wi-Fi fingerprint
  - Wi-Fi information sensed at a location

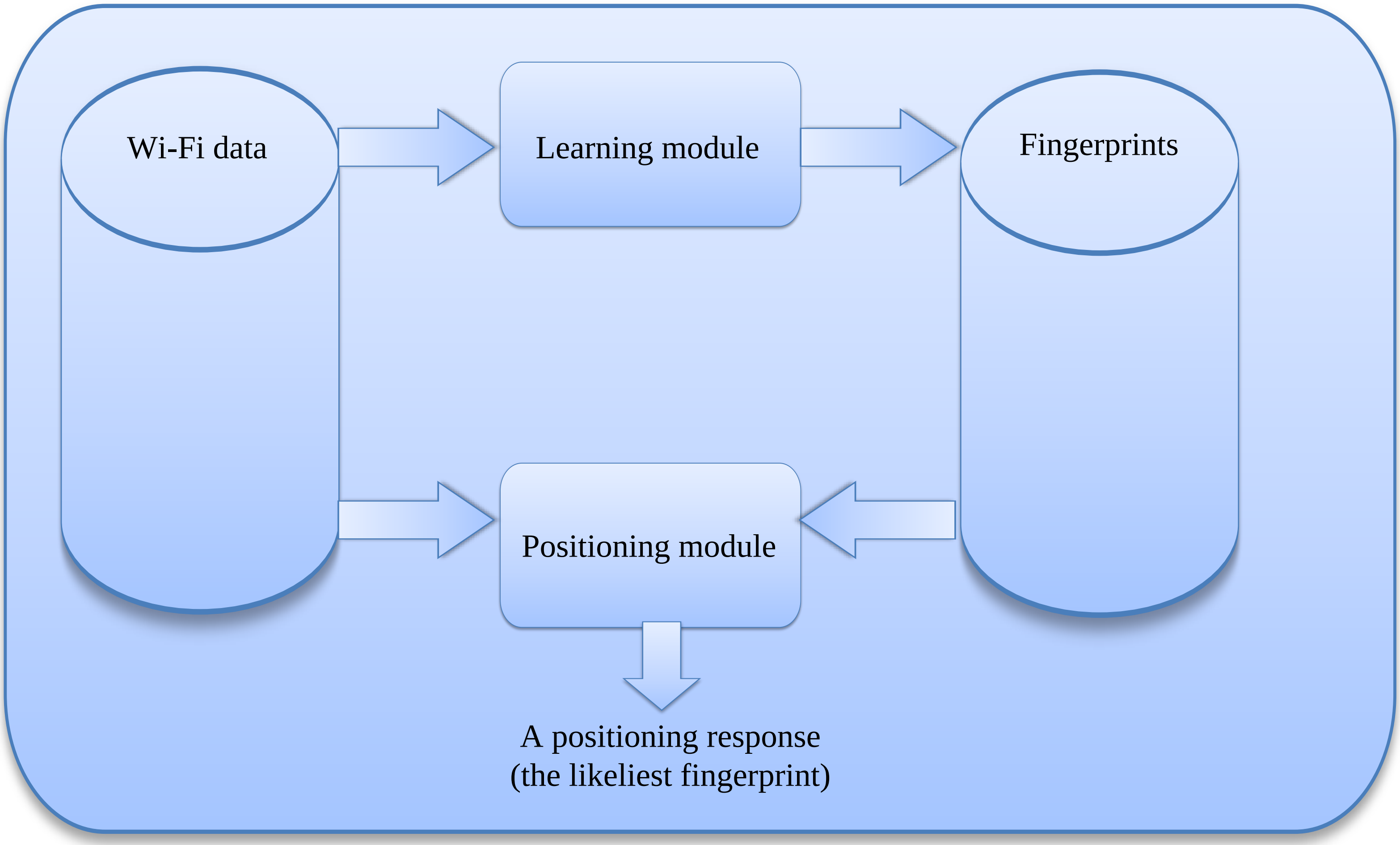


Concepts

ComTec

- Wi-Fi fingerprint
  - Wi-Fi information sensed at a location
- Wi-Fi fingerprinting
  - practice of recording the fingerprint of Wi-Fi signal
  - the method of finding the particular pattern of Wi-Fi signal in a particular location
- Wi-Fi fingerprint database
  - a database which contains the generated fingerprints
  - radio map

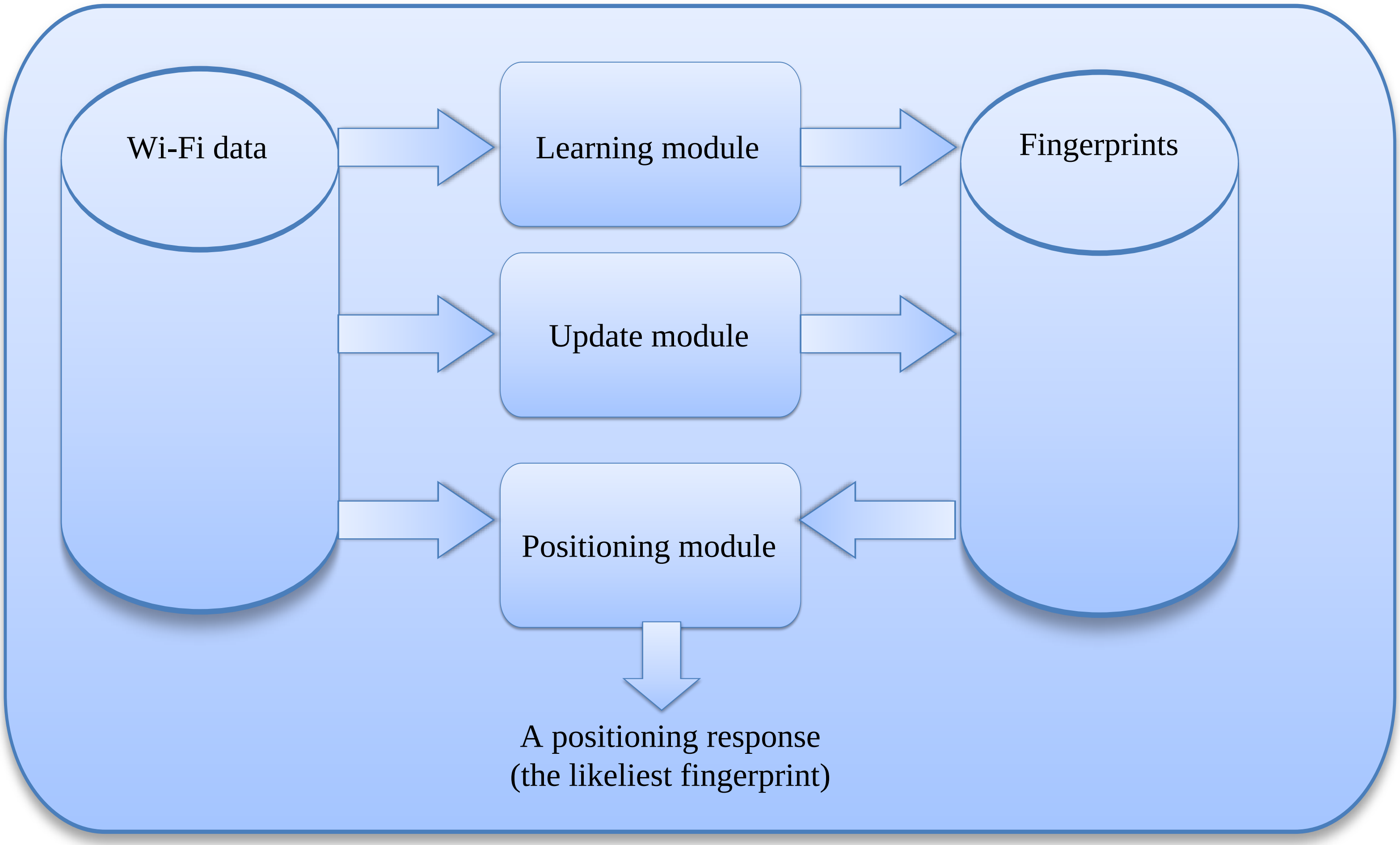
Overview of Wi-Fi fingerprinting System



Source: [1]

Overview of Wi-Fi fingerprinting System

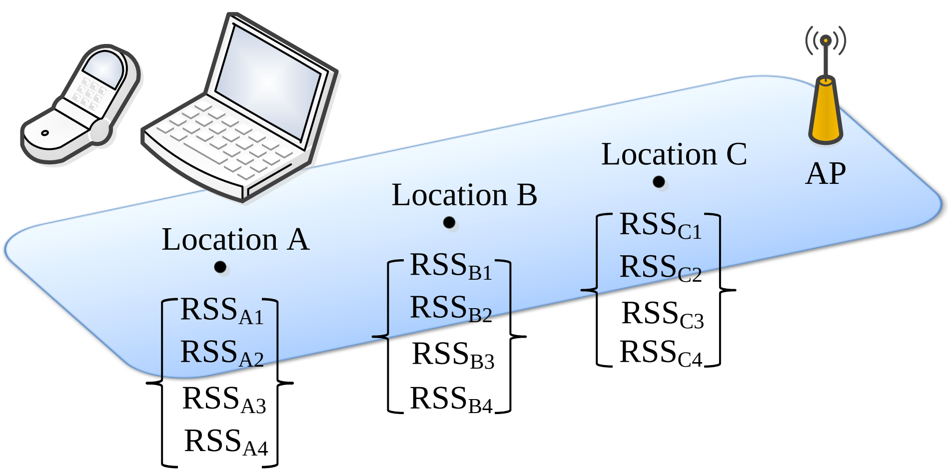
ComTec



Source: [1]

Overview of Wi-Fi fingerprinting System

- Learning phase (training, offline phase)
  - Collecting Wi-Fi data
  - Generating Wi-Fi fingerprint: building a database containing Wi-Fi fingerprints

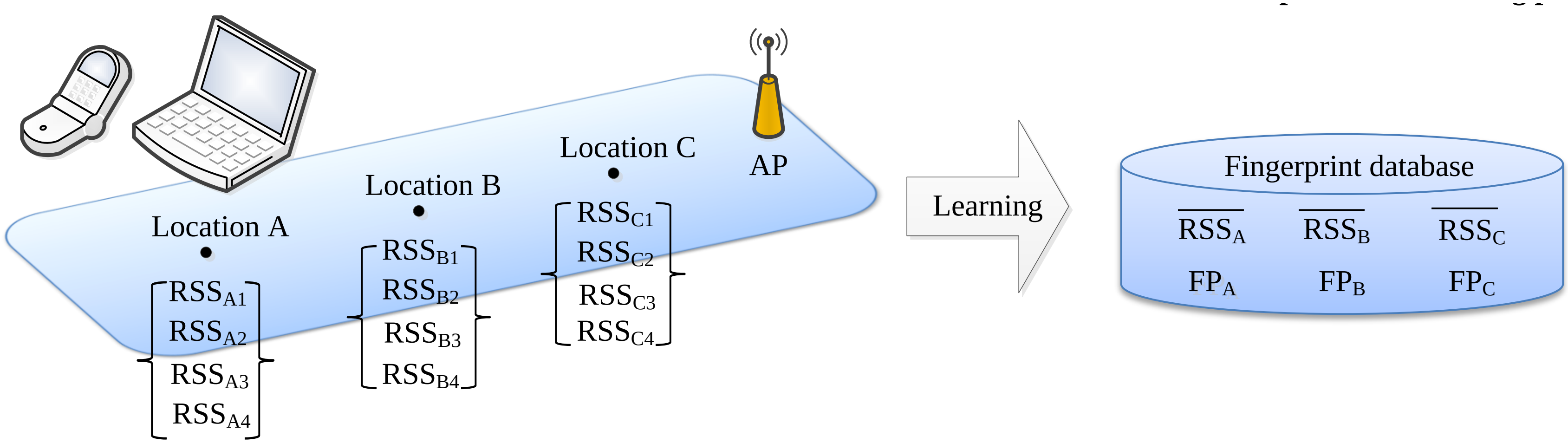


Source: [1]

Overview of Wi-Fi fingerprinting System

ComTec

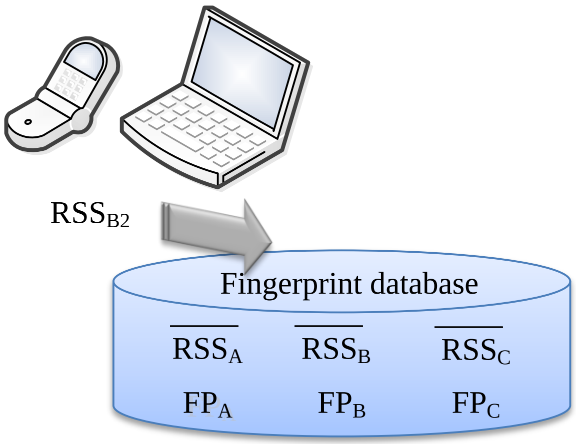
- Learning phase (training, offline phase)
  - Collecting Wi-Fi data
  - Generating Wi-Fi fingerprint: building a database containing Wi-Fi fingerprints



Source: [1]

Overview of Wi-Fi fingerprinting System

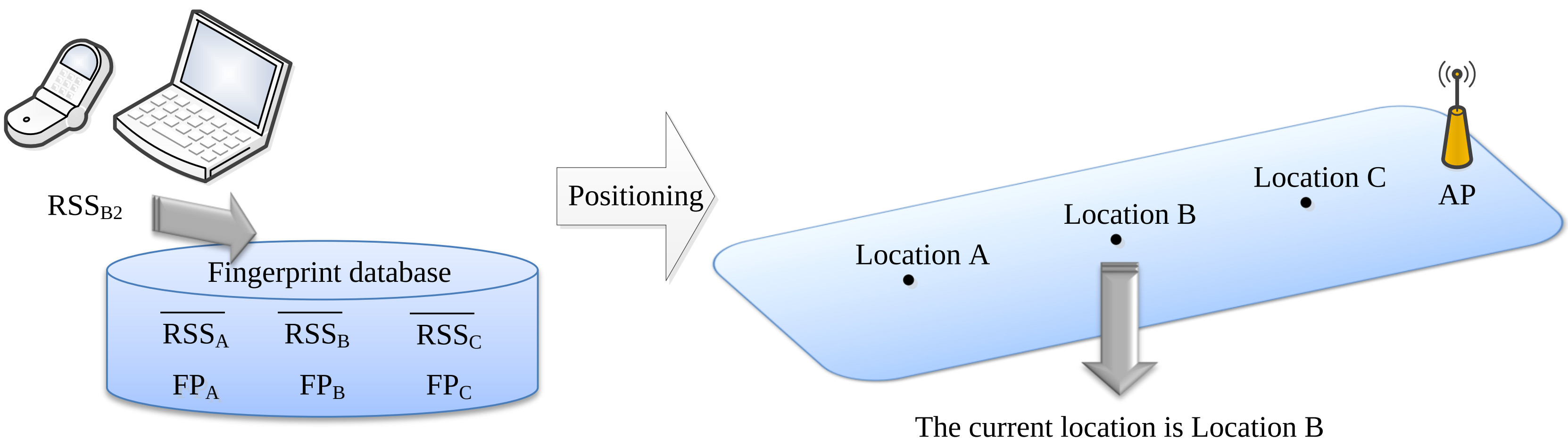
- Positioning phase (testing, online phase)
  - Comparing the current Wi-Fi scans with the FP database to infer the location



Source: [1]

Overview of Wi-Fi fingerprinting System

- Positioning phase (testing, online phase)
  - Comparing the current Wi-Fi scans with the FP database to infer the location



Source: [1]

Components of Wi-Fi fingerprinting System

- Wi-Fi APs
  - broadcast Wi-Fi signal
- Mobile devices
  - collect the Wi-Fi RSS from APs
  - manage fingerprint database
  - execute the learning and positioning algorithms
  - user interface (UI)
- Users
  - carry the mobile device in their daily lives
  - interact with the mobile device through the UI.

Components of Wi-Fi fingerprinting System

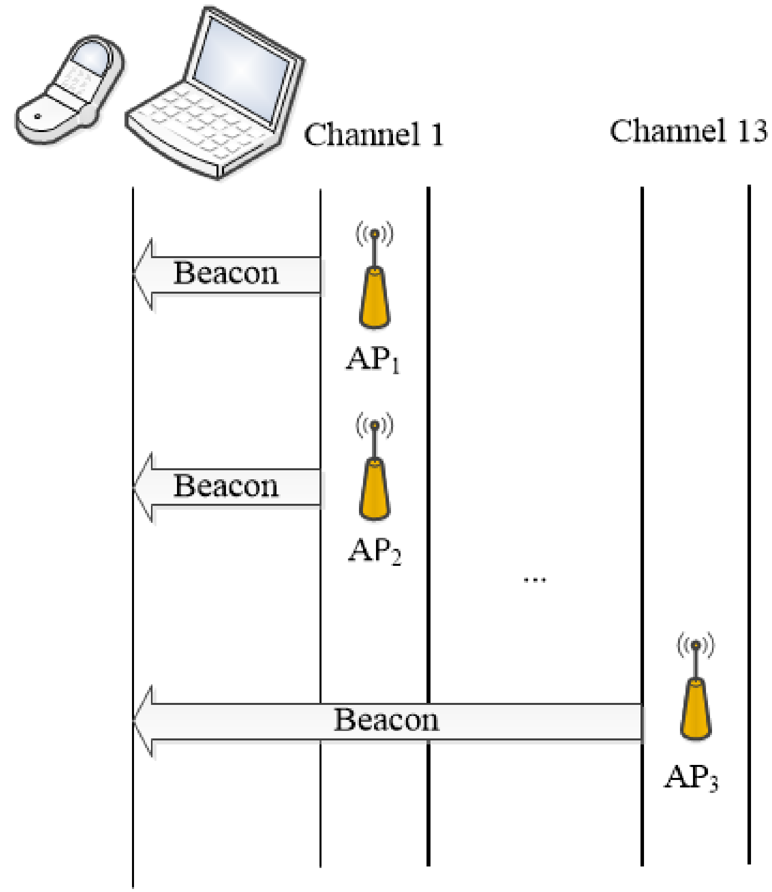
- Wi-Fi APs
  - broadcast Wi-Fi signal
- Mobile devices
  - collect the Wi-Fi RSS from APs
  - manage fingerprint database
  - execute the learning and positioning algorithms
  - user interface (UI)
- Users
  - carry the mobile device in their daily lives
  - interact with the mobile device through the UI.

May be done on a computer

Scanning methods

ComTec

- Passive scanning
  - Moves to each channel in the channel list
  - Listen to the beacon frames from nearby APs



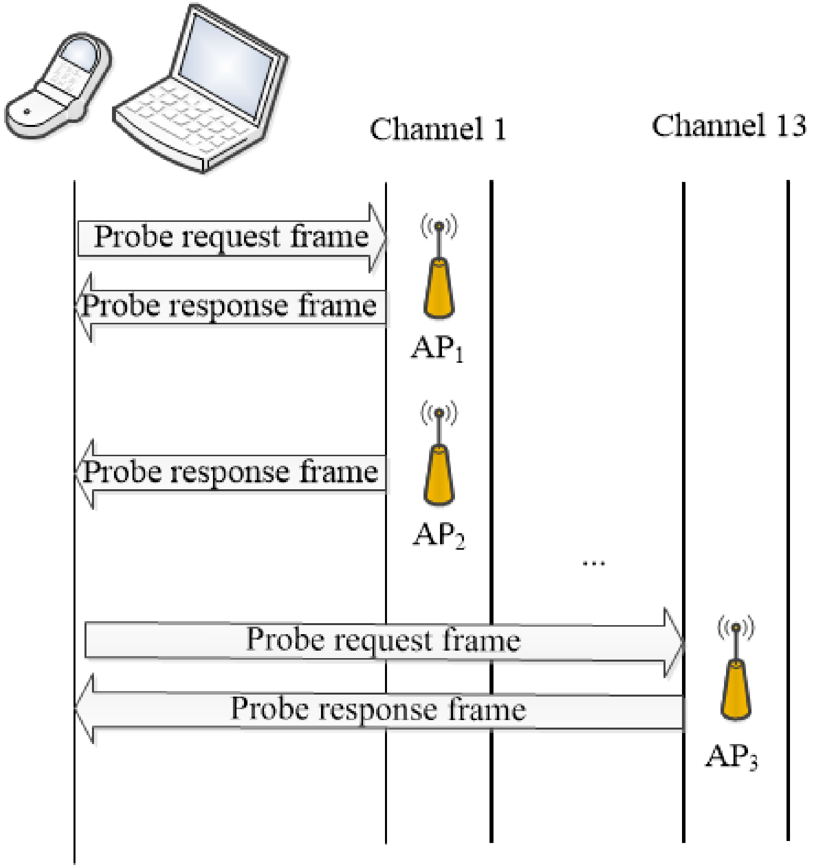
Passive scanning

Source: [1]

Scanning methods

ComTec

- Active scanning:
  - Mobile device transmits a Probe Request frame on each of the channel
  - Wait to receive response from any APs
  - Move on to the next channel
  - This process is iterated until all channels have been scanned



Active scanning

Source: [1]

Wi-Fi collection approaches

- Grid-based
  - split a selected area into many small grids
  - measure Wi-Fi signal in those grids
- Interpolation-based
  - measure Wi-Fi data in observed locations
  - based on the fingerprints of the observed locations, interpolate fingerprints for other locations
- Path survey (War-driving)
  - collect Wi-Fi data when moving
  - record the walking distances at the same time

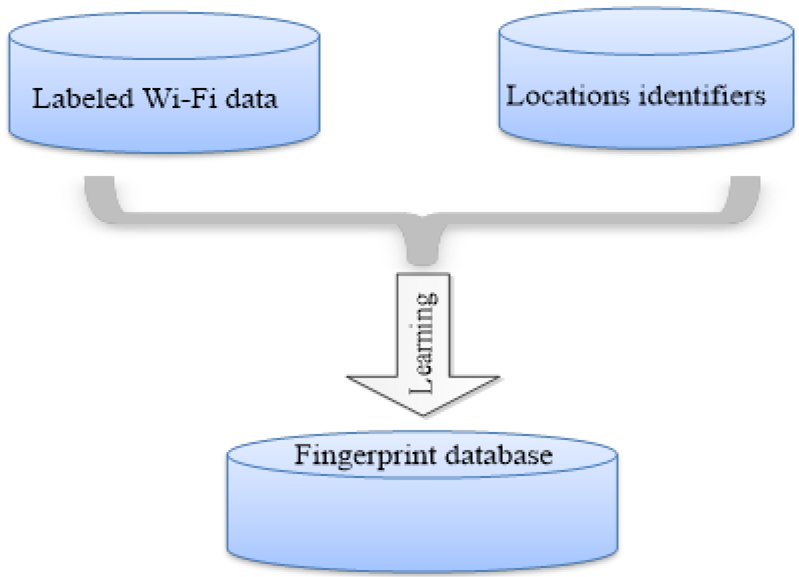
Wi-Fi collection approaches

- Dead Reckoning
  - infer location based on previously determined locations
  - combine with the traveled distance and the direction of movement
- Place learning
  - discover the significant locations where people spend much time by analyzing various sensor data
- Crowdsourcing
  - Using raw Wi-Fi data, fingerprints contributed by different users, devices

Learning approaches

ComTec

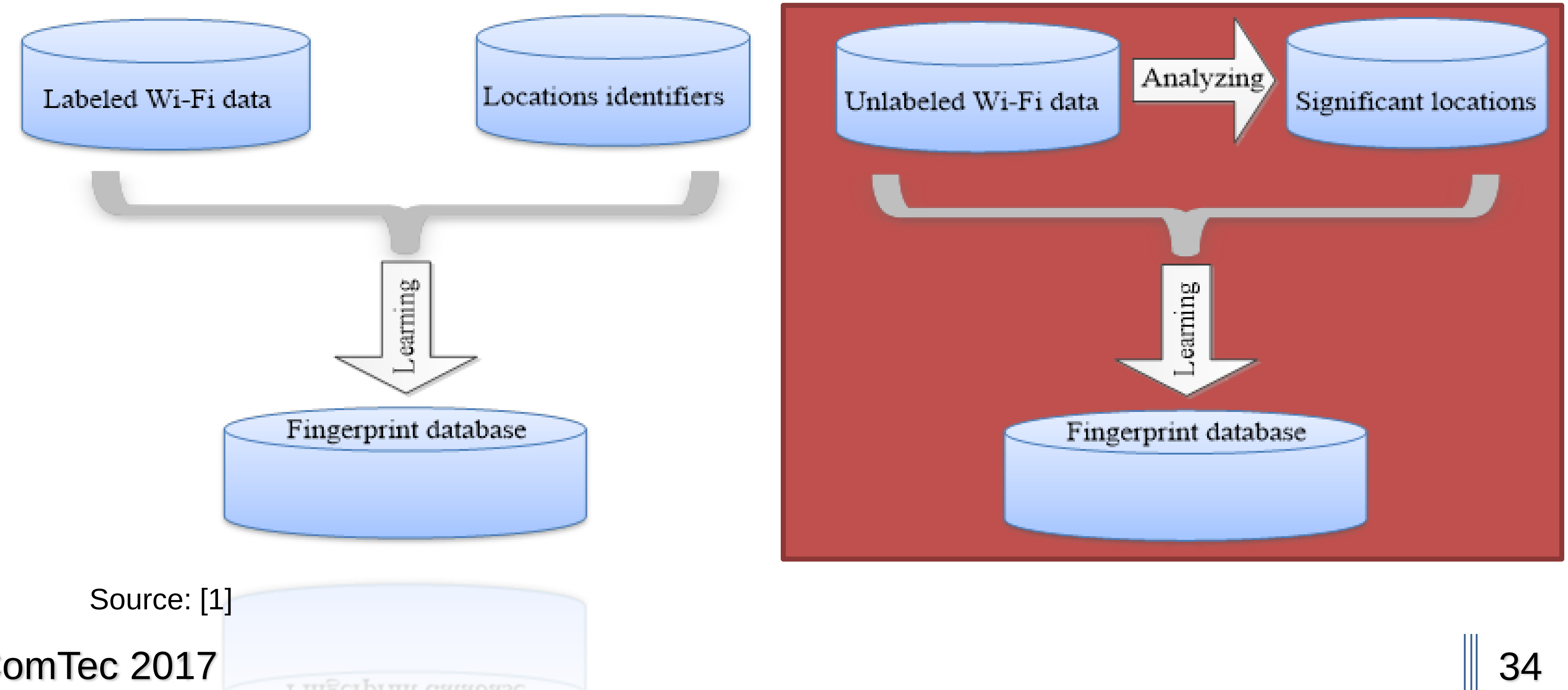
- Supervised
- Semi supervised



Source: [1]

Learning approaches

- Supervised
- Semi supervised
- Unsupervised



Source: [1]

Fingerprint feature

ComTec

MAC$_i$: MAC address of AP$_i$
$\overline{RSS_i}$: averaged received signal strength of AP$_i$

- Deterministic technique
  - Mean RSS value

$$FP_l = \{MAC_1:\overline{RSS_1}, \dots\ MAC_p:\overline{RSS_p}\} \mid (p \in \mathbb{N}_+)$$

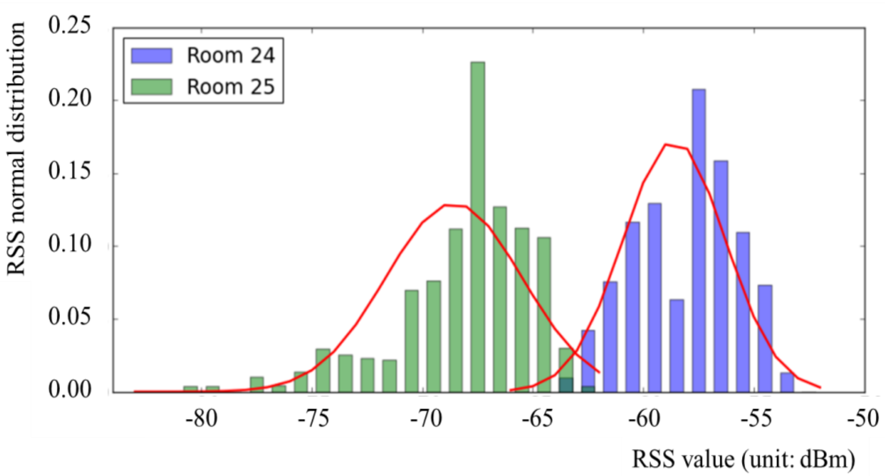
  - Wi-Fi scan

$$S_l^i = \{MAC_1: RSS_1^i, \dots, MAC_p: RSS_p^i\} \mid (i, p \in \mathbb{N}_+)$$

  - RSS range

```
00:24:fe:04:b2:d3 : [-58,-66]
00:24:fe:ac:72:02 : [-71,-79]
```

  - RSS histogram



Source: [1]

Fingerprint feature

ComTec

- Deterministic technique
  - Wi-Fi ratio: the RSS ratio between pairs of APs

$$r(AP_i, AP_j) = \frac{RSS_i}{RSS_j} \mid (i, j \in \mathbb{N}_+;\ i < j)$$

  - RSS difference: the differences between pairs of APs

$$d(AP_i, AP_j) = RSS_i - RSS_j \mid 1 \leq i < j \leq n$$

  - AP rank: rank values are assigned to APs based on their RSS

$$FP_l = \{1: MAC_i, \ldots, p: MAC_j\} \mid (RSS_i \geq RSS_j;\ i, j, p \in \mathbb{N}_+)$$

-The response rate: the times of sensing the AP divided by the times of scans.

MAC_i: MAC address of AP_i
RSS_i: received signal strength of AP_i

Fingerprint feature

- Probabilistic technique
  - Gaussian distribution
    - In case the RSS values of an AP are distributed as a Gaussian model.
    - A statistical signature, as a fingerprint, is generated for each location by using the Gaussian model.
  - Non-Gaussian distribution
    - In case the RSS distribution is non-Gaussian, an improved double-peak Gaussian distribution has been proposed to approximate the RSS distribution.
  - Log-normal distribution
    - The RSS distribution may be asymmetric

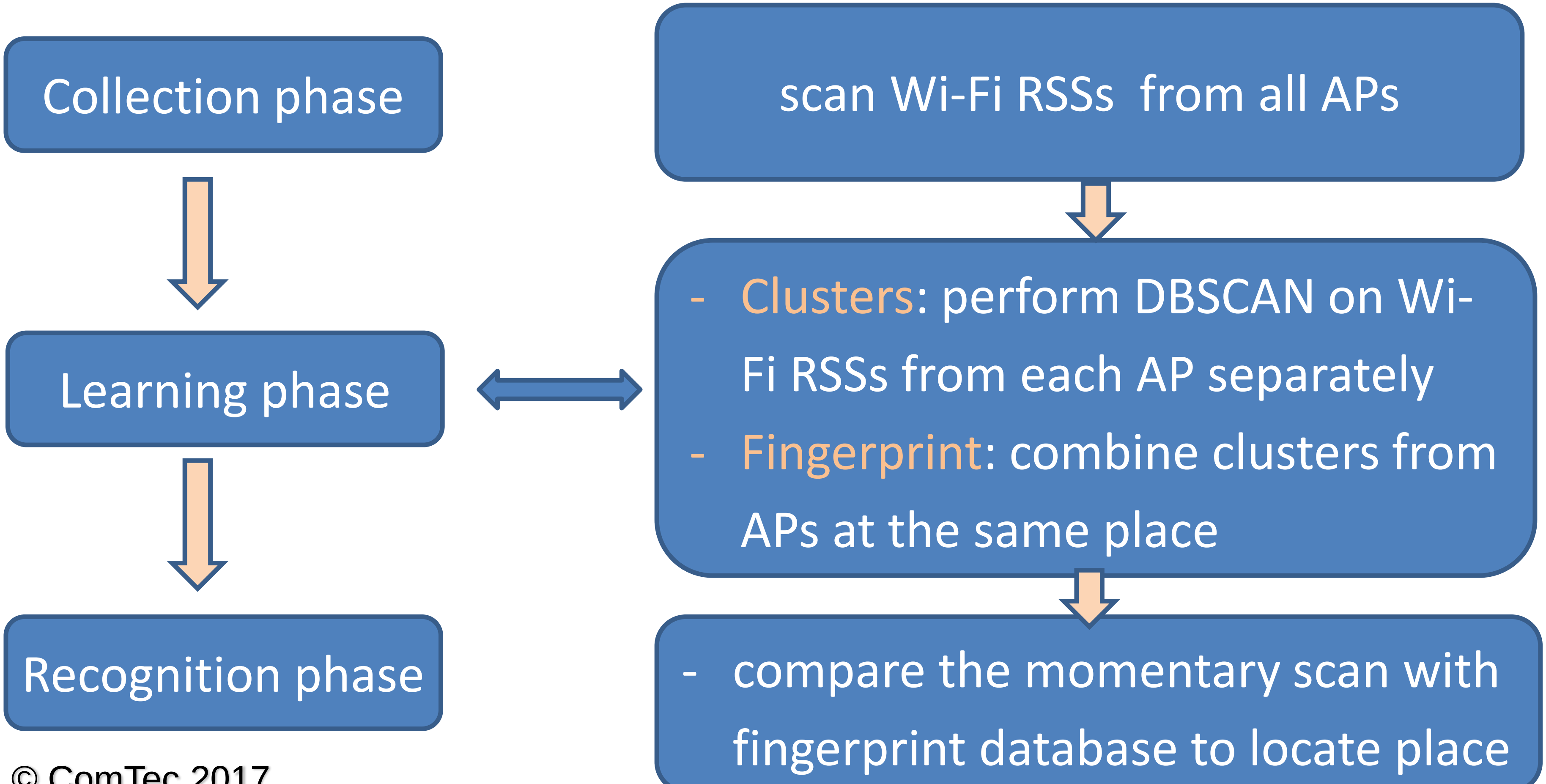
Accuracy

ComTec

- In meter
  - how close is the estimated location to the true location indicated as *error distance (m)*
- Room level
  - locating a user within a room

DCCLA algorithm

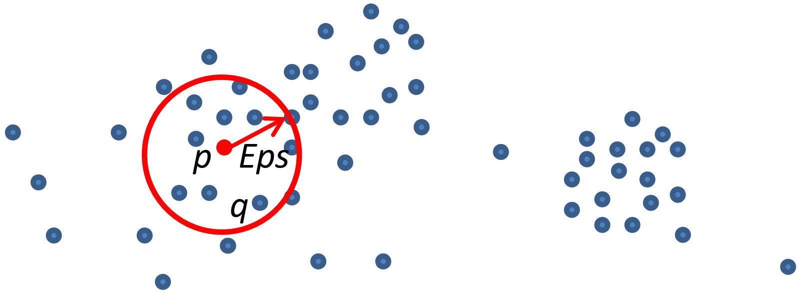
- Wi-Fi Fingerprinting-based localization
- Unsupervised Density-based Clustering method



Density-based clustering

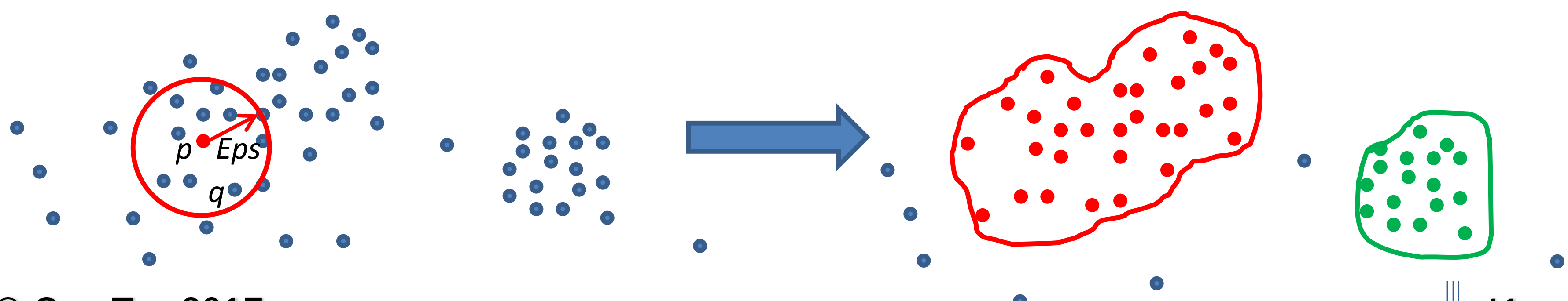
ComTec

- the *density-based spatial clustering of applications with noise* (DBSCAN) algorithm utilizes concepts of *neighborhood* and *neighborhood-density to extract clusters*.
- 2 parameters
  - *Eps*: radius of the neighborhood
  - *MinPts*: minimum number of points in a cluster



Density-based clustering

- the *density-based spatial clustering of applications with noise* (DBSCAN) algorithm utilizes concepts of *neighborhood* and *neighborhood-density to extract clusters*.
- 2 parameters
  - *Eps*: radius of the neighborhood
  - *MinPts*: minimum number of points in a cluster

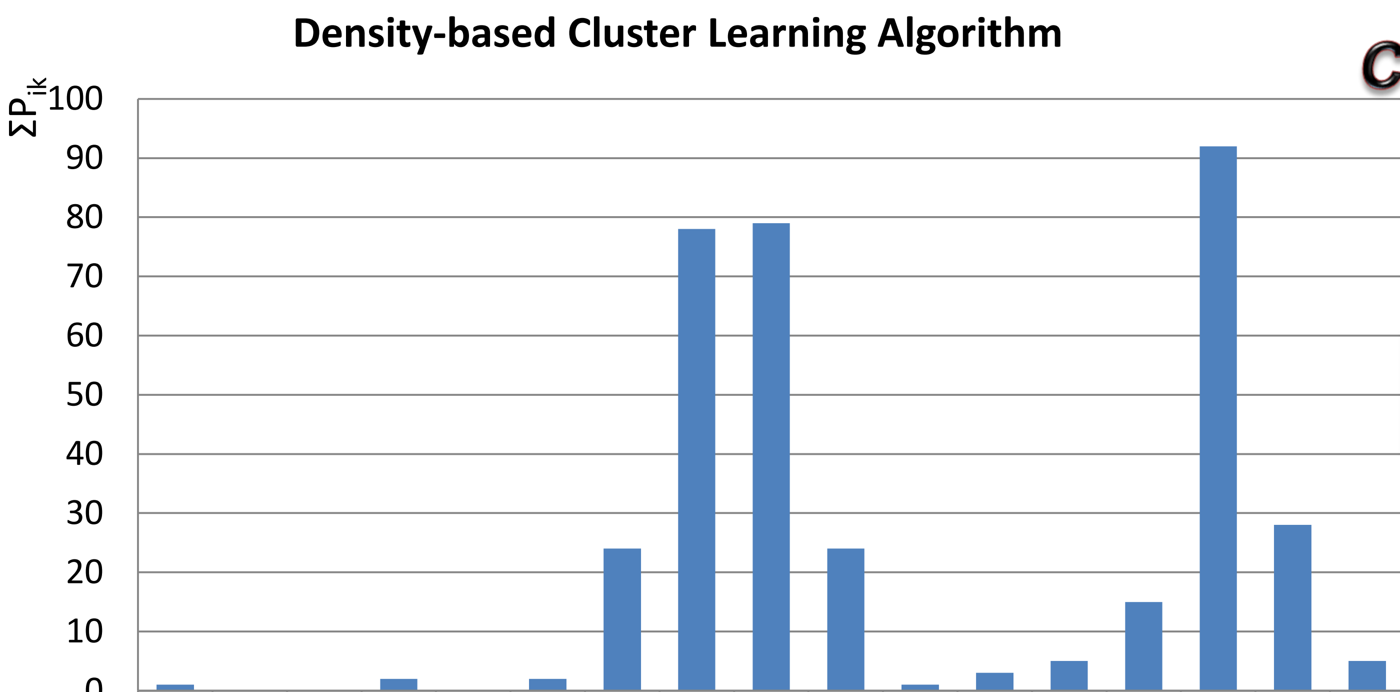


Wi-Fi clustering using DBSCAN

- Discover clusters
  - perform DBSCAN on each AP to discover high-density clusters.
  - cluster: consecutive RSS range with high RSS density
- Fingerprint
  - Clusters appear at the same time from different APs

Wi-Fi clustering using DBSCAN

ComTec



Clusters ?

NR = 2
MinPts = 120

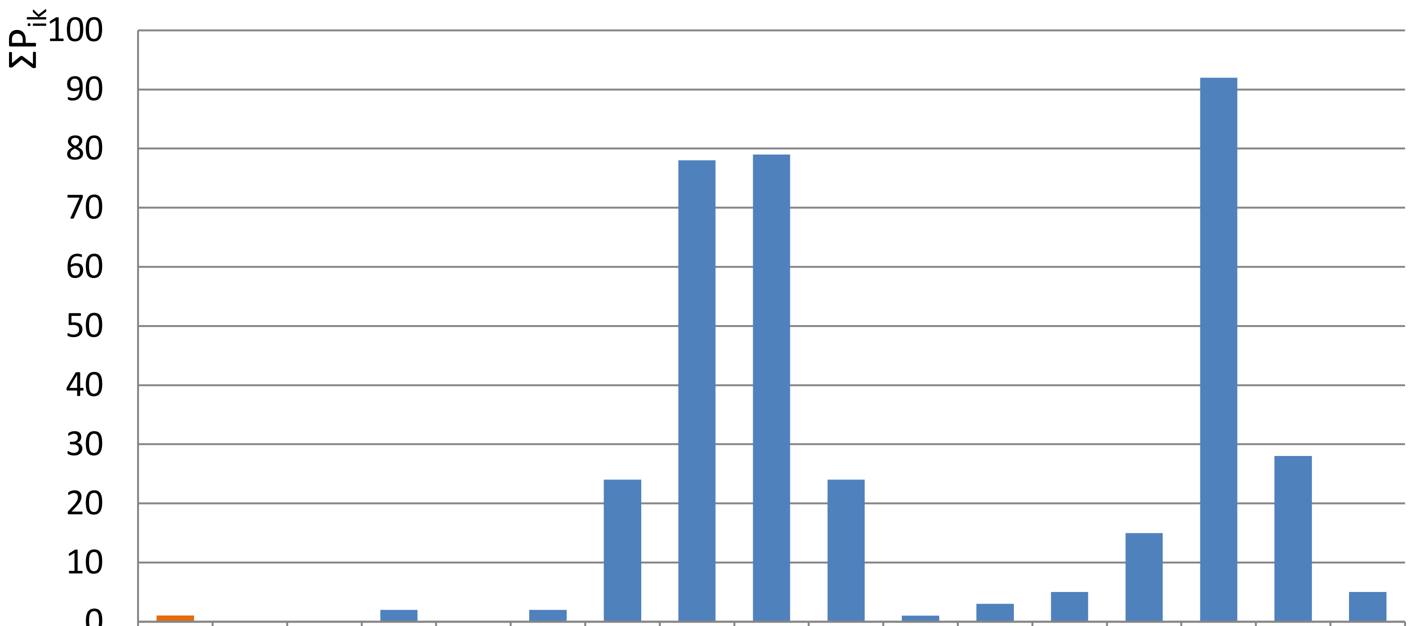
	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62	RSSI$_{ik}$ (dBm)
$\sum_{k=m}^{m} P_{ik}$ →	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5	

Data in a Wi-Fi database: collected in room 2408 between 09:00 - 09:30 on 15.08.2012. *MAC$_i$ = 00:0f:34:46:b5:0a*
$\sum_{k=-78}^{-62} P_{ik}$ = 359

Wi-Fi clustering using DBSCAN

ComTec

Density-based Cluster Learning Algorithm



NR = 2
MinPts = 120

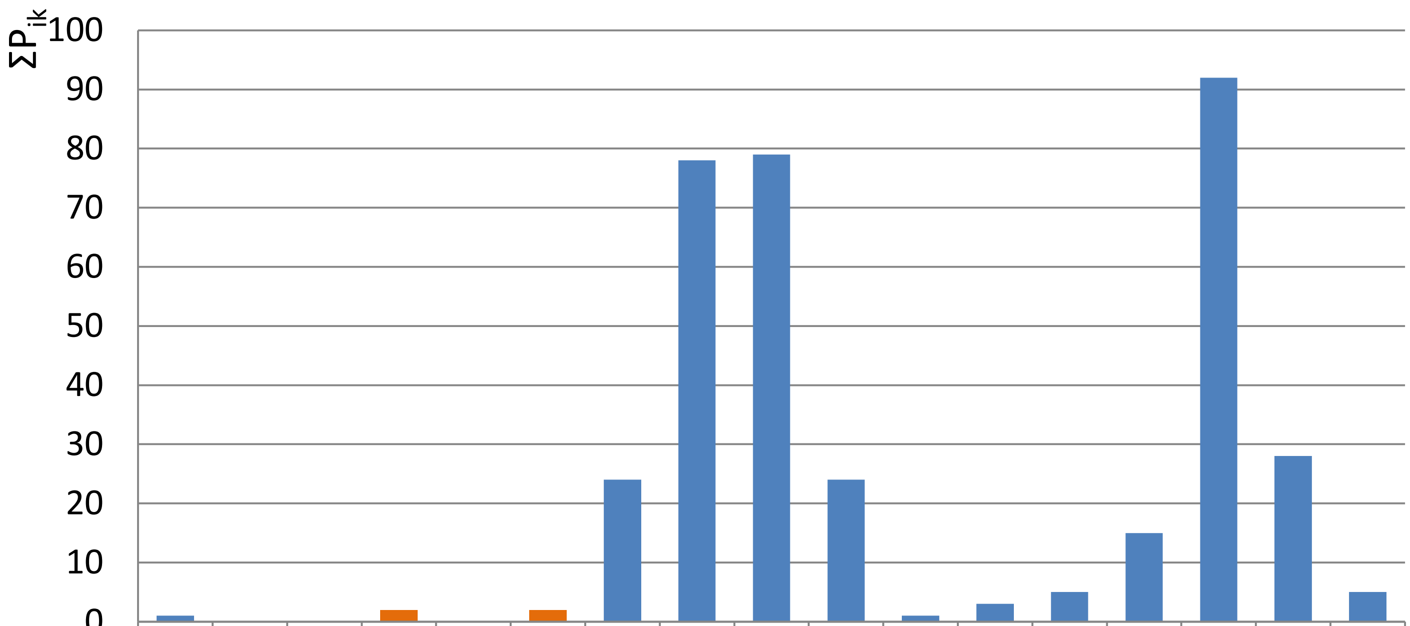
RSSI$_{ik}$ (dBm)	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62
$\sum_{k=m}^{m} P_{ik}$	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5

$$\rho\ (P_{ik}) = \sum_{k=-78}^{-76} P_{ik} = 1 < 120$$

Wi-Fi clustering using DBSCAN

ComTec

Density-based Cluster Learning Algorithm



$RSSI_{ik}$ (dBm)	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62
$\sum_{k=m}^{m} P_{ik}$	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5

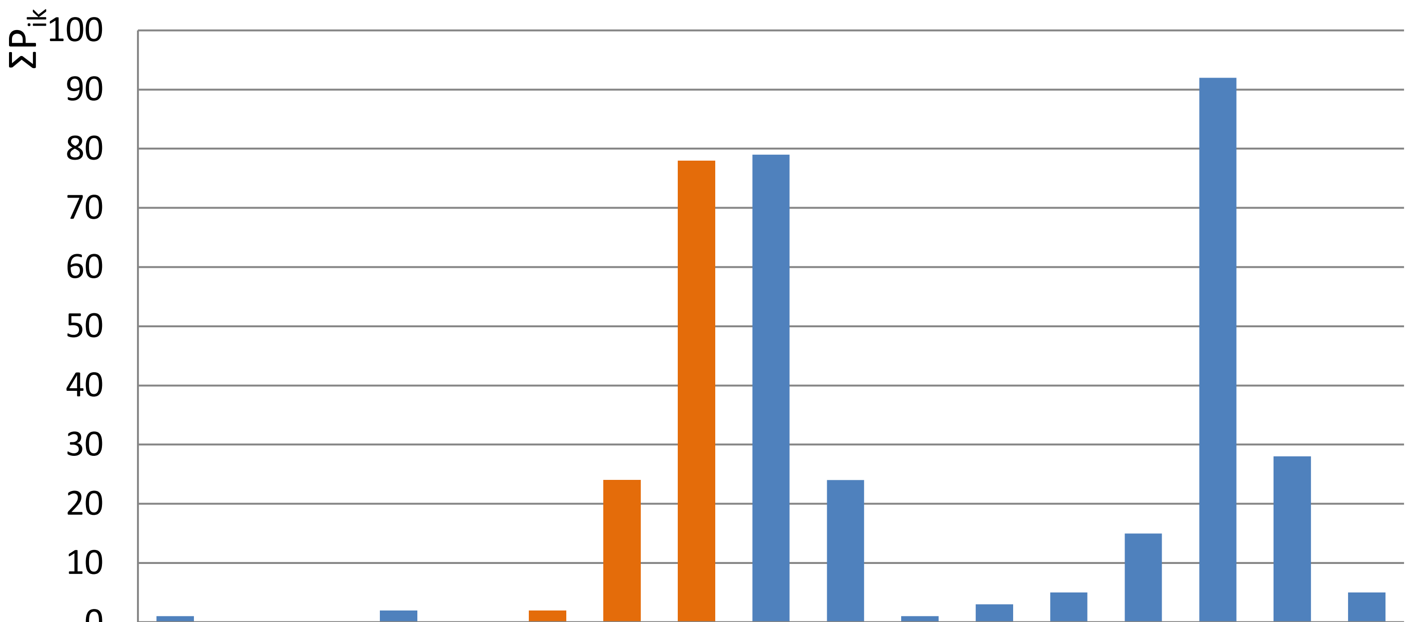
NR = 2
MinPts = 120

$$\rho\,(P_{ik}) = \sum_{k=-75}^{-73} P_{ik} = 4 < 120$$

Wi-Fi clustering using DBSCAN

ComTec

Density-based Cluster Learning Algorithm



	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62	← $RSSI_{ik}$ (dBm)
$\sum_{k=m}^{m} P_{ik}$ →	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5	

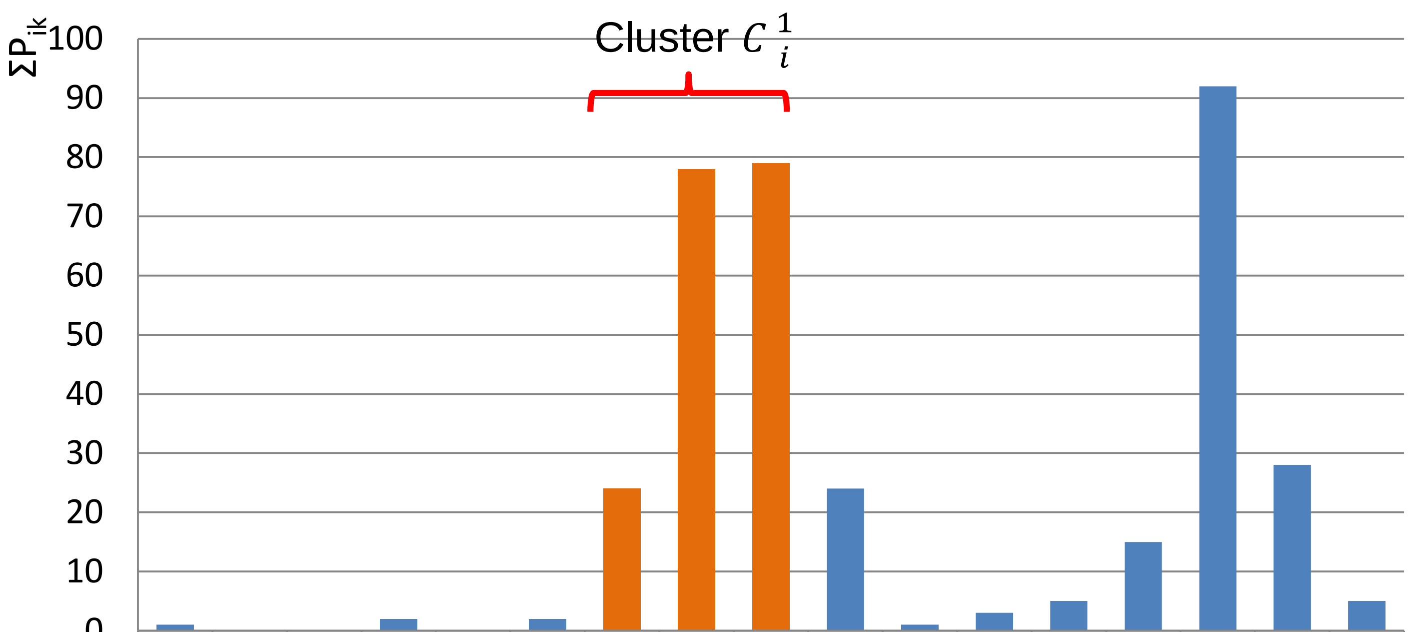
NR = 2
MinPts = *120*

$$\rho\ (P_{ik}) = \sum_{k=-73}^{-71} P_{ik} = 104 < 120$$

Wi-Fi clustering using DBSCAN

ComTec

Density-based Cluster Learning Algorithm

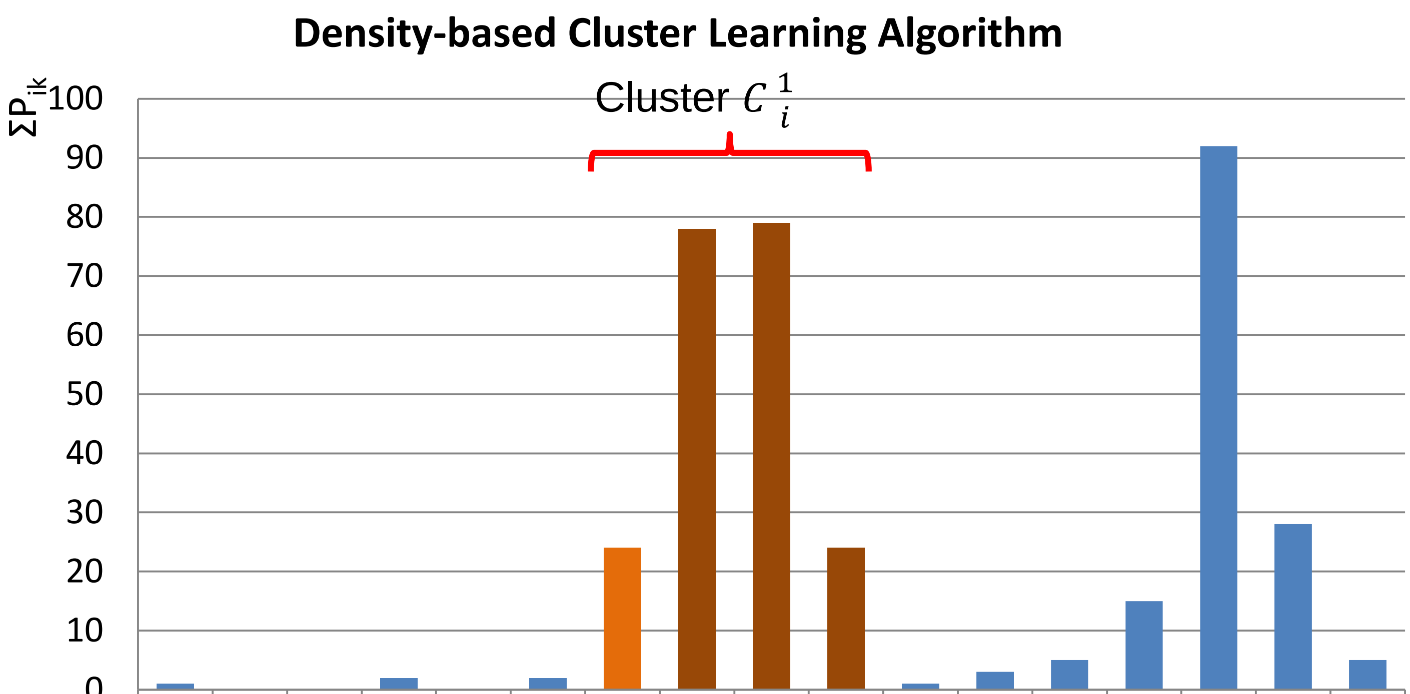


RSSI$_{ik}$ (dBm)	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62
$\sum_{k=m}^{m} P_{ik}$	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5

NR = 2
MinPts = 120

$$\rho\ (P_{ik}) = \sum_{k=-72}^{-70} P_{ik} = 181 \geq 120$$

Wi-Fi clustering using DBSCAN

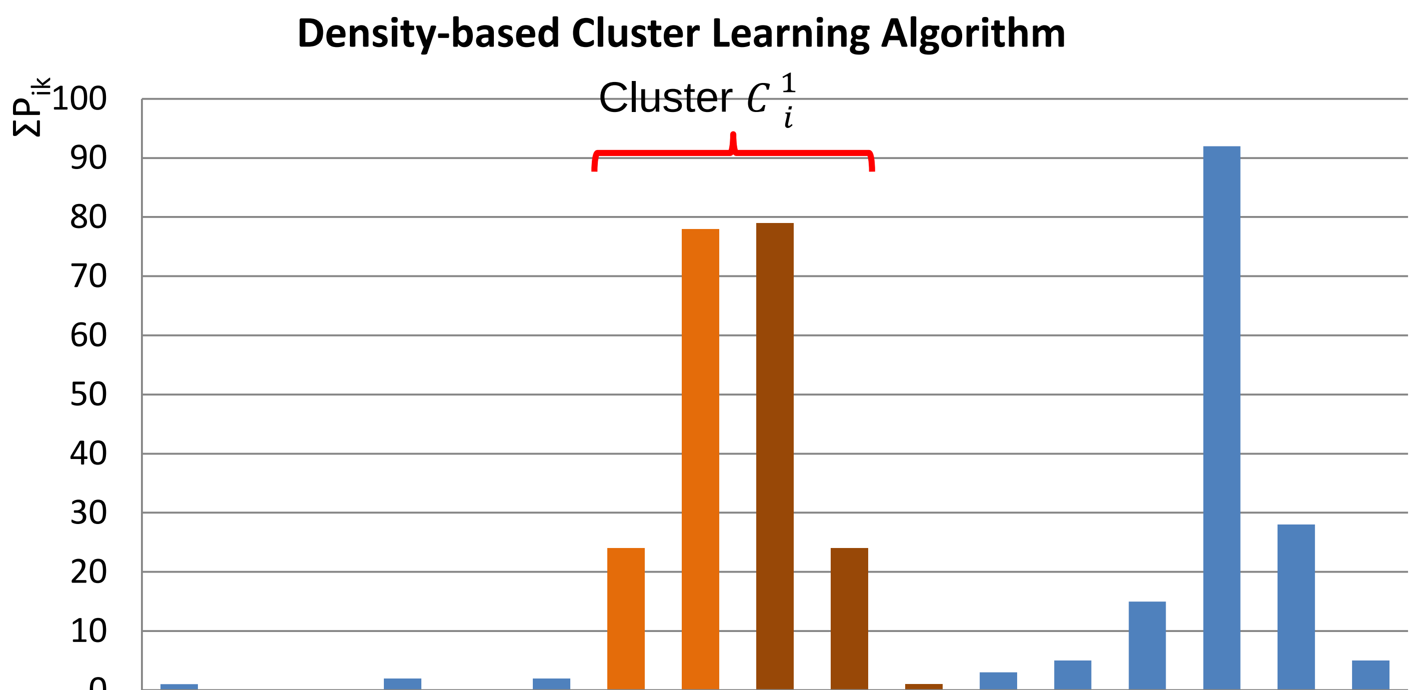


	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62	← $RSSI_{ik}$ (dB)
$\sum_{k=m}^{m} P_{ik}$ →	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5	

NR = 2
MinPts = 120

$$\rho\ (P_{ik}) = \sum_{k=-71}^{-69} P_{ik} = 181 \geq 120$$

Wi-Fi clustering using DBSCAN



RSSI$_{ik}$ (dBm)	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62
$\sum_{k=m}^{m} P_{ik}$	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5

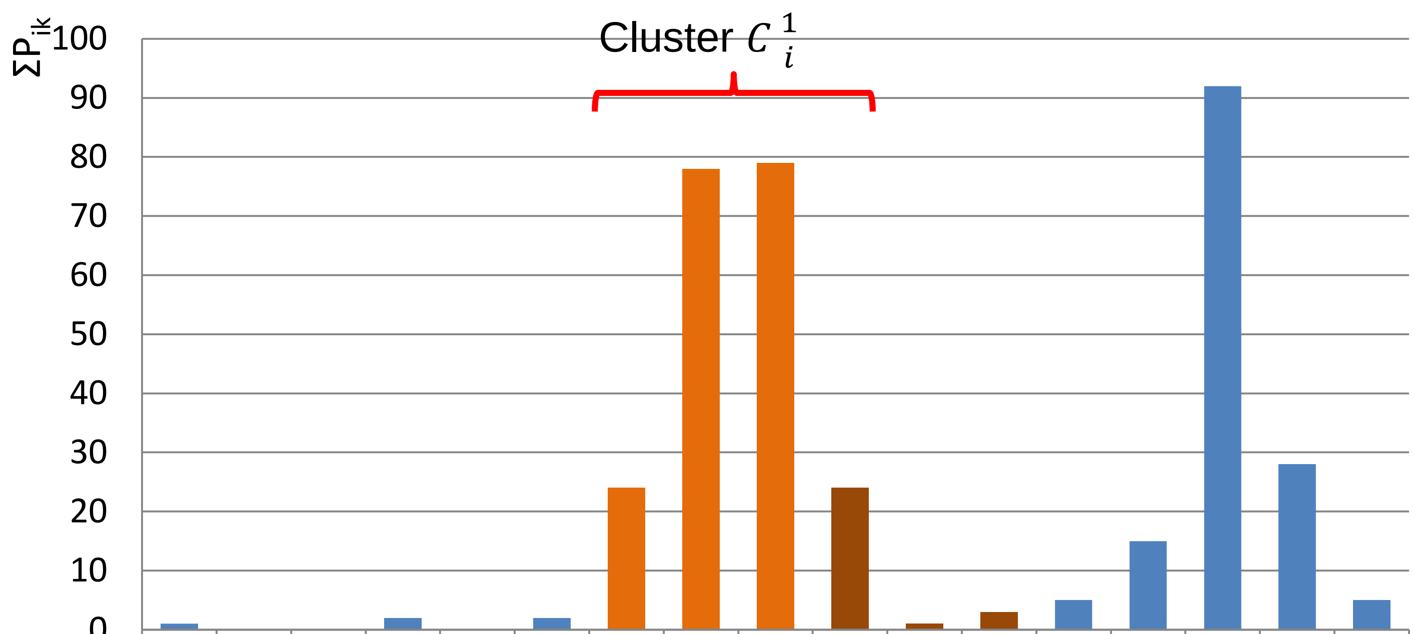
NR = 2
MinPts = 120

$$\rho\ (P_{ik}) = \sum_{k=-70}^{-68} P_{ik} = 104 < 120$$

Note: if $\rho\ (P_{ik}) < 120$, the density check will continue, since P_{ik} with RSSI$_{ik}$=-69 dB still belongs to the cluster C_i^1

Wi-Fi clustering using DBSCAN

Density-based Cluster Learning Algorithm



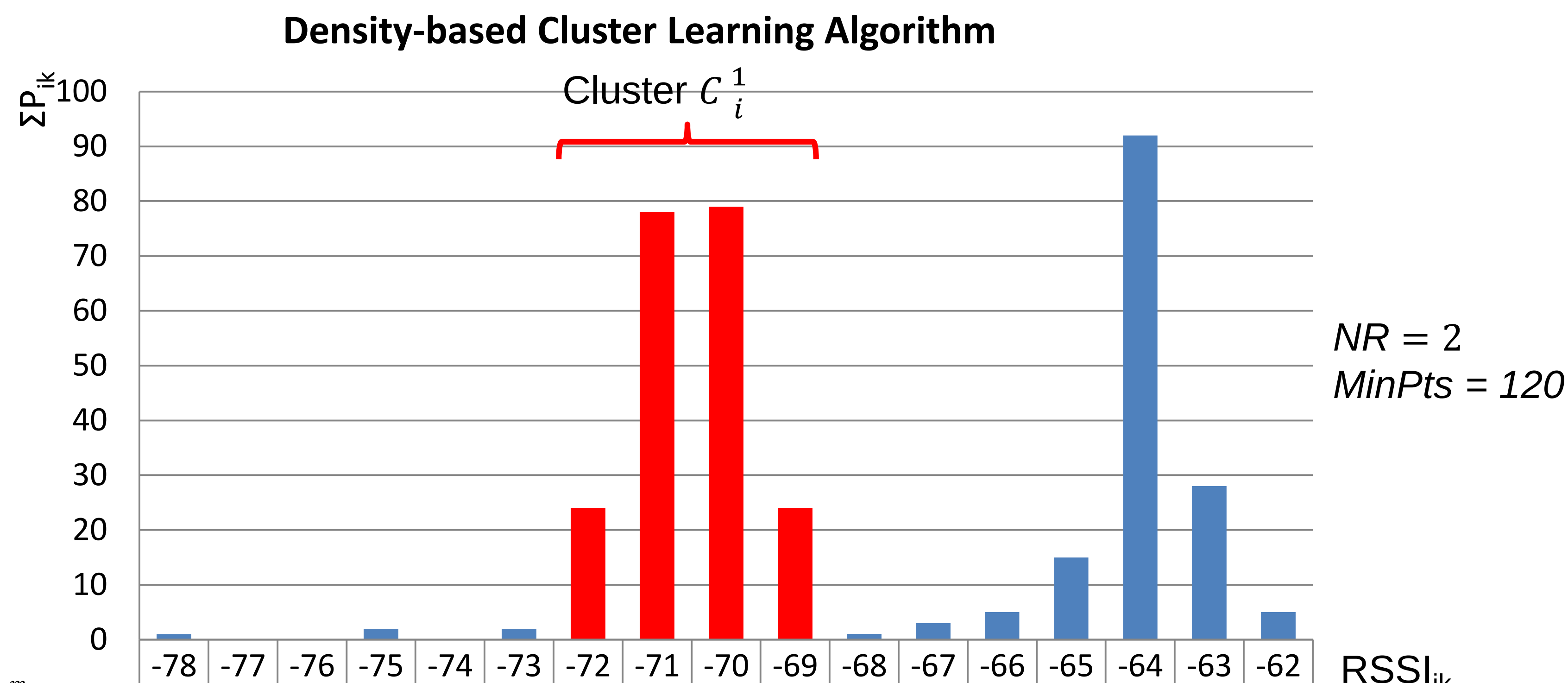
$RSSI_{ik}$ (dBm)	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62
$\sum_{k=m}^{m} P_{ik}$	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5

NR = 2
MinPts = 120

$$\rho\ (P_{ik}) = \sum_{k=-69}^{-67} P_{ik} = 28 < 120$$

Wi-Fi clustering using DBSCAN

ComTec

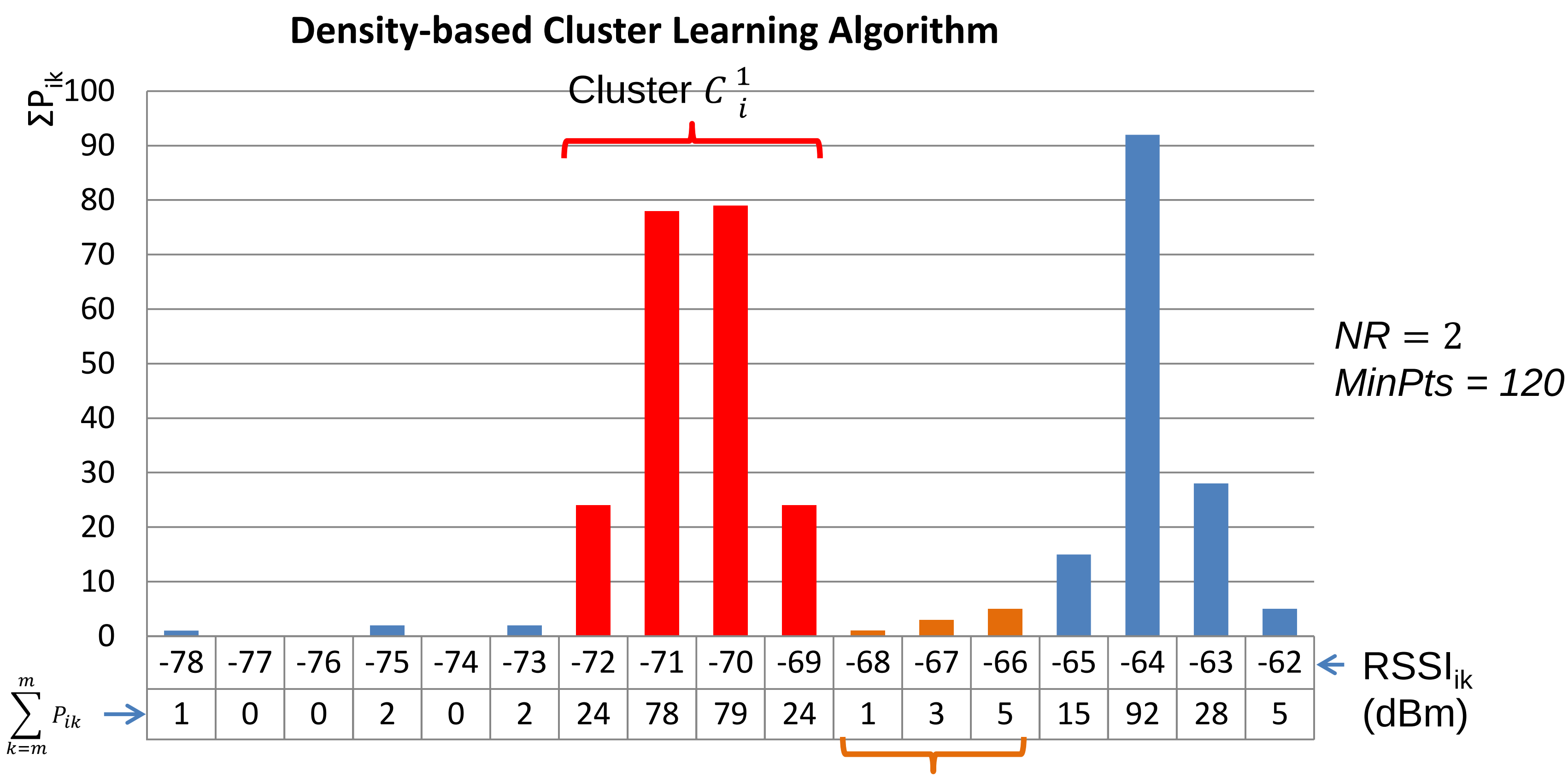


$RSSI_{ik}$ (dBm)	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62
$\sum_{k=m}^{m} P_{ik}$	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5

$$C_i^1 = \{MACi,\ -72dB, -69dB\}$$

Wi-Fi clustering using DBSCAN

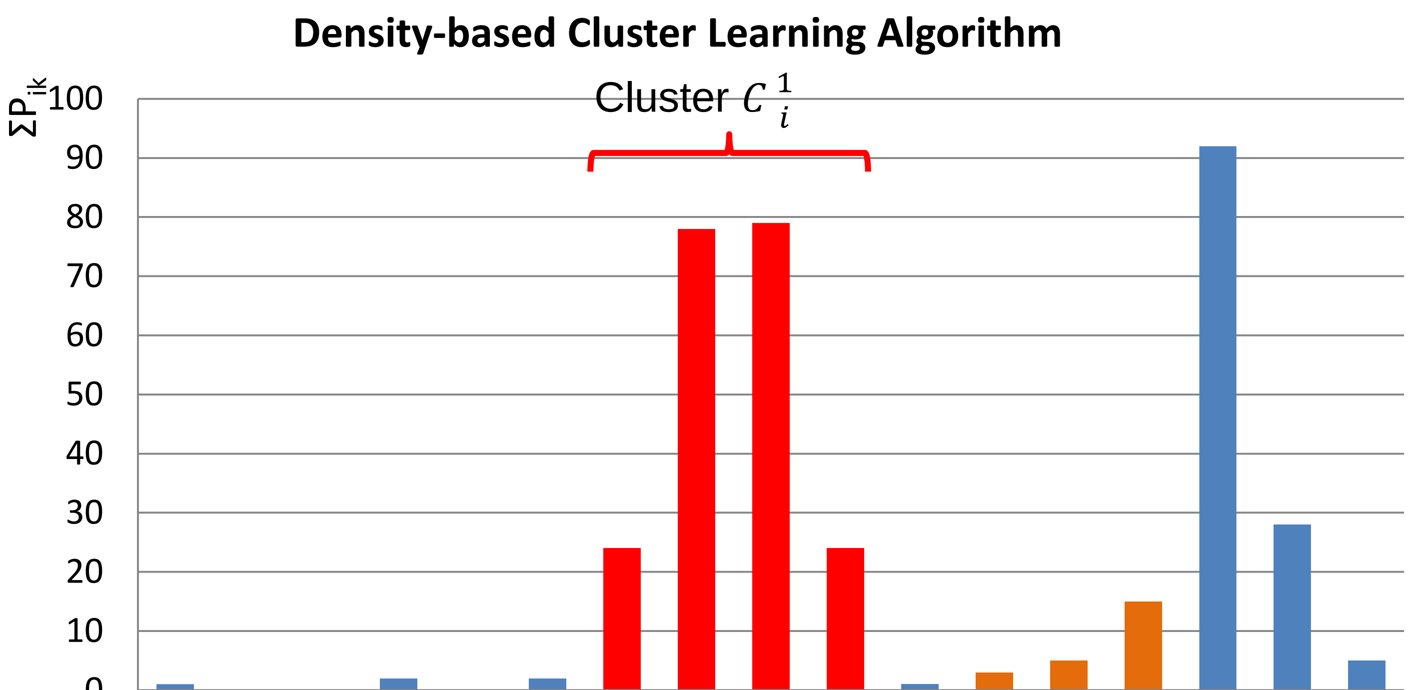
ComTec



$$\rho\,(P_{ik}) = \sum_{k=-68}^{-66} P_{ik} = 9 < 120$$

Wi-Fi clustering using DBSCAN

ComTec



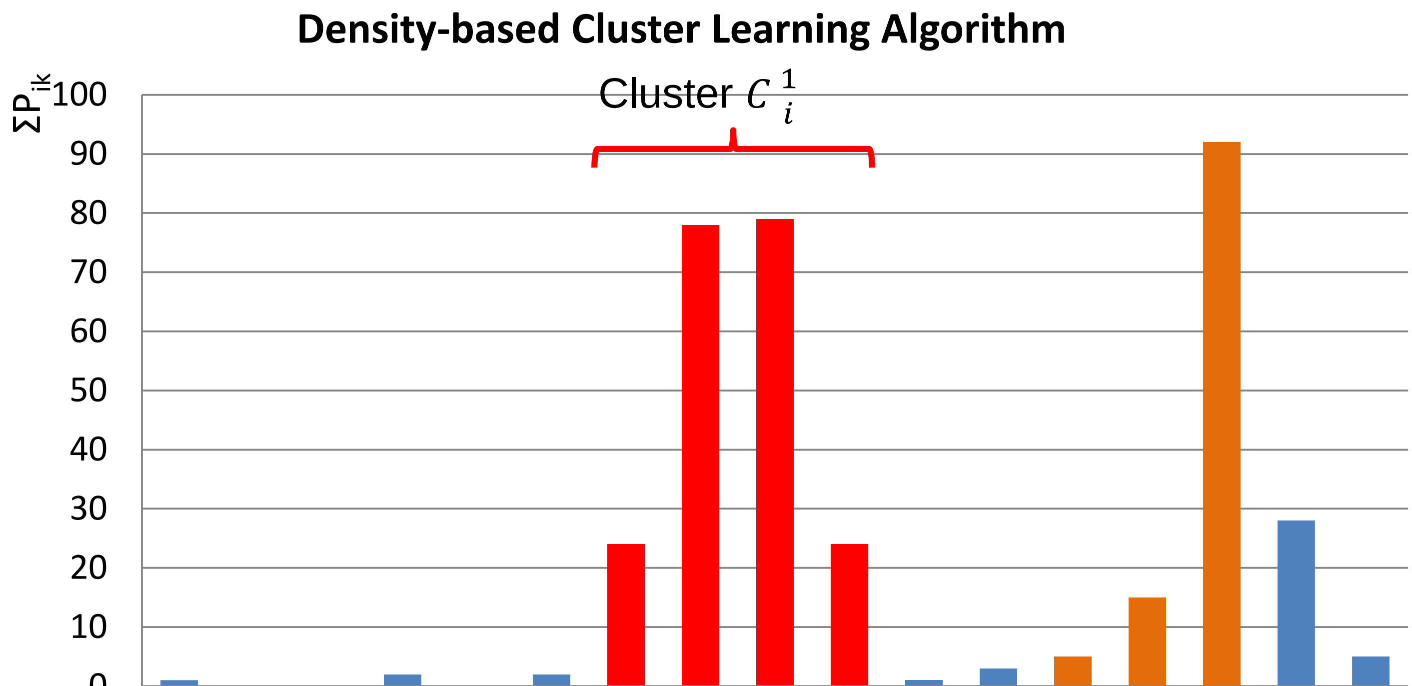
	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62	← $RSSI_{ik}$ (dBm)
$\sum_{k=m}^{m} P_{ik}$ →	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5	

NR = 2
MinPts = 120

$$\rho(P_{ik}) = \sum_{k=-67}^{-65} P_{ik} = 23 < 120$$

Wi-Fi clustering using DBSCAN

ComTec



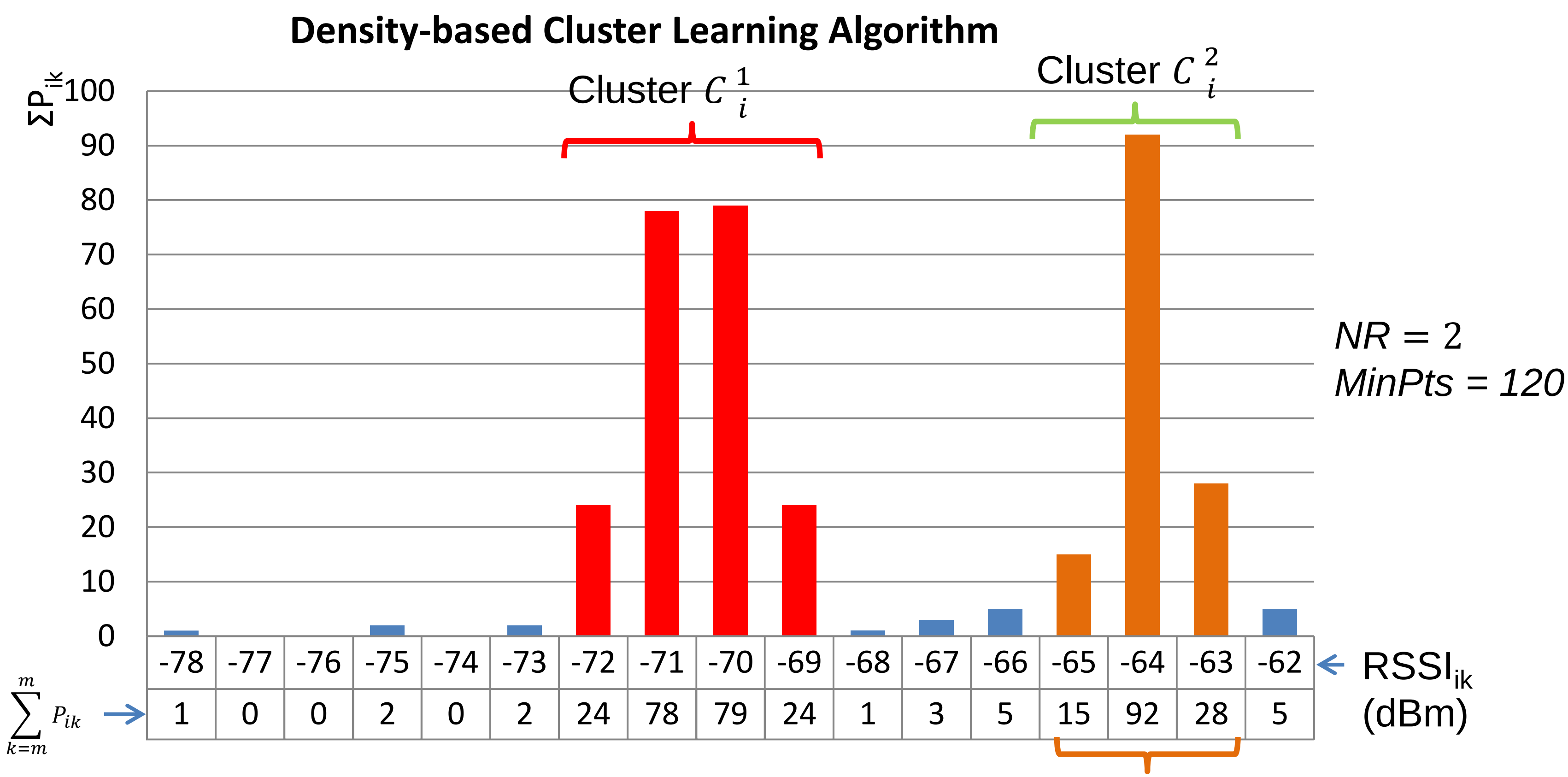
	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62
$\sum_{k=m}^{m} P_{ik}$	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5

$RSSI_{ik}$ (dBm)

NR = 2
MinPts = 120

$$\rho\ (P_{ik}) = \sum_{k=-66}^{-64} P_{ik} = 112 < 120$$

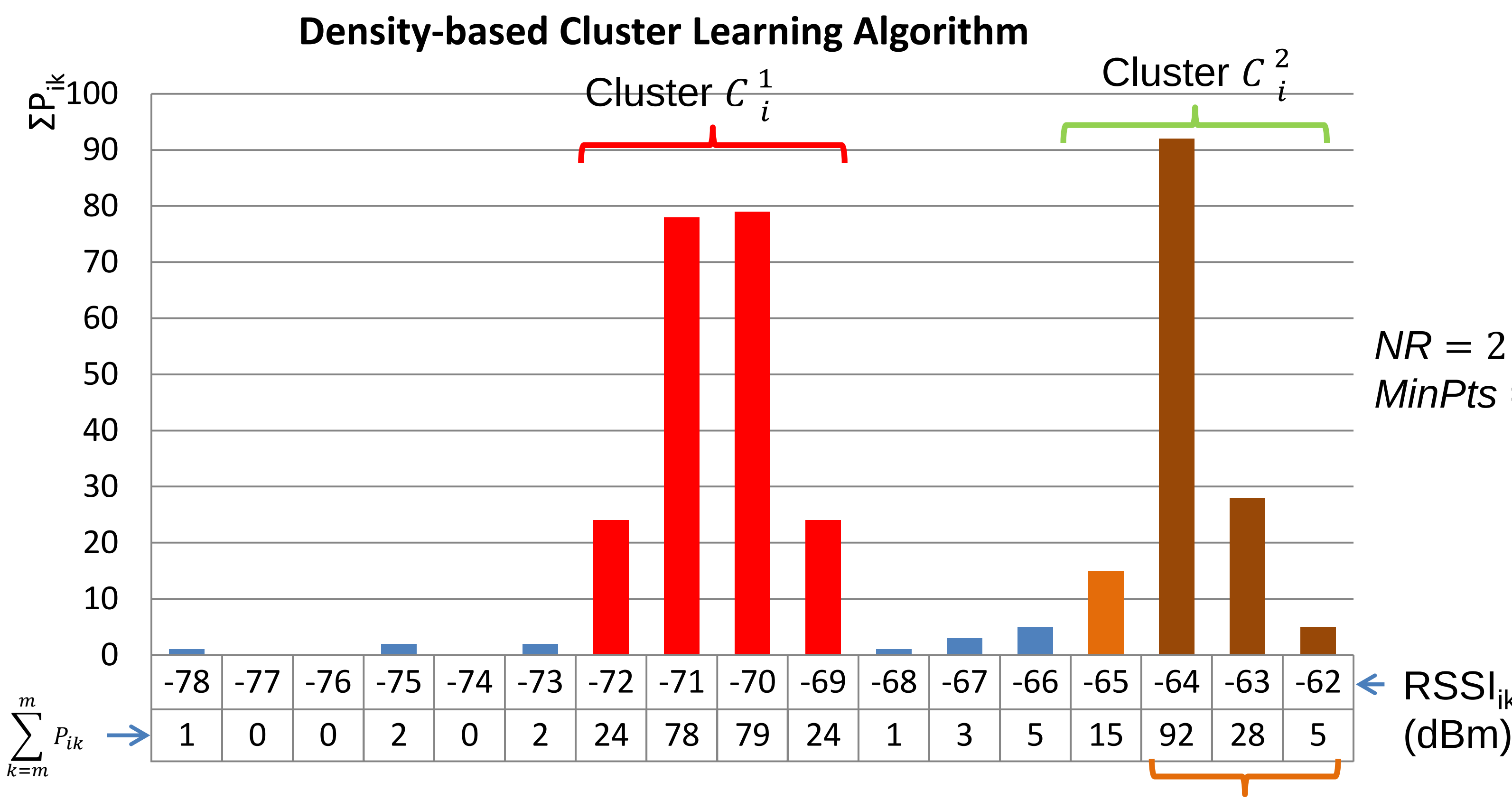
Wi-Fi clustering using DBSCAN



$$\rho\ (P_{ik}) = \sum_{k=-65}^{-63} P_{ik} = 135 \geq 120$$

Wi-Fi clustering using DBSCAN

ComTec

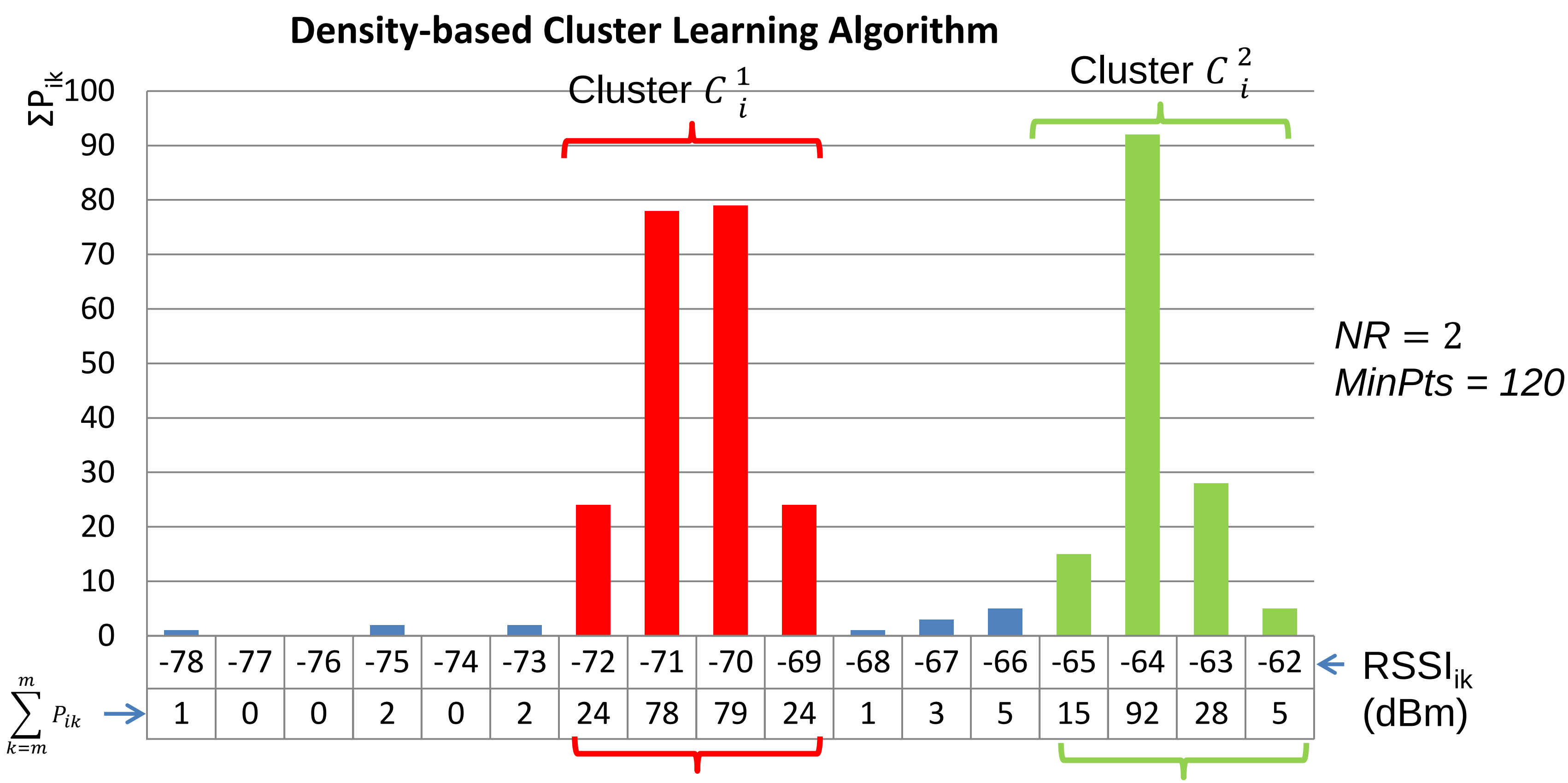


$RSSI_{ik}$ (dBm)	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62
$\sum_{k=m}^{m} P_{ik}$	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5

NR = 2
MinPts = 120

$$\rho(P_{ik}) = \sum_{k=-64}^{-62} P_{ik} = 125 \geq 120$$

Wi-Fi clustering using DBSCAN

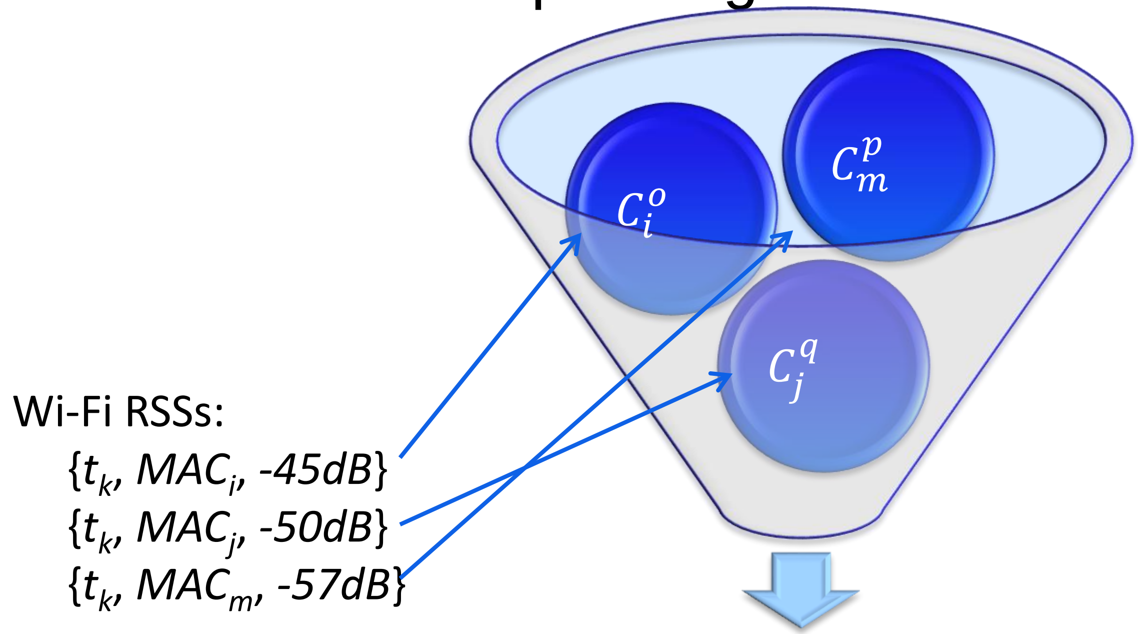


$RSSI_{ik}$ (dBm)	-78	-77	-76	-75	-74	-73	-72	-71	-70	-69	-68	-67	-66	-65	-64	-63	-62
$\sum_{k=m}^{m} P_{ik}$	1	0	0	2	0	2	24	78	79	24	1	3	5	15	92	28	5

$C_i^1 = \{MACi,\ -72dB, -69dB\}$ $\quad$ $C_i^2 = \{MACi,\ -65dB, -62dB\}$

Positioning phase

– Compare the current Wi-Fi RSSs with learned Wi-Fi fingerprints
– The user returns to a location corresponding to this fingerprint



$$FP_k = \left\{C_i^o, \dots, C_j^q\right\}$$

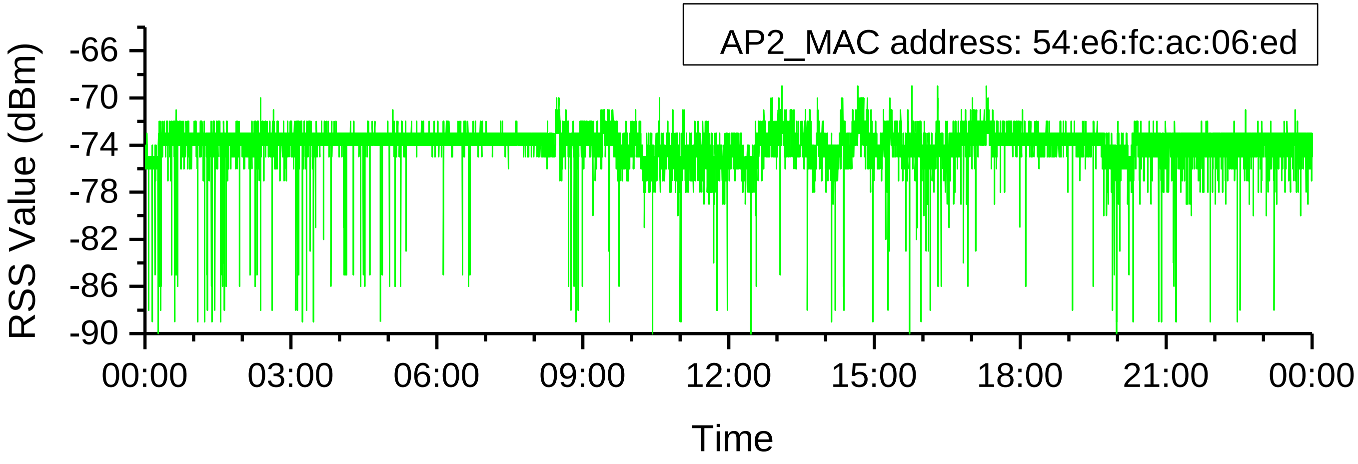
Source: [1]

Common Positioning approaches

- Nearest neighbor
  - Euclidean distance, Manhattan distance
- Probabilistic framework
  - Estimating the likelihood function, e.g., histogram, kernel, Gaussian
- Neural network
- Support vector machine
- Filtering approaches

RSS fluctuation

- Human presence
- Obstacles
- Change of environment

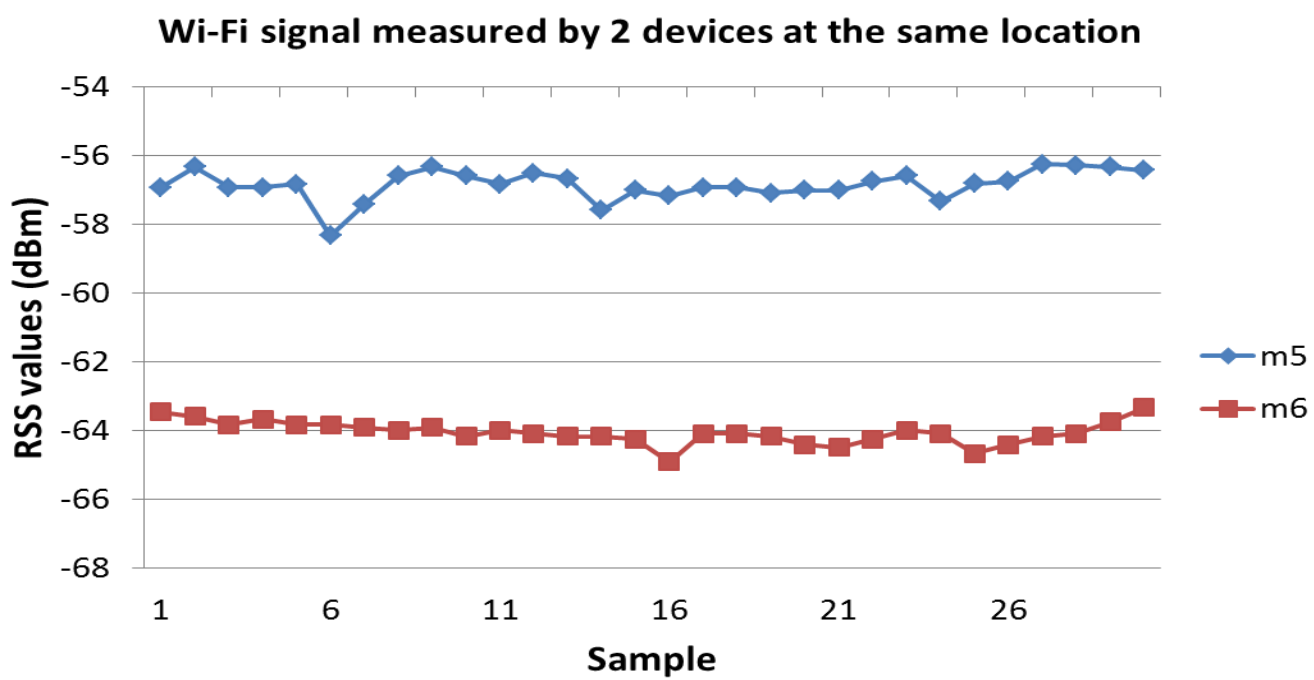


Source: [1]

Device heterogeneity

ComTec

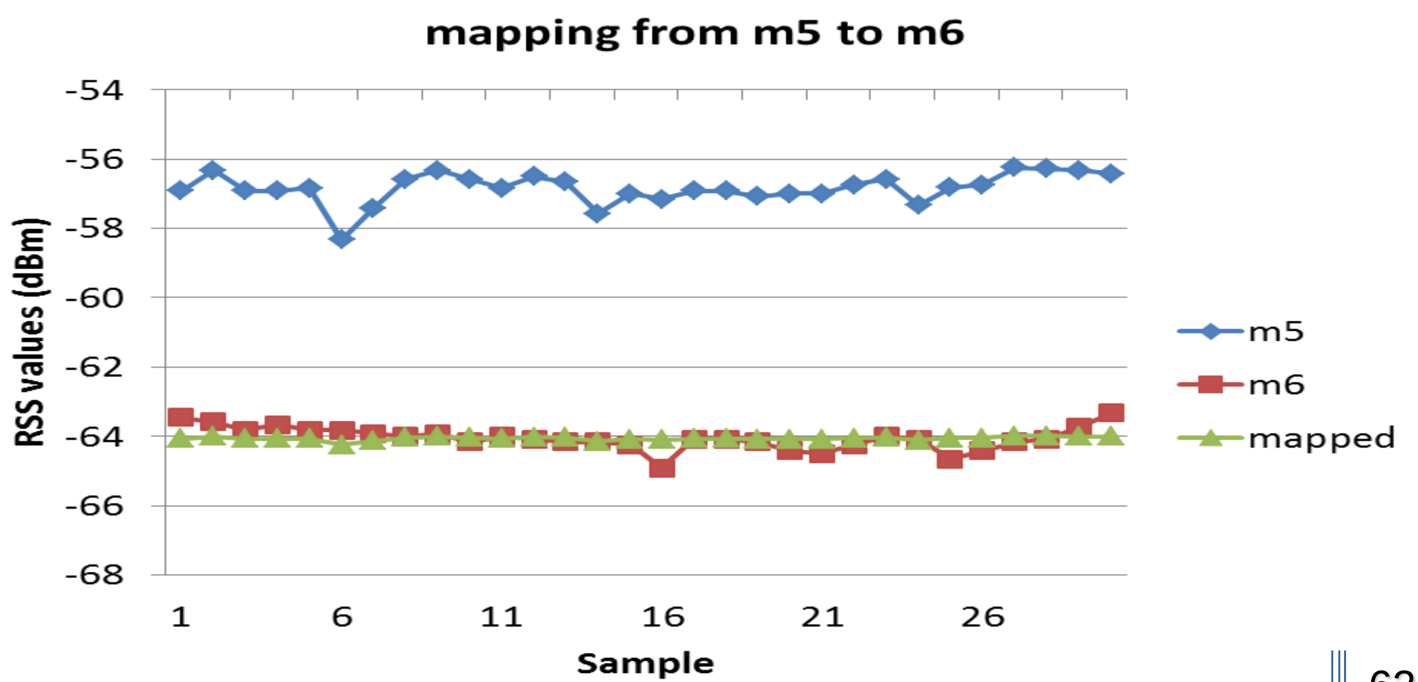
- Different devices are used in the training phase and positioning phase
- Those devices may have different Wi-Fi chipset, different antenna gain



Solving Device heterogeneity

ComTec

- Mapping device
  - Manual calibration
  - Linear transformation method
  - Ranking method



Solving Device heterogeneity

ComTec

- Extract features that do not depend on hardware
- Spatial mean normalization (SMN)

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \hat{x}_i = F(x_i|1,...,N)) = x_i - \bar{x}$$

- Signal strengths difference (SSD)

$$d(AP_i, AP_j) = RSS_i - RSS_j$$

- Hyperbolic Location Fingerprinting (HLF)

$$r(AP_i, AP_j) = \frac{RSS_i}{RSS_j} \qquad nlr(AP_i, AP_j) = \log r(AP_i, AP_j) - \log(\frac{1}{RSS_{max}})$$

2.4 GHz vs 5 GHz

ComTec

- The free space path loss index depends on the frequency of the signal

$$P_L\,(\mathrm{dB}) = 10\log_{10}\frac{P_t}{P_r} = -10\log_{10}\frac{G_l\lambda^2}{(4\pi d)^2}.$$

- The output power limit is different
- Different characteristic:
  - Scattering inside the building
  - Damping through water
  - Wall penetration
  - Fast fading

Other factors

- Energy consumption
- Response rate
- Fingerprint database
- Different frequency
- Number of scan per location estimate
- Influence of the number of Wi-Fi APs
- Combination with other sensors

References

1. Y. Xu, “Autonomous Indoor Localization Using Unsupervised Wi-Fi Fingerprinting,” PhD. Dissertation, Kassel University Press, 2015.
2. H. Liu, H. Darabi, P. Banerjee, and J.Liu, “Survey of Wireless Indoor Positioning Techniques and Systems”, IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, Vol. 37, No. 6, November 2007.
3. S. Das, T. Teixeira, and S. F. Hasan, “Research Issues related to Trilateration and Fingerprinting Methods”, International Journal of Research in Wireless Systems, Volume 1, Issue 1, November, 2012.
4. E. Ester, H. Kriegel, J. Sander and X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”, 2nd International Conference on Knowledge Discovery and Data Mining, 1996.
5. S. Shin, A. G. Forte, A. S. Rawat, and H. Schulzrinne, “Reducing MAC Layer Handoff Latency in IEEE 802.11 Wireless LANs”, Proceeding MobiWac 2004.
6. B. Bing (Editor), “Wireless Local Area Networks – The New Wireless Revolution,” Wiley-Interscience, 2002.
7. A. Goldsmith, “Wireless Communications,” Cambridge University Press, 2005.