

Article

Localization in Low Power Wide Area Networks using Wi-Fi Fingerprints

Thomas Janssen, Maarten Weyn and Rafael Berkvens *

IDLab, University of Antwerp – imec, Groenenborgerlaan 171, 2020 Antwerpen, Belgium

* Correspondence: rafael.berkvens@uantwerpen.be; Tel.: +32-3-265-1864

1 **Featured Application:** Localization of a wireless track and trace device communicating over
2 Sigfox and using nearby Wi-Fi access points as references.

3 **Abstract:** Supply chain management requires regular updates of the location of assets, which can
4 be enabled by low power wide area networks, such as Sigfox. While it is useful to localize a device
5 simply by its communication signals, this is very difficult to do with Sigfox because of wide area
6 and ultra narrowband nature. On the other hand, installing a satellite localization element on the
7 device greatly increases its power consumption. We investigated using information about nearby
8 Wi-Fi access points as a way to localize the asset over the Sigfox network, so without connecting
9 to those Wi-Fi networks. This paper reports the location error that can be achieved by this type of
10 outdoor localization. By using a combination of two databases, we could localize the device on all
11 36 test locations with a median location error of 39 m. This shows that the localization accuracy of
12 this method is promising enough to warrant further study, most specifically the minimal power
13 consumption.

14 **Keywords:** localization; internet of things; low power wide area networks; Wi-Fi; sigfox;
15 fingerprinting

16 1. Introduction

17 The location of an asset at a given time is an important piece of information in supply chain
18 management [1–3]. Hillbrand and Schoech [4] divide tracking and tracing of assets into two categories:
19 discrete and continuous. The discrete category relies on barcodes and Radio Frequency Identification
20 (RFID) approaches to know when a specific asset is at a predefined location. The continuous category
21 relies on a Global Navigation Satellite System (GNSS) such as the Global Positioning System (GPS) to
22 establish the location of the asset and a cellular data connection to communicate this location with
23 the management software. Kandel *et al.* [5] define a subcategory to continuous tracking, called the
24 quasi-continuous tracking. In this subcategory, the management software continuously tracks a vehicle,
25 while the vehicle has knowledge about the assets it carries through discrete tracking. Oliveira *et al.* [6]
26 present and discuss a number of recent quasi-continuous tracking solutions.

27 In an Internet of Things scenario, assets would have a permanent connection to the internet
28 themselves, rather than through a hub in the transportation vehicle. This allows the owner of the
29 assets to be independent of the transporter. While this connection can in theory be a cellular data
30 connection, in practice the energy requirements for such connections limit its applicability only to
31 assets that are large enough to carry additional batteries for this connection. Low Power Wide Area
32 Networks (LPWAN) propose to reduce the energy requirement of the mobile data connection, albeit at
33 reduced data rates [7].

34 Sigfox [8] is one of the major players in the LPWAN market. It uses an ultra narrow band (UNB)
35 channel of 100 Hz with a Differential Binary Phase Shift Keying (BPSK) modulation operating on
36 the 868 MHz telecommunication frequency in Europe. This frequency does not require a license for
37 operating Short Range Devices (SRD), defined as devices with a limited transmission power. To comply
38 with these regulations, a Sigfox device can only transmit 140 messages of 12 B each on a single day at
39 100 bit/s. The communication range is around 30 km outside cities and 3 km inside cities [7,9]. In this
40 article, we present our idea of enabling localization of a device using the 12 B payload limit of Sigfox.

41 Localization based on wireless communication can be split into three categories: proximity;
42 geometry, or trilateration or triangulation; and pattern recognition, also referred to as scene analysis or
43 fingerprinting [10]. Proximity localization assumes that a mobile unit is at or near the location of the
44 gateway with which it is connected. It depends on the range of the communication method. Proximity
45 localization is often used in Radio Frequency Identification (RFID) scenarios, which have a limited
46 range [11]. If the range increases, the mobile unit can be in communication with multiple gateways,
47 which allows more sophisticated estimation of the mobile unit's location. These types of estimation
48 quickly approach the trilateration technique. Trilateration, a type of geometric localization, estimates
49 the distance between the mobile unit and multiple gateways. These distance estimates can be based
50 on the time of flight of the signal [12], or based on the attenuation of the signal when received by
51 either the mobile unit or the gateways [13]. Triangulation, as opposed to trilateration, estimates the
52 angle of arrival of the received signal through an antenna array. Pattern recognition, most commonly
53 known in the localization community as fingerprinting, consists of comparing a snapshot of the current
54 environment with a database of such snapshots linked to known locations [14].

55 Wi-Fi fingerprinting is excellently reviewed by He and Chan [15]. It consists of an offline phase,
56 during which a site survey is conducted, resulting in a fingerprint database, and an online phase, during
57 which a mobile unit is actually localized by comparing its fingerprint measurements with the database.
58 A fingerprint, in the case of Wi-Fi, is a list of all nearby Wi-Fi access points, paired with their received
59 signal strength at the mobile unit. One thing to consider in Wi-Fi fingerprinting is in how to compare
60 the fingerprint created by the mobile unit with the fingerprints in the database. Torres-Sospedra
61 *et al.* [16] provide an overview of 51 distance metrics that could be used for this comparison. Another
62 consideration in Wi-Fi fingerprinting is the challenge of constructing the fingerprint database. While the
63 traditional approach is to collect the fingerprints during a labor intensive site survey, some researchers
64 have looked into crowd sourcing the database construction [17].

65 Since Sigfox is an ultra-narrowband technology, localization based on its communication signal
66 is very challenging. Gezici *et al.* [18] show that the Cramér-Rao Lower Bound (CRLB) for estimating
67 the range between mobile device and base station through the time delay of the signal is reversely
68 proportional to the bandwidth. They also indicate that the CRLB for estimating that range through
69 the received signal strength is proportional to the distance between the mobile unit and the base
70 station; since Sigfox is a wide area network, we can assume that the CRLB will generally be rather
71 high. Sallouha *et al.* [19] demonstrate a classification or fingerprinting solution to Sigfox localization,
72 employing GPS enabled devices to create the fingerprints. They achieve a classification accuracy of
73 78 % in fingerprints with a radius of 150 m, separated by 2.5 times this distance from other fingerprints.

74 In this article, we present our approach to Sigfox localization, based on sensing nearby Wi-Fi
75 access points. Due to the limitations of Sigfox, we can transmit only two Basic Service Set Identifiers
76 (BSSID), which are six bytes each. This leaves no payload space to transmit any received signal strength
77 information. Since we do not wish to invest time in constructing the Wi-Fi access point database,
78 we will use online crowd sourced services. While crowd sourced services have obvious drawbacks,
79 especially being noisy, Sapiezynski *et al.* [20] achieved a 15 m median location error for 73 % of their
80 test scans. The aim of the research in this article is to establish the expected location error for our
81 approach.

82 The article continues as follows. Section 2 outlines the localization approach, discusses the online
83 Wi-Fi databases, and presents the measurement locations. Section 3 describes the localization results,

84 and the frequency of missing an access point. Section 4 analyses and discusses those results. Finally,
85 Section 5 draws the conclusion.

86 2. Methods and Materials

87 This section is split into the localization approach and a discussion of the online Wi-Fi access
88 point databases. It also includes an overview of the test locations.

89 2.1. Localization Approach

90 The general approach of Wi-Fi fingerprinting localization can be divided into two steps: the
91 training and operational step. The first, training step is to discover networks or Wireless Access Points
92 (WAP) and store their BSSID in a database, usually linked with the received signal strength at that
93 location. There is no need to connect the Wi-Fi device with these access points. The second, operational
94 step is requesting from the database the location of newly discovered WAP during operation. The
95 database will reply with the location of each WAP. We will not construct the database ourselves, but
96 utilize databases that are available online; thus, we will only concern ourselves with the operational
97 step.

98 To discover nearby access points, a laptop or a smartphone with a Wi-Fi adapter is required.
99 A Python script saves the BSSID and received signal strength of each discovered access point. In
100 Windows, this can easily be done by running the following command: `netsh wlan show networks`
101 `mode=bssid`. The Python script then saves the output to a file, together with the current location. The
102 current location is determined by a GPS module. To be more accurate, all locations are chosen to be
103 recognized on a map, so that a visual check of the GPS location can be performed. This procedure is
104 executed on 36 locations in and around Antwerp, Belgium.

105 When the data is collected, a HTTP-request to a database will be performed for every BSSID.
106 Afterwards, the database will respond with a message, which includes the coordinate of the access
107 point. If the BSSID is not found in the database, an error message will be sent. Subsequently, all possible
108 combinations of two BSSIDs are generated and the mean coordinate of every pair is calculated. If only
109 one of the two BSSIDs was found in the database, the coordinate of that BSSID will be considered as
110 the mean coordinate. Finally, the distance error can be calculated. This is the distance between the
111 coordinate of the position where the measurement took place and the mean coordinate of the pair
112 of BSSIDs. This distance is calculated using the Haversine function, which takes into account the
113 curvature of the Earth [21].

114 Additionally, we calculate the probability of having no match of the BSSID in the database:

$$P(-nB) = \frac{\binom{-B}{n}}{\binom{W}{n}}, \quad (1)$$

115 where $P(\cdot)$ is the probability operator; W is the number of access point seen at a location; B are
116 the BSSIDs in W that are found in the database, $-B$ are the BSSIDs that are not found in the database;
117 n is the number of BSSIDs that can be used for localization—in our case, this is maximum 2 because of
118 the limitations of Sigfox communication; $-nB$ are the combinations of n BSSIDs that are not found in
119 the database. This probability is interesting, because if none of the BSSIDs of the Wi-Fi access points
120 that are seen at a location can be found in the database, the location of the device cannot be estimated.

121 In an attempt to improve the accuracy of the results, we develop an additional strategy: to send
122 only the combination of BSSIDs that have the strongest signal strength. It is not possible to send the
123 RSSI values together with the BSSIDs in one Sigfox message, but we can use it in the sending device to
124 determine the right combination of BSSIDs. Although this strategy is likely to consume more energy
125 from the device (to pick the right combination of BSSIDs), it may decrease the mean location errors.

Table 1. Summary of largest Wi-Fi fingerprinting databases

Database name	Amount of Wi-Fi access points	Reference
Combrain Positioning Service	1,296,000,000	[22]
LocationAPI by UnwiredLabs	1,500,000,000	[23]
Mozilla Location Service	30,000,000,000	[24]
Navizon	<i>unknown</i>	[25]
WiGLE	4,750,076,107	[26]

126 Another improvement is attempted by combining the WiGLE and LocationAPI databases. The
 127 most straightforward combination is to take the average of the locations returned by each database. If a
 128 database does not return a location for one or both access points, these locations are simply disregarded
 129 in the calculation of the average. Since LocationAPI returns a single location for two BSSIDs, we
 130 account for that location twice, so that it receives an equal weight compared to the two results from
 131 the WiGLE database. In fact, we do not know if LocationAPI knows about only one of the two access
 132 points; in this case the weight of that location is erroneously doubled, but this cannot be prevented.

133 This straightforward combination neglects the fact that the location error from the LocationAPI
 134 database is generally much better than that of the WiGLE database. Thus, another approach for
 135 combination is to choose the location provided by the LocationAPI database, unless the BSSIDs could
 136 not be found in that database; in which case the WiGLE database is consulted for a location estimate.
 137 While we will show the location error results of the first approach, this second combination approach
 138 will be analyzed further.

139 2.2. Wi-Fi Databases

140 There is a variety of fingerprinting databases to work with. Some companies send their so called
 141 wardrivers all over the world to gather information about Wi-Fi access points. However, there is a
 142 trend to create more and more crowd-sourced databases [20]. People can just create a free account and
 143 upload data of the access points in their neighborhood. The most popular and largest databases are
 144 summarized in Table 1.

145 Databases with information about Wi-Fi access points are growing at a significant rate. This can
 146 be stated by observing that, for example, the Navizon database is growing at a rate of 500,000 data
 147 points every day [27]. It is interesting to note that this database also keeps track of the amount of cell
 148 phone towers. One of the many reasons of this significant rate is the Internet of Things (IoT) and the
 149 world of ubiquitous computing we are living in today. Navizon even pays people [28] to contribute to
 150 the database. The one million registered users can just download the app, drive around in their car
 151 and earn a pay-out of \$15 for every 10,000 access points.

152 One database used for this research is called WiGLE [26]. WiGLE is short for Wireless Geographic
 153 Logging Engine. The website is an open source platform; everyone can view registered BSSIDs around
 154 the world on a map. If a free account is created, requests can be performed and data can be uploaded
 155 to the database. At this moment, WiGLE wardrivers have discovered over 4.5 billion Wi-Fi access
 156 points around the world. This database is chosen because it is easy to work with and has a decent
 157 amount of access points.

158 Another database used for this research is LocationAPI [23]. LocationAPI is a website hosted by
 159 UnwiredLabs, an Indian corporation focusing on positioning systems. At this moment, LocationAPI
 160 wardrivers have discovered over 1.5 billion Wi-Fi access points worldwide. Its database contains a
 161 large number of wireless access points in Belgium.

162 To obtain an objective look at the error or accuracy of Wi-Fi localization through a Sigfox network,
 163 the characteristics of both databases are compared. WiGLE has a larger worldwide database. This does
 164 not mean that the WiGLE results are more accurate, because there are more access points discovered in
 165 the USA than in Belgium. WiGLE claims wardrivers have discovered 1.7 million WAPs in Belgium. On

```

"resultCount":1,
"last":1,
"success":true,
"results":[{"trilat":51.17709351,
            "trilong":4.41607475,
            "ssid":"UAntwerpen",
            "qos":1,
            "transid":"20110605-00081",
            "firsttime":"2011-05-31T17:11:13.000Z",
            "lasttime":"2013-10-28T11:32:52.000Z",
            "lastupdt":"2013-10-28T09:33:29.000Z",
            "netid":"00:0b:86:26:86:16",
            "name":null,
            "type":"infra",
            "comment":null,
            "wep":"2",
            "channel":6,
            "bcninterval":0,
            "freenet":"?",
            "dhcp":"?",
            "paynet":"?",
            "userfound":false}],
"first":1

```

Figure 1. WiGLE returns detailed information about an access point.

```

'status': 'ok',
'balance': 89,
'lat': 51.17786625,
'lon': 4.42454131,
'accuracy': 10

```

Figure 2. LocationAPI returns only the average location of multiple BSSIDs, but also an estimated accuracy.

166 the other hand, while LocationAPI contains a smaller amount of access points, this database includes
 167 11 million WAPs in Belgium.

168 A request to the WiGLE database results in a string with a lot of information. It returns the latitude
 169 and longitude, but also data like security details and uptime indicators, see Figure 1. LocationAPI
 170 returns less information, but it is important to note that the data it does return is more useful: each
 171 result includes the mean coordinate of multiple BSSIDs, plus the estimated accuracy in meters. A nice
 172 extra feature is the 'balance' parameter. This informs us on how many successful requests we have left,
 173 see Figure 2. However, it is not possible to request the location of a Wi-Fi access point in LocationAPI.

174 The documentation of both services helps to get started with performing requests. However,
 175 the documentation of WiGLE is harder to find and understand, since it is still under development.
 176 The LocationAPI documentation is very easy to find, more human readable and it contains examples.
 177 Table 2 presents a summary of all the properties and characteristics of the two databases.

178 2.3. Measurement locations

179 Measurements are executed at 36 different locations in and around Antwerp. The locations are
 180 chosen so that there is a variety of more and less densely populated areas. At every location, a script

Table 2. Comparison of the properties and characteristics of the WiGLE and LocationAPI databases

	WiGLE	LocationAPI
Amount of access points worldwide	4,750,076,107	1,500,000,000
Amount of access points in Belgium	1,700,000	11,000,000
Accuracy given in request	No	Yes
Balance given in request	No	Yes
Number of BSSIDs in a request	1 BSSID	At least 2 BSSIDs
Number of coordinates in a response	1 coordinate per WAP	1 mean coordinate
Documentation	More complex	More intuitive

181 is run to save all BSSIDs and RSSI values of the access points in the vicinity, together with the exact
 182 coordinate of the measurement location. Figure 3 shows a map of the different measurement locations.

183 3. Results

184 This section summarizes the location estimate results by database. First the WiGLE database,
 185 then the LocationAPI database, and finally the combination of both databases. For each database, we
 186 discuss the overall location error, the probability of finding the BSSIDs, and the location error when
 187 transmitting only the two BSSIDs with the highest received signal strength.

188 3.1. WiGLE Results

189 The location estimates when localizing using the WiGLE database are prone to large errors, see
 190 Figure 4. The overall mean location error is 6.24 km, while the overall median location error is 77 m.
 191 The 75th percentile location error is larger than 1 km only at locations 12 and 14; at location 14, it is
 192 5.12 km. Ten locations have an error greater than 2 km, which are clipped in Figure 4. One of those is
 193 location 8. The mean location error is 101 km and the median location error is 45 m. To represent the
 194 measurement, a cumulative distribution of the errors at this location is shown in Figure 5. There are
 195 some combinations of BSSIDs with a significant location error: 802 km and 1633 km. However, more
 196 than 90 % of the BSSID combinations have an error smaller than 119 m, see Figure 5b.

197 The average probability of finding the BSSIDs in the WiGLE database of a single WAP at one of
 198 the 36 locations is 32.6 %. This is calculated using Equation (1) with $n = 1$. The average probability of
 199 finding neither of two BSSIDs in the WiGLE database is 13.2 %, which is calculated with $n = 2$ or:

$$P(-2B) = \frac{\binom{-B}{2}}{\binom{W}{2}}. \quad (2)$$

200 A graph of the probability of finding neither BSSID in the WiGLE database is shown in Figure 6.
 201 When only sending the combination of BSSIDs with the highest RSSI values to the WiGLE
 202 database, the mean error of all measurement locations decreases from 6.24 km to only 103 m. The
 203 median error decreases from 77 m to 66 m. The probability that neither of those BSSIDs could be found,
 204 however, is 11.1 %; no location can be estimated at those locations. Figure 7 shows the distribution of
 205 the error.

206 3.2. LocationAPI Results

207 The location estimates when localizing using the LocationAPI database are much less prone to
 208 large errors than with the WiGLE database, see Figure 8. The overall mean location error is 21.06 km,
 209 while the overall median location error is 37 m. The 75th percentile location error is never larger than
 210 0.5 km. Seven locations have an error greater than 2 km, which is clipped in Figure 8. One of those is
 211 location 1. The mean location error is 546 km and the median location error is 23 m. The cumulative
 212 distribution of the errors at this location is shown in Figure 9. There is one combination of BSSIDs with

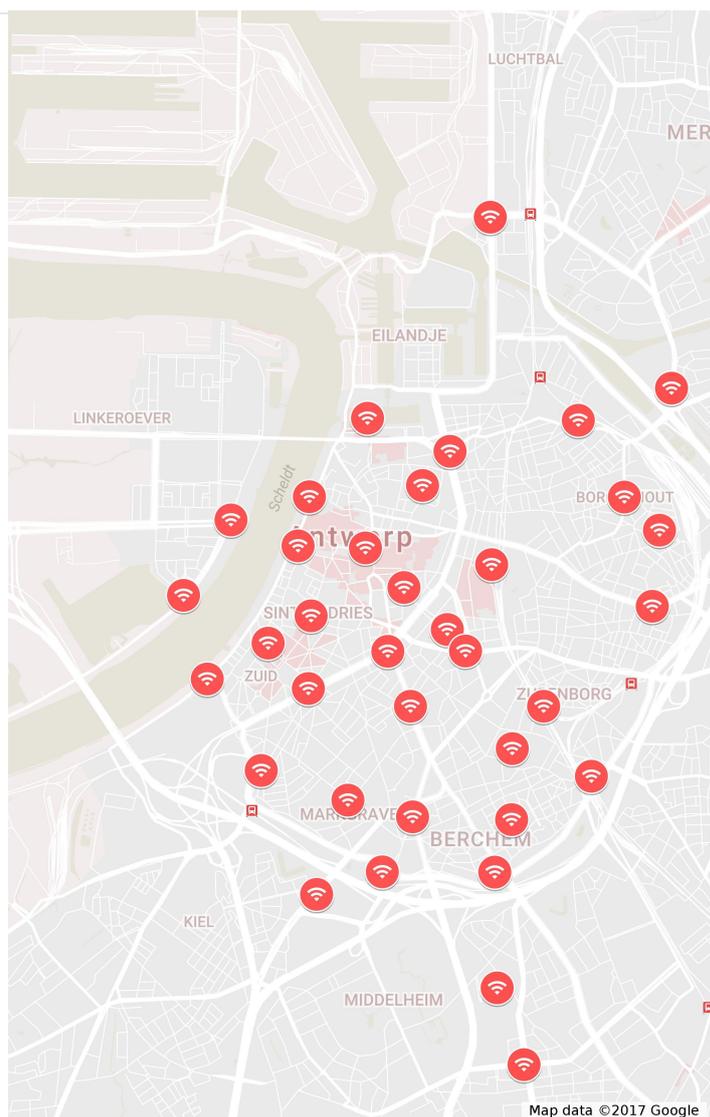


Figure 3. Measurement locations in Antwerp, ©2017 Google. See Figure S1 for an interactive version.

213 a significant location error: 8543 km. However, more than 93.6 % of the BSSID combinations have an
 214 error smaller than 43 m, see Figure 9b.

215 The average probability of finding neither BSSIDs in the LocationAPI database of a combination
 216 of WAP seen at one of the 36 location is 12.5 %. This is calculated as in Equation (2). Since it is not
 217 possible to request a single BSSID to the LocationAPI database, the chance of match of a single BSSID
 218 cannot be calculated. A graph of the probability of finding neither BSSIDs in the LocationAPI database
 219 is shown in Figure 6.

220 When only sending the combination of BSSIDs with the highest RSSI values to the LocationAPI
 221 database, the mean error of all measurement locations decreases from 16 km to only 114 m. The
 222 median error decreases slightly, from 37 m to 33 m. The probability that neither of those BSSIDs can be
 223 found, however, is again 19.4 %; no location can be estimated at those locations. Figure 7 shows the
 224 distribution of the error.

225 3.3. Combination of WiGLE and LocationAPI

226 A straightforward solution to improve the results is to combine the WiGLE and LocationAPI
 227 databases. As a first approach, we propose to combine the location estimates from both database by
 228 taking their average; this result can be seen in Figure 10. The overall mean location error is 31.79 km,

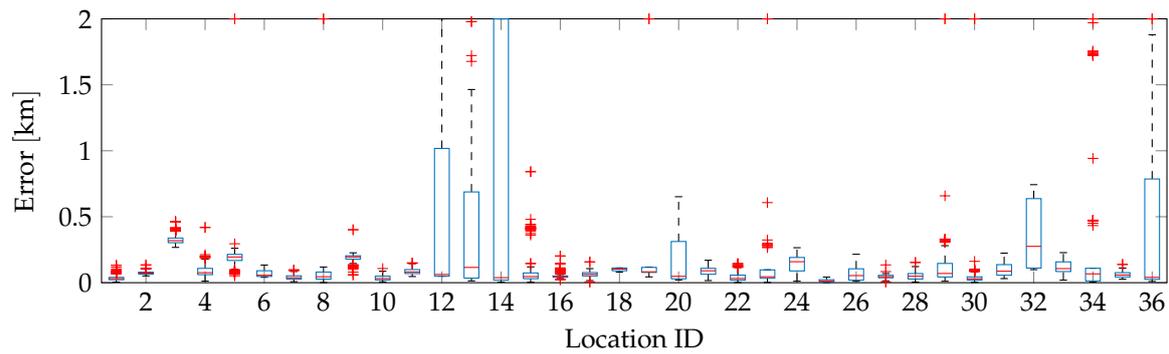


Figure 4. Overview of the location error on all 36 locations using the WiGLE database. Errors larger than 2 km are clipped.

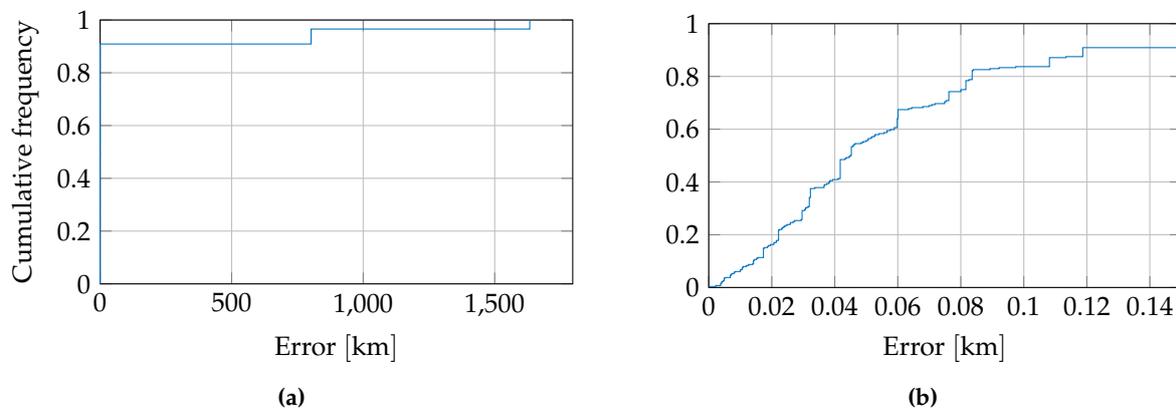


Figure 5. Cumulative distribution of the location error at the eighth measurement location, using the WiGLE database. A detail of Figure (a) is shown in Figure (b). The mean location error is 101 km, the median location error is 45 m.

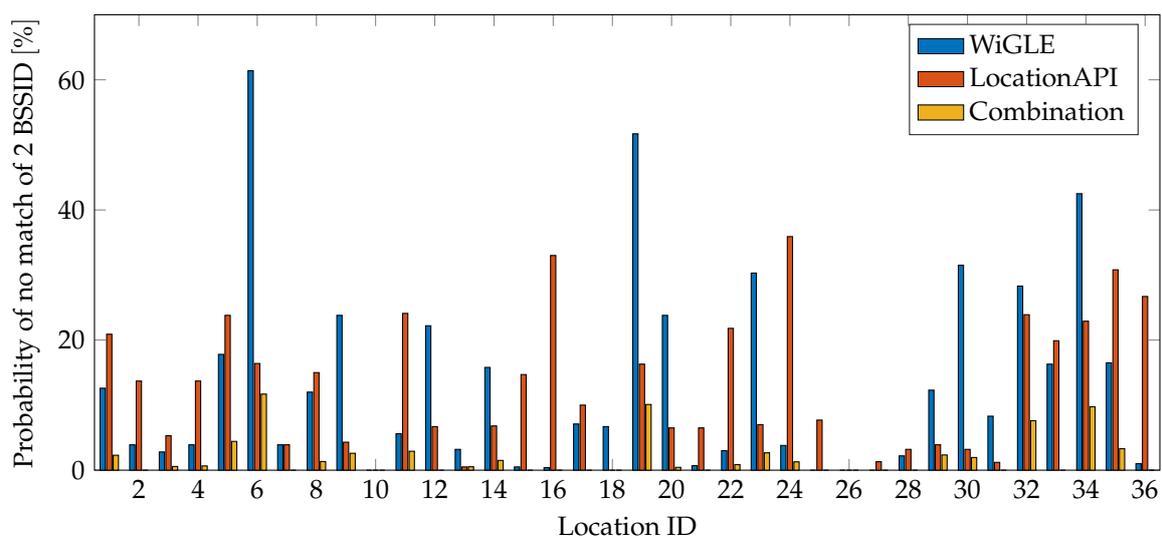


Figure 6. Probability of finding no result for either of two BSSIDs in the WiGLE and LocationAPI database. The mean chance of finding no result over all locations is 13.2% for WiGLE, 12.5% for LocationAPI, and 1.9% for the combination.

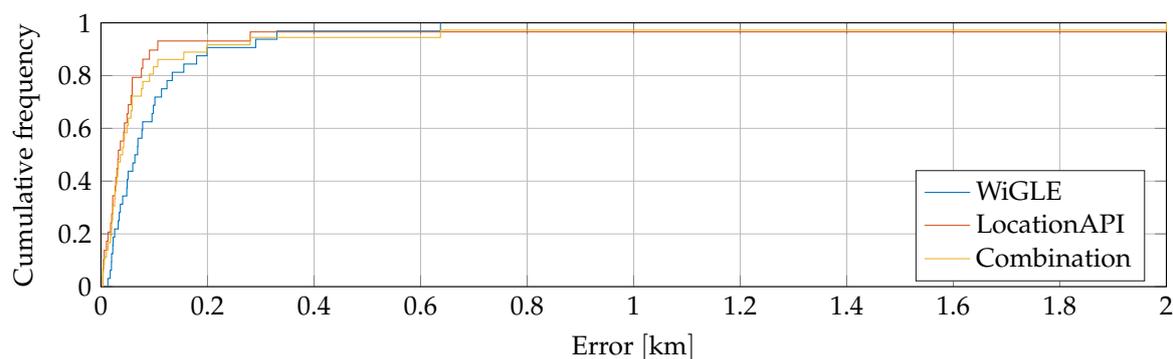


Figure 7. Cumulative distribution of the location error when the two BSSIDs are transmitted that had the highest RSSI. In the WiGLE database, the mean location error is 103 m, the median location error is 66 m, and no location estimate can be found at 4 test locations. In the LocationAPI database, the mean location error is 114 m, the median location error is 33 m, and no location estimate can be found at 7 test locations. In the combination of the databases, the mean location error is 125 m, the median location error is 39 m, and an estimate can be found at all 36 test locations. One outlier of LocationAPI and the combination is at 2.01 km.

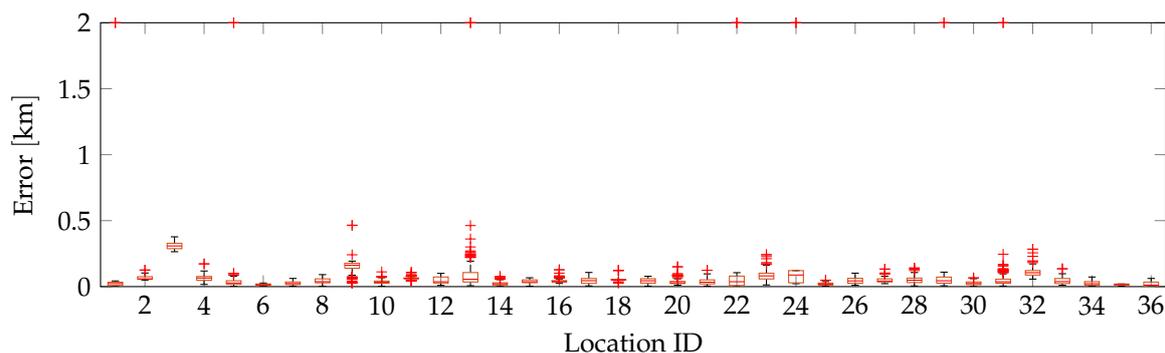


Figure 8. Overview of the location error on all 36 locations using the LocationAPI database. Errors larger than 2 km are clipped.

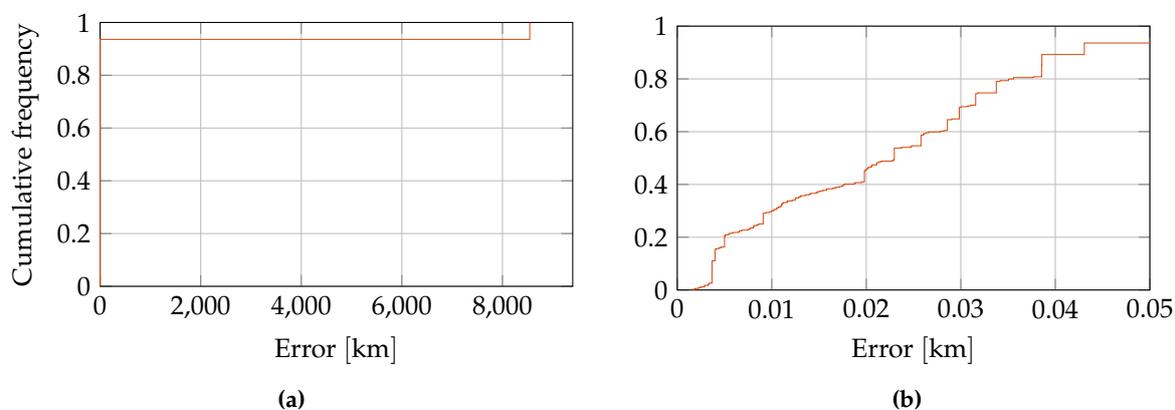


Figure 9. Cumulative distribution of the location error at the first measurement location, using the LocationAPI database. A detail of Figure (a) is shown in Figure (b). The mean location error is 546 km, the median location error is 23 m.

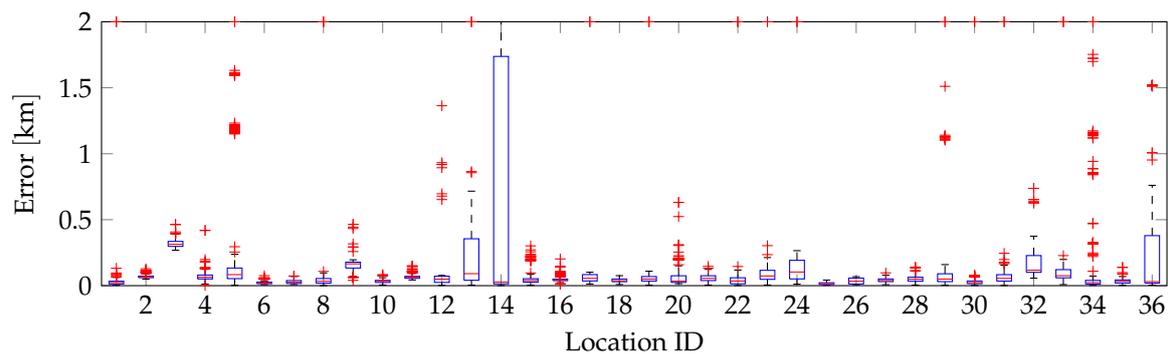


Figure 10. Overview of the location error on all 36 locations using the combination of the two databases by averaging. Errors larger than 2 km are clipped.

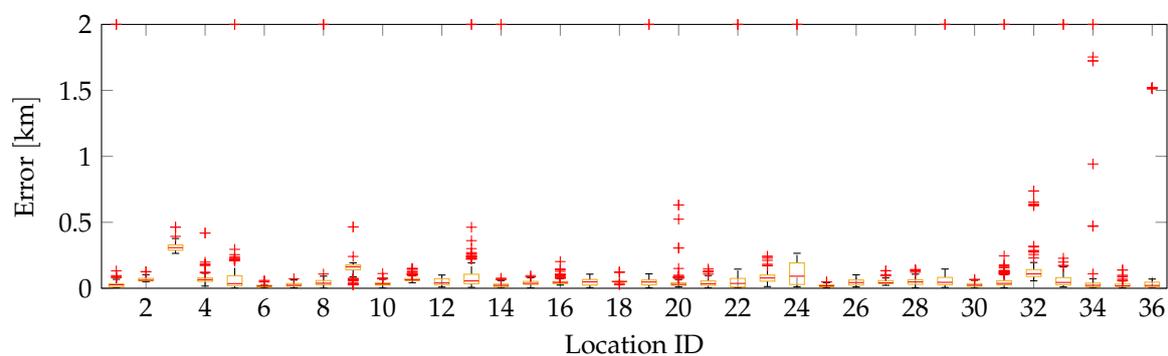


Figure 11. Overview of the location error on all 36 locations using the combination of the two databases by first choosing LocationAPI and then WiGLE. Errors larger than 2 km are clipped.

229 while the overall median location error is 53 m. The 75th percentile location error is only once larger
 230 than 1 km. Fifteen locations have an error greater than 2 km, which is clipped in Figure 10.

231 The location error overview of the second approach, which favors the LocationAPI database
 232 since it generally has a lower location error, is shown in Figure 11. The overall mean location error is
 233 24.17 km, while the overall median location error is 40 m. The 75th percentile location error is never
 234 larger than 0.5 km. Twelve locations have an error greater than 2 km, which is clipped in Figure 11.

235 The average probability of finding neither BSSIDs in the LocationAPI database, nor in the WiGLE
 236 database, of a combination of WAP seen at one of the 36 locations is 1.9%. This is calculated as in
 237 Equation (2). A graph of the probability of finding neither BSSIDs in the combination of databases is
 238 shown in Figure 6.

239 When only sending the combination of BSSIDs with the highest RSSI values to the combination of
 240 databases, the mean error of all measurement locations decreases from 24.17 km to only 125 m. The
 241 median error decreases slightly from 40 m to 39 m. Using this method, a location could be estimated at
 242 all 36 test locations. Figure 7 shows the distribution of the error.

243 4. Discussion

244 The WiGLE and LocationAPI Wi-Fi fingerprint databases both contain a lot of entries, worldwide
 245 as well as in Belgium. However, the results of the databases differ from each other at some locations.

246 4.1. Comparison of databases

247 When considering the median errors, the LocationAPI results are almost always better than the
 248 ones of WiGLE. At only four of the 36 measurement locations, the requests to WiGLE yield a smaller

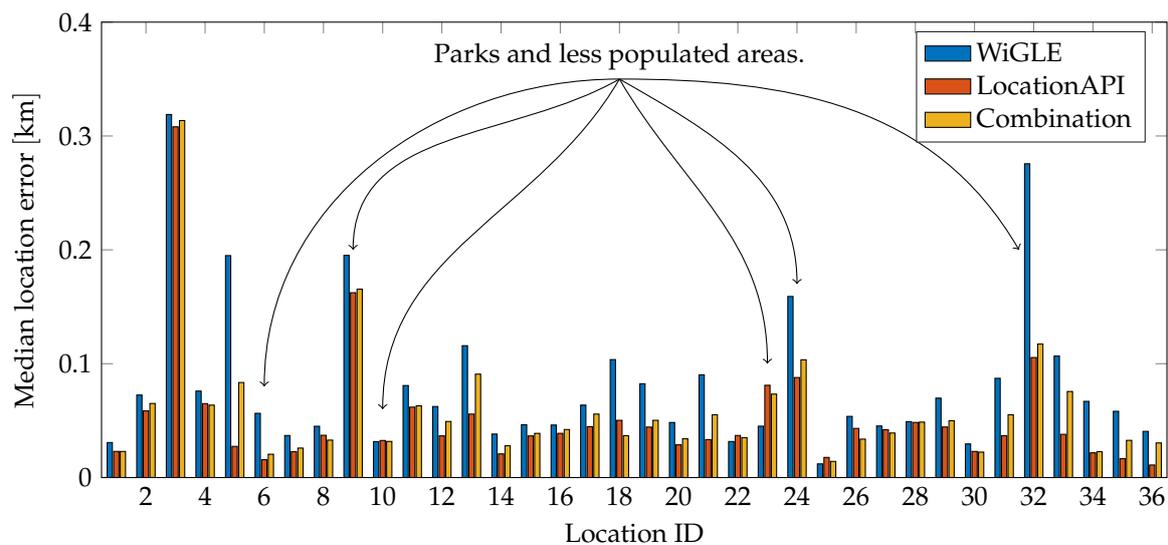


Figure 12. Median location error at each test location for both the WiGLE and LocationAPI databases, and for their combination. The location error is not consequently larger in less populated areas. See Figure S2 for an interactive map; its color legend is the same as in this figure.

249 median location error, see Figure 12. The mean distance errors are in the range of 15 m to 546 km. This
 250 variety of mean errors is, as mentioned in the results, due to outliers in the location estimates.

251 The probability that a BSSID combination is sent that contains two BSSIDs that result in no match
 252 is for both databases around 13%. However, there are large differences in this probability on specific
 253 locations, see Figure 6. This how the combination of the two databases reduces the probability to 1.9%.

254 4.2. Reasons for flaws in the results

255 When looking at the size of the location errors, we observe that a large error often occurs because
 256 a single BSSID leads to a wrong latitude and longitude pair. This causes a large error in every
 257 combination which contains that BSSID. When not considering that single 'failed BSSID', the errors
 258 will remain quite small and the distance error would decrease. Therefore, we also always calculate the
 259 median location error, which is more resistant to such outliers.

260 There is sure to be a fault during the resolving of some BSSIDs to coordinates. Therefore, when
 261 localizing a device using Wi-Fi and the WiGLE or LocationAPI database, one should keep in mind that
 262 there could be a 'corrupt' BSSID in the databases. When requesting a BSSID, the servers might respond
 263 with a wrong coordinate of the wireless access point. The distance between the GPS coordinate and
 264 the mean coordinate can vary from a few meters to over several thousand kilometers. Of course, this
 265 influenced the mean error of every measurement where such a flaw occurred.

266 For example, at location 1, when using the LocationAPI database, there is a large deviation at the
 267 errors of one or more BSSIDs. Only 5% of all combinations result in an error that is equal to 8543 km.
 268 The coordinates that are returned are located in Brazil... The remaining 95% of the errors are always
 269 below 44 m. Therefore, the median error is only 23 m, while the mean error is 546 km.

270 The reliability of the databases can definitely be questioned. For example, anyone can
 271 anonymously upload data to the databases. Hopefully, there is a kind of check before the data
 272 is actually made available for other users. Otherwise users can put wrong data into the database,
 273 for any number of reasons. The managers of the WiGLE and LocationAPI database admit that there
 274 could be geographical errors in their databases. However, an administrator of the LocationAPI website
 275 stated the following [29]:

276 “A lot of effort has gone into enabling our algorithms to auto-adapt based on the type
277 of geography requested, correct for unclean data-sets (in case of crowd-sourced data) by
278 intense cleansing of both contributor and customer data.”

279 Since it is not possible for companies like WiGLE of UnwiredLabs to verify each coordinate of
280 every uploaded wireless access point, a localization application should have some built-in control
281 mechanisms. Imagine, for example, a user uploading 100 coordinates of access points in a specific,
282 relatively small region. If 99 of the coordinates are only a few meters away from each other, and there
283 is one coordinate a few kilometers further, this could indicate that this single coordinate is the result of
284 the bad resolving in the database. The information of this access point should be left out of the data, to
285 avoid sending a combination with the coordinate of that ‘corrupt’ BSSID.

286 Another problem which must be tackled is that, even though MAC addresses are unique
287 worldwide, there will always be MAC spoofers. MAC spoofers manipulate their MAC addresses for
288 any number of reasons. Therefore, there should be a Duplicate Address Detection (DAD) mechanism
289 in the localization application. Using DAD, it becomes possible to eliminate the entry with the ‘false
290 location’ from the database if the ‘real’ location is already registered multiple times in the database.

291 Furthermore, it needs to be mentioned that there could be small distance errors, due to the
292 limitations of the hardware. The GPS and the Wi-Fi adapter of the laptop have a certain refresh speed,
293 so the user location or list of wireless access points is not instantly updated when moving.

294 4.3. Population density

295 Usually, a lot of Wi-Fi networks can be discovered in the heart of a city like Antwerp. In the
296 countryside or in open fields or parks, it is harder to detect a Wi-Fi network, because there are just
297 less WAP available in such areas. Therefore, it is interesting to investigate whether or not there is a
298 relationship between the kind of area and the distance error. Of the 36 locations, there are six locations
299 which can be categorized as less densely populated areas: locations 6, 9, 10, 23, 14, and 32. These
300 locations are indicated in Figure 12.

301 There could be a relationship between the size of the median error and the kind of area where
302 the measurement took place. Three out of the five biggest median errors were measured at a less
303 densely populated location. However, the largest error was measured at a densely populated location
304 and there are also errors in less densely populated areas that are smaller than the mean error. So, we
305 cannot be sure if there is a relationship between the distance error and the kind of location. Despite the
306 limitation of to the use of only two BSSIDs, the amount of access points in the vicinity does play a role.
307 The chance of having no match in the database increases in less populated areas from 13 % to 20 %.

308 5. Conclusions

309 We wanted to know how accurately we can localize a mobile device that communicates over Sigfox
310 by observing nearby Wi-Fi access points. We used two online databases, WiGLE and LocationAPI, of
311 Wi-Fi access points locations to search for the access points found by the mobile device. Because of the
312 limitations in Sigfox communication, only two BSSIDs can be searched for in a single transmission,
313 with a maximum of 140 transmissions per day.

314 Using the WiGLE database, we achieved a median location error of 45 m, with a 86.8 % chance of
315 finding the combination of BSSIDs transmitted by the device. Using the LocationAPI database, we
316 achieved a median location error of 23 m, with a 87.5 % chance of finding the combination transmitted
317 by the device. Both databases have outliers well above 100 km. However, only in the WiGLE database
318 the 75th percentile location error is larger than 1 km, and this only at two of the test locations.

319 To increase the chance of finding the combination of BSSIDs transmitted by the device, we
320 combined the results of both databases. Since the LocationAPI database is more accurate than the
321 WiGLE database, we chose to combine them by only using the WiGLE database when the combination
322 of BSSIDs was not found in the LocationAPI database. With this method, we achieved a median
323 location error of 40 m, with a 98.1 % chance of finding the combination transmitted by the device.

324 Since only 140 transmissions are allowed per day, and only two BSSIDs can be sent in a single
325 transmission, we additionally explored a strategy for selecting the combination of BSSIDs. The strategy
326 consists of selecting the two BSSIDs with the highest received signal strength. Using our combination
327 of both databases, the median location error is 39 m, and this combination could be found on all 36 test
328 locations.

329 After this experiment with real data, we want to increase the reliability by building a device that
330 can both sense the Wi-Fi access points and transmit two BSSIDs over Sigfox. This will not only enable
331 us to confirm our results in a closed experiment, but also allow us to measure how much energy is
332 exactly required for this type of localization, so that we know if it is a viable way to track and trace
333 assets in an Internet of Things scenario.

334 **Supplementary Materials:** The following are available online at www.mdpi.com/link—Figure S1: Dynamic map
335 of test locations, Figure S2: Dynamic map of median error at the test locations. Results for WiGLE, LocationAPI,
336 and the combination are colored as in the other figures: WiGLE is blue, LocationAPI is red, and the combination is
337 yellow.

338 **Acknowledgments:** Part of this research was funded by the Flemish FWO SBO S004017N IDEAL-IoT (Intelligent
339 DEense And Longe range IoT networks) project, and by the ICON project MuSCLe-IoT. MuSCLe-IoT is a project
340 realized in collaboration with imec, with project support from VLAIO (Flanders Innovation and Entrepreneurship).
341 Project partners are imec, Flash Private Mobile Networks, Engie M2M, Sensolus and Aertssen.

342 **Author Contributions:** R.B. and M.W. conceived and designed the experiments; T.J. performed the experiments;
343 T.J. and R.B. analyzed the data; T.J. and R.B. wrote the paper.

344 **Conflicts of Interest:** The authors declare no conflict of interest. The funding sponsors had no role in the design
345 of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the
346 decision to publish the results.

347 References

- 348 1. Shamsuzzoha, A.; Ehres, M.; Tenkorang, R.A.; Nguyen, D.; Helo, P.T. Performance evaluation of tracking
349 and tracing for logistics operations. *International Journal of Shipping and Transport Logistics* **2013**, *5*, 31–54.
- 350 2. Asioli, D.; Boecker, A.; Canavari, M. On the linkages between traceability levels and expected and actual
351 traceability costs and benefits in the Italian fishery supply chain. *Food Control* **2014**, *46*, 10–17.
- 352 3. Stranieri, S.; Cavaliere, A.; Banterle, A. Voluntary traceability standards and the role of economic incentives.
353 *British Food Journal* **2016**, *118*, 1025–1040.
- 354 4. Hillbrand, C.; Schoech, R. Shipment Localization Kit: An Automated Approach for Tracking and Tracing
355 General Cargo. International Conference on the Management of Mobile Business (ICMB 2007). IEEE, 2007,
356 pp. 1–7.
- 357 5. Kandel, C.; Klumpp, M.; Keusgen, T. GPS based track and trace for transparent and sustainable global
358 supply chains. Concurrent Enterprising (ICE), 2011 17th International Conference on; IEEE: Aachen,
359 Germany, 2011; pp. 1–8.
- 360 6. Oliveira, R.R.; Cardoso, I.M.; Barbosa, J.L.; da Costa, C.A.; Prado, M.P. An intelligent model for logistics
361 management based on geofencing algorithms and RFID technology. *Expert Systems with Applications* **2015**,
362 *42*, 6082–6097.
- 363 7. Raza, U.; Kulkarni, P.; Sooriyabandara, M. Low Power Wide Area Networks: An Overview. *IEEE*
364 *Communications Surveys & Tutorials* **2017**, *PP*, 1–19.
- 365 8. Sigfox S.A.. Sigfox, 2017. Available online: <https://www.sigfox.com/en> (accessed on 10 June 2017).
- 366 9. Centenaro, M.; Vangelista, L.; Zanella, A.; Zorzi, M. Long-range communications in unlicensed bands: the
367 rising stars in the IoT and smart city scenarios. *IEEE Wireless Communications* **2016**, *23*, 60–67.
- 368 10. Berkvens, R. Uncertainty of localization using electromagnetic fingerprints. PhD thesis, Universiteit
369 Antwerpen, 2017.
- 370 11. Bolić, M.; Rostamian, M.; Djurić, P.M. Proximity Detection with RFID: A Step Toward the Internet of
371 Things. *IEEE Pervasive Computing* **2015**, *14*, 70–76.
- 372 12. Neiryck, D. An IEEE 802.15.4A Ultra-Wideband Transceiver for Real Time Localisation and Wireless
373 Sensor Networks. In *Frequency References, Power Management for SoC, and Smart Wireless Interfaces*;

- 374 Baschirotto, A.; Makinwa, A.A.K.; Harpe, P., Eds.; Springer International Publishing: Cham, Switzerland,
375 2014; chapter 17, pp. 297–310.
- 376 13. Fang, S.H.; Hsu, Y.T.; Shiao, Y.; Sung, F.Y. An Enhanced Device Localization Approach Using Mutual
377 Signal Strength in Cellular Networks. *IEEE Internet of Things Journal* **2015**, *2*, 596–603.
- 378 14. Bensky, A. *Wireless Positioning Technologies and Applications, Second Edition*, 2 ed.; Artech House, 2016; p.
379 424.
- 380 15. He, S.; Chan, S.H.G. Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons.
381 *IEEE Communications Surveys & Tutorials* **2016**, *18*, 466–490.
- 382 16. Torres-Sospedra, J.; Moreira, A.; Knauth, S.; Berkvens, R.; Montoliu, R.; Belmonte, O.; Trilles, S.; João
383 Nicolau, M.; Meneses, F.; Costa, A.; Koukofikis, A.; Weyn, M.; Peremans, H. A realistic evaluation of indoor
384 positioning systems based on Wi-Fi fingerprinting: The 2015 EvAALETRI competition. *Journal of Ambient
385 Intelligence and Smart Environments* **2017**, *9*, 263–279.
- 386 17. Zhang, L.; Valaee, S.; Xu, Y.B.; Ma, L.; Vedadi, F. Graph-Based Semi-Supervised Learning for Indoor
387 Localization Using Crowdsourced Data. *Applied Sciences* **2017**, *7*, 467.
- 388 18. Gezici, S.; Zhi Tian.; Giannakis, G.; Kobayashi, H.; Molisch, A.; Poor, H.; Sahinoglu, Z. Localization via
389 ultra-wideband radios: a look at positioning aspects for future sensor networks. *IEEE Signal Processing
390 Magazine* **2005**, *22*, 70–84.
- 391 19. Sallouha, H.; Chiumento, A.; Pollin, S. Localization in Long-range Ultra Narrow Band IoT Networks
392 using RSSI. 2017 IEEE International Conference on Communications (ICC); IEEE: Paris, France, 2017; p. 6,
393 [1703.02398].
- 394 20. Sapiezynski, P.; Gatej, R.; Mislove, A.; Lehmann, S. Opportunities and Challenges in Crowdsourced
395 Wardriving. Proceedings of the 2015 ACM Conference on Internet Measurement Conference - IMC '15;
396 ACM Press: Tokyo, Japan, 2015; pp. 267–273.
- 397 21. Robusto, C.C. The Cosine-Haversine Formula. *The American Mathematical Monthly* **1957**, *64*, 38.
- 398 22. Combain. Combain Positioning Solutions, 2017. Available online: <https://combain.com/> (access on 3 July
399 2017).
- 400 23. UnwiredLabs. LocationAPI: Reliable mobile mapping, 2017. Available online: <http://locationapi.org/>
401 (access on 3 July 2017).
- 402 24. Mozilla. Mozilla Location Service, 2017. Available online: <https://location.services.mozilla.com/> (access
403 on 3 July 2017).
- 404 25. Navizon. Navizon: WiFi and Cell-ID location database with Global coverage, 2017. Available online:
405 <https://www.navizon.com/> (access on 3 July 2017).
- 406 26. WiGLE. WiGLE: Wireless Network Mapping, 2017. Available online: <https://wigo.net/> (access on 30
407 June 2017).
- 408 27. Navizon. Navizon technology: How it works, 2017. Available online: [https://www.navizon.com/
409 navizon-how-it-works](https://www.navizon.com/navizon-how-it-works) (access on 30 June 2017).
- 410 28. Business Wire. Make Your Smart Phone Pay For Itself by Running Navizon, 2017. Available online: [http://
411 www.businesswire.com/news/home/20100128005942/en/Smart-Phone-Pay-Running-Navizon](http://www.businesswire.com/news/home/20100128005942/en/Smart-Phone-Pay-Running-Navizon) (access
412 on 3 July 2017).
- 413 29. UnwiredLabs. The Unwired Labs Blog: Going from “Accurate” to “Really accurate”, 2015. Available
414 online: <http://locationapi.org/blog/> (access on 9 May 2017).