2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 24-27 September 2018, Nantes, France

Outdoor Fingerprinting Localization using Sigfox

Thomas Janssen, Michiel Aernouts, Rafael Berkvens and Maarten Weyn
IDLab - Faculty of Applied Engineering
University of Antwerp - imec
Groenenborgerlaan 171, 2020 Antwerp, Belgium;
Email: thomas.janssen@student.uantwerpen.be, {michiel.aernouts, rafael.berkvens, maarten.weyn}@uantwerpen.be

***Abstract*—The Internet of Things (IoT) has caused the modern society to connect everything in our environment to a network. In a myriad of IoT applications, smart devices need to be located. This can easily be done by satellite based receivers. However, there are more energy-efficient localization technologies, especially in Low Power Wide Area Networks (LPWAN). In this research, we discuss the accuracy of an outdoor fingerprinting technique using a large outdoor Sigfox dataset which is openly available. A *k*NN (*k* Nearest Neighbors) algorithm is applied to our fingerprinting database. 31 different distance functions and four RSS data representations are evaluated. Our analysis shows that a Sigfox transmitter can be located with a mean estimation error of 340 meters.**

***Keywords*—IoT, LPWAN, Sigfox, Localization, Fingerprinting**

I. Introduction

In recent years, the Internet of Things (IoT) has been connecting numerous everyday objects and devices to large communication networks. In order to establish communication between these devices and IoT users, Low Power Wide Area Network (LPWAN) standards such as Sigfox [1], LoRaWAN [2] and NB-IoT [3] have been developed. Generally, LPWAN standards enable IoT devices to exchange small messages over long distances while maintaining a very low power consumption, which allows these devices to work for multiple years on small batteries. This meets the need of an endless amount of applications such as home automation, health care solutions and smart cities.

Many IoT applications benefit from context awareness. For example, a mobile environmental monitoring device has to send an alert if a certain air quality threshold is exceeded, but the threshold of a dense urban area will be different from the threshold in rural areas. Concisely, the application has to adapt its behavior based on context information such as the current location of the IoT device. Hence, localization methods have to be applied to obtain this context information.

The most common examples of localization technologies are Global Navigation Satellite Systems (GNSS) such as Global Positioning System (GPS). Although GNSS receivers achieve highly accurate location estimations, they generally consume a significant amount of power. For instance, a GPS receiver on a mobile phone consumes 320 mW on average [4]. Therefore, GNSS localization is not suitable for long-term battery-powered IoT applications. On the other hand, a device can be located via its communication link by applying wireless positioning techniques such as Angle of Arrival (AoA), Time Difference of Arrival (TDoA), or methods based on Received Signal Strength (RSS) such as ranging and fingerprinting. Consequently, no additional messages or power consumption are required to obtain a location estimate.

In this paper, we present our research on outdoor Sigfox fingerprinting. Similar to previous research on Wi-Fi fingerprinting [5], we use an openly available Sigfox fingerprinting dataset to analyze the effect of distance functions and RSS representations on the location estimation error. The remainder of this paper is structured in the following way. Section II illustrates the related work of this paper. Section III explains the dataset that was used, and describes how the dataset was used to analyze outdoor Sigfox fingerprinting algorithms. In Section IV, we show and discuss the results of this analysis. Finally, Section V concludes the paper, summarizing our main observations.

II. Related work

To date, there already exist a number of studies on localization using fingerprinting techniques. While most of this research is carried out in indoor environments, we try to extend the range of applications to outdoor environments. Therefore, a Sigfox network is used. In the following paragraphs, we will briefly explain the main concepts of Sigfox and fingerprinting techniques.

A. Sigfox

Sigfox is a LPWAN standard that uses Ultra-Narrow Bandwidth (UNB) modulation to send small messages via the license-free sub-GHz ISM band (868 MHz in Europe, 915 MHz in the US). Depending on environmental characteristics, base stations can receive these messages over distances of 10 km to 50 km. In Europe, duty cycle regulations of the 868 MHz ISM band limits Sigfox devices to 36 transmission seconds per hour, and 6 seconds time on air per package. As the UNB modulation imposes a limited data rate of 100 bps, a Sigfox device can send 6 messages per hour, with a maximum payload of 12 bytes per message [6]. Despite these limitations, Sigfox is a suitable LPWAN standard for many IoT applications that are not time-critical, e.g. monitoring water meters, air quality sensing, etc. In our previous research, we explored a localization approach which combines open-source WiFi BSSID databases with Sigfox. Every 10 minutes, a mobile node obtained its current nearest WiFi BSSIDs and simulated a 12 byte Sigfox message which contains the 2 BSSIDs with the highest RSSI. By comparing these BSSIDs to open-source WiFi BSSID databases, the location of the mobile node could be obtained with estimation errors of 23 to 45 meter [7]. Of course, this approach requires Sigfox messages to be dedicated for localization purposes. In this paper, we use the physical characteristics of the Sigfox communication link to estimate the location of a transmitting Sigfox device.

In Belgium, a nation-wide Sigfox network is deployed by Engie M2M [8]. In the first part of our research, we used this network to collect a large database of Sigfox messages in Antwerp, Belgium. The collection methodology of this database will be explained further in Section III. In a previous research, we discussed the accuracy of Wi-Fi fingerprinting using this network as well.

B. Fingerprinting

Fingerprinting is a localization method that uses signal characteristics such as a Received Signal Strength Indicator (RSSI) to estimate the location of a transmitting device. In a first offline phase, a fingerprinting database is built, which contains RSSI measurements at known training locations in a predetermined area. During the online phase, a mobile transmitter can be located by comparing the real-time RSSI measurements to the fingerprints in the training database, e.g. by using probabilistic methods, machine learning or by applying a pattern matching technique such as k Nearest Neighbors (kNN) analysis. Up to now, fingerprinting is mostly used in constricted indoor environments with wireless technologies such as Wi-Fi, Bluetooth Low Energy (BLE), etc [9], [10]. Sallouha et al. evaluated a fingerprinting method with Sigfox in an outdoor environment [11]. In their implementation, RSSI measurements are used to estimate the distance between a transmitting Sigfox device and the responding base stations to classify in which area the transmitter is located. Within such a class, the location estimation is improved by distance estimation between end-devices and ground truth GPS devices. However, the experiments were conducted in a small area which contains only two classes. In our research, we use a Sigfox dataset with a spatial spread that covers the entire city center of Antwerp, Belgium. With this dataset we investigated the fingerprinting algorithm using different distance functions and RSSI data representations.

III. Methods and Materials

Several steps needed to be taken in order to execute our experiments successfully. In this section, we describe the procedure of every experiment and explain why we performed each particular step.

A. Sigfox dataset

In order to implement fingerprinting in a large outdoor area, an extensive measurement campaign has to be conducted to create a large training database that covers the entire area. Previous work by Aernouts et al. provided the research community with large LPWAN fingerprinting datasets in outdoor areas [12]. One of these datasets holds 14,378 Sigfox messages which were collected in the city center of Antwerp, using the Sigfox network which is deployed by Engie M2M. These messages were transmitted by devices that were mounted on 20 cars of the Belgian postal services. The cars commute through the entire city center on a daily basis, which benefits the spatial spread of the dataset. Every twelve minutes, the current GPS coordinates of such a device are sent via a Sigfox message. Together with the RSSI measurements of every Sigfox base station that received the message, the GPS coordinates of the message were stored in the Sigfox dataset. Consequently, the dataset consists of 14,378 rows where each row represents a unique training sample. The first 84 columns represent all Sigfox base stations which are present in the dataset. If a base station received a message, the RSSI for that base station is filled in in its respective column. Otherwise, an out-of-range value of -200 dBm is inserted. The last columns in the dataset show the receiving time and GPS coordinates of a message.

For our experiments, we divided the Sigfox dataset in a training set, evaluation set and test set. The size of these subsets are respectively 70%, 15% and 15% of the complete dataset.

B. Distance functions and RSS representations

The distance between two points in space needs to be calculated in order to perform a fingerprinting algorithm. So far, investigations have been confined primarily to the Euclidean distance and Received Signal Strength (RSS) data expressed in dBm. However, various options can be considered to improve the results of the distance calculations. Cha et al. described several distance functions and similarities [13]. Torres-Sospedra et al. implemented all these distance functions for an indoor fingerprinting positioning system using Wi-Fi [5]. Based on this research, we investigated the accuracy of Sigfox in outdoor environments. In total, 31 different distance functions are implemented. Some distance functions are excluded, since some of them are equivalent or too advanced for our experiments. Besides, in distance-based methods (such as kNN), some distance functions are equivalent and some similarities occur since two distances can only differ by a single constant. In general, distance functions are categorized into families, based on their similarities. The 31 implemented distances are listed below by family.

- *Minkowski family*: Euclidean, Manhattan, Minkowski 3, Minkowski 4, Minkowski 5, Chebyshev
- *L1 family*: Gower, Sørensen, Soergel, Kulczynski, Canberra, Lorentzian
- *Intersection family*: Intersection, Wave Hedges, Czekanowski, Motyka
- *Inner product family*: Jaccard, Dice
- *Fidelity family*: Hellinger, Matusita, Squared Chord
- *Squared L2 family*: Squared Euclidean, Pearson χ^2, Neyman χ^2, Squared χ^2, Probabilistic χ^2, Divergence, Clark, Additive Symmetric χ^2
- *Combinations*: Kumar-Johnson, Average(L_1, L_∞)

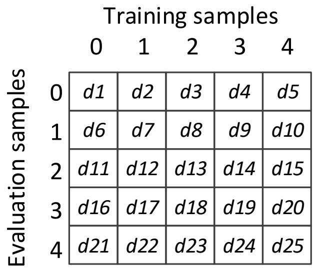

Fig. 1. Representation of a distance matrix

Additionally, Torres-Sospedra et al. list four alternatives to represent the signal strength values: *positive, normalized, exponential* and *powed* values. For the mathematical background of these data representations, the interested reader is referred to the paper of Torres-Sospedra [5]. Note that we use the same values for the parameters required in the exponential and powed representations, so α equals 24 and β equals e. As mentioned before, an out-of-range value of -200 dBm is put into the fingerprinting database for base stations that did not receive a Sigfox message.

C. Distance matrices

In order to perform a fingerprinting algorithm, distance matrices need to be generated in the next step. Each row in a distance matrix represents an evaluation sample, while each column represents a training sample. Thus, for every evaluation sample, the distance to every training sample is stored in the corresponding cell in the distance matrix. This is visualized in Figure 1. For every distance function and for every RSS representation, a distance matrix is generated.

D. Fingerprinting algorithm

Given the dataset and the distance matrices, we can start processing all this data and apply a fingerprinting algorithm using Python scripts. Many algorithms are available, e.g. k Nearest Neighbors (kNN), Support Vector Machines (SVM) [14] and Neural Networks (NN). For our experiments, we apply the kNN classification algorithm. In this algorithm, the coordinates of the k nearest base stations are used to estimate the position of the device of interest. In practice, a sample of signal strengths is compared to the training samples in the fingerprinting database. Using the kNN algorithm, we can then obtain a location estimate for the transmitting device. In pursuance of achieving the highest possible accuracy, we find the optimal k for every distance function and RSS representation, with k ranging from 1 to 16. The optimal k is defined as k where the smallest distance is measured.

E. Error calculations

After applying the fingerprinting algorithm, the error between the estimated position and the actual position of the device is calculated. This is done by calculating the Vincenty distance, which takes into account the curvature of the Earth.

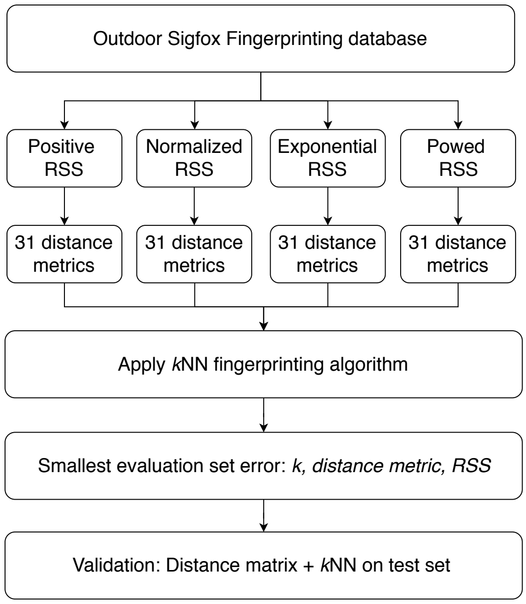

Fig. 2. Overview of localization experiments

For every distance function and RSS representation, the mean distance error is calculated for every k in the evaluation set and for the optimal k in the test set, which is discussed in the next section.

F. Validation

Finally, in order to validate the parameters of the fingerprinting algorithm and the results, a test dataset is used. From this set, a distance matrix is generated for the distance function and RSS representation with the smallest distance error and optimal k. In that way, we can compare the distance error of the test dataset to the distance error of the evaluation dataset, for the optimal k and distance function. It these errors are similar, we can assume our results to be valid.

In Figure 2, an overview of our experiments is shown.

IV. Results

In this section, we will discuss the results of the experiments and compare the different distance functions and RSS representations with each other. In order to validate our results, we will use a test set for the case with the smallest error, optimal k and RSS representation.

In Table I, the optimal k and mean error are listed for every distance function and every RSS representation. In almost all cases, the lineal representations (positive and normalized) yield similar results. Furthermore, we can observe a decrease of the mean error for the exponential and powed representations. In our experiment, using one of these two representations always results in a smaller distance error as when using a lineal representation, except for the Kumar-Johnson distance. When comparing the exponential and powed representations, the exponential RSS representation always yields the smallest distance error. The Sørensen, Soergel, Kulczynski, Czekanowski and Motyka distance functions all

TABLE I
FINGERPRINTING RESULTS FOR THE SIGFOX URBAN DATASET, SHOWING OPTIMAL k AND SMALLEST ERROR IN METERS.

Measure	Positive RSS		Normalized RSS		Exponential RSS		Powed RSS	
	k	Error	k	Error	k	Error	k	Error
Euclidean	10	667	10	667	6	329	9	361
Manhattan	7	537	7	537	6	333	6	356
Minkowski 3	9	729	9	729	6	341	7	373
Minkowski 4	13	752	13	752	6	357	6	394
Minkowski 5	14	762	14	762	6	367	7	409
Chebyshev	7	906	7	906	5	413	4	494
Gower	7	537	7	537	6	333	6	356
Sørensen	9	539	9	539	6	322	7	346
Soergel	9	539	9	539	6	322	7	346
Kulczynski	9	539	9	539	6	322	7	346
Canberra	9	687	9	687	8	488	7	543
Lorentzian	6	460	6	524	6	333	6	357
Intersection	7	537	7	537	6	333	6	356
Wavehedges	7	613	7	613	7	453	7	481
Czekanowski	9	539	9	539	6	322	7	346
Motyka	9	539	9	539	6	322	7	346
Jaccard	11	677	11	677	7	325	6	352
Dice	11	677	11	677	7	325	6	352
Hellinger	9	775	9	775	8	417	9	567
Matusita	9	775	9	775	8	417	9	567
Squared Chord	9	775	9	775	8	417	9	567
Squared Euclidean	6	789	10	667	6	329	9	361
Pearson χ^2	3	1122	3	1113	3	575	2	847
Neyman χ^2	4	1056	6	1048	3	739	4	830
Squared χ^2	12	747	12	747	6	391	8	477
Probabilistic Symmetric χ^2	10	758	12	747	6	391	8	477
Divergence	10	707	10	707	9	588	10	682
Clark	9	687	9	687	8	487	7	543
Additive Symmetric	6	955	6	948	7	555	6	939
Kumar-Johnson	6	950	6	940	7	636	6	947
Average(L_1, L_∞)	8	853	6	730	6	390	6	443

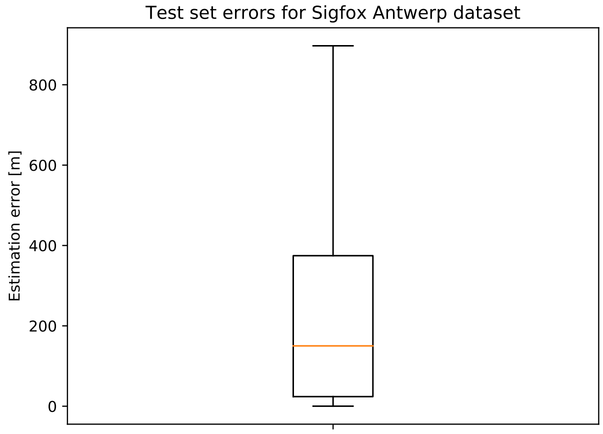

Fig. 3. Box plot of estimation errors of the Sigfox urban test dataset.

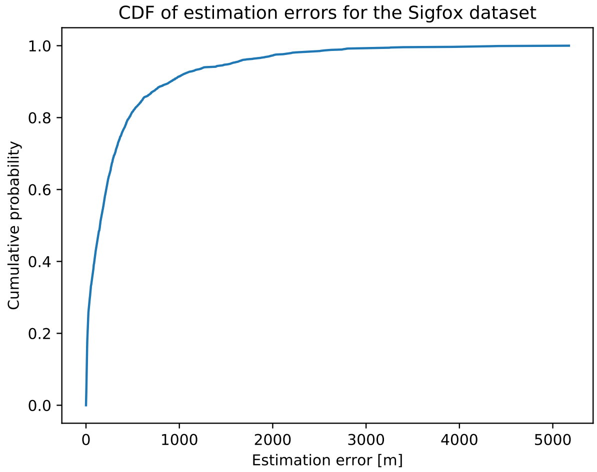

Fig. 4. Cumulative distribution function of estimation errors of the Sigfox urban test dataset.

yield the exact same errors. Using the exponential RSS representation and k equal to 6, the smallest distance error can be achieved. In these cases, the mean estimation error is 322 meters.

Since Sørensen distance requires the least computational power, we chose this distance function for our test set calculations. Thus, the test set distance matrix is created using exponential RSS representation, the Sørensen distance and k equal to 6. When performing the same fingerprinting algorithm as with the evaluation set, the mean test set error is 340 meters, which is close to the mean error of 322 meters that we obtained

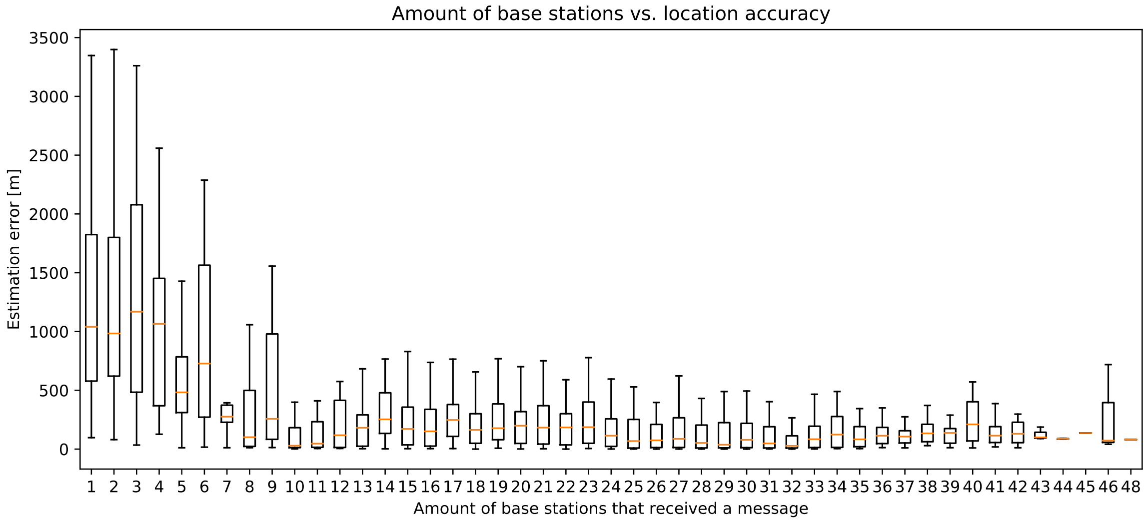

Fig. 5. Box plots of estimation errors of the Sigfox urban test dataset.

with the evaluation set. Therefore, we can assume our results to be valid. In Figure 3, a box plot of the test set errors is shown and in Figure 4, a cumulative distribution function (CDF) of the errors is shown. The median error is 150 meters and the 95th percentile yields an error of 1548 meters.

In Figure 5, the correlation between the amount of base stations that received a test message and the estimation errors for that message is shown. In general, we can observe that if more base stations receive a message, the estimation error decreases. Consequently, Sigfox network operators should take the amount of base stations into account to improve the location estimation accuracy.

V. Conclusion

Due to the proliferation of IoT devices, new energy-efficient technologies are needed to locate them. In this research, we performed experiments to discuss the accuracy of Sigfox localization, using a kNN fingerprinting algorithm. Therefore, We investigated 31 distance metrics and four alternatives to represent RSS data. The results show a mean localization error of 340 meters, in the optimal case where $k = 6$ and the Sørensen distance is used in combination with the exponential RSS representation.

In our future work, we will explore several options to improve the location estimation accuracy of the fingerprinting algorithm. First, we found that increasing the amount of base stations in a given area decreases the distance error. Second, we will investigate fine-tuning of the parameters α and β in the exponential and powed RSS representations, respectively. Third, the kNN algorithm can be improved by implementing a weighted variant of this algorithm. Finally, more studies are required to optimize the size of the datasets. We will repeat this analysis on other LPWAN datasets by Aernouts et al, including a LoRaWAN dataset that was also collected in the city center of Antwerp [12].

Acknowledgment

Part of this work was funded by the MuSCLe-IoT (Multimodal Sub-Gigahertz Communication and Localization for Low-power IoT applications) project, co-funded by imec, a research institute founded by the Flemish Government, with project support from VLAIO (contract number HBC.2016.0660). The MuSCLe-IoT industry partners are Flash, Sensolus, Engie M2M, and Aertssen.

References

[1] Sigfox, “Sigfox - The Global Communications Service Provider for the Internet of Things (IoT),” 2018. [Online]. Available: https://www.sigfox.com/en

[2] LoRa Alliance, “LoRaWAN: Wide Area Networks for IoT.” [Online]. Available: https://www.lora-alliance.org/

[3] Y. P. Wang, X. Lin, A. Adhikary, A. Grövlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi, “A Primer on 3GPP Narrowband Internet of Things,” *IEEE Communications Magazine*, vol. 55, no. 3, pp. 117–123, 2017.

[4] M. B. Kjasrgaard, “Location-based services on mobile phones: minimizing power consumption,” *IEEE Pervasive Computing*, vol. 11, no. 1, pp. 67–73, 2012.

[5] J. Torres-Sospedra, R. Montoliu, S. Trilles, . Belmonte, and J. Huerta, “Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems,” *Expert Systems with Applications*, vol. 42, no. 23, pp. 9263–9278, 12 2015.

[6] B. Vejlgaard, M. Lauridsen, H. Nguyen, I. Z. Kovács, P. Mogensen, and M. Sørensen, “Coverage and Capacity Analysis of Sigfox, LoRa, GPRS, and NB-IoT,” in *Vehicular Technology Conference*. IEEE, 2017.

[7] T. Janssen, M. Weyn, and R. Berkvens, “Localization in Low Power Wide Area Networks Using Wi-Fi Fingerprints,” *Applied Sciences*, vol. 7, no. 9, p. 936, 9 2017.

[8] Engie M2M, “Engie M2m: Enabling your IoT solution,” 2018. [Online]. Available: http://www.engiem2m.be/homepage

[9] S. He and S.-H. G. Chan, “Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.

[10] R. Faragher and R. Harle, “Location Fingerprinting With Bluetooth Low Energy Beacons,” *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2418–2428, 11 2015.

[11] H. Sallouha, A. Chiumento, and S. Pollin, “Localization in long-range ultra narrow band IoT networks using RSSI,” in *IEEE International Conference on Communications*. IEEE, 5 2017, pp. 1–6.

[12] M. Aernouts, R. Berkvens, K. Van Vlaenderen, and M. Weyn, "Sigfox and LoRaWAN Datasets for Fingerprint Localization in Large Urban and Rural Areas," *Data*, vol. 3, no. 2, 4 2018.

[13] S.-h. Cha, "Comprehensive Survey on Distance / Similarity Measures between Probability Density Functions," *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 1, no. 4, pp. 300–307, 2007.

[14] D. Tran and T. Nguyen, "Localization In Wireless Sensor Networks Based on Support Vector Machines," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 7, pp. 981–994, 7 2008.