

Design of a Guideline for Range-based Localization Algorithms Evaluation using Multiple Linear Regressions

Dhouha El Houssaini^{1,2}, Zaid Abdullah¹, Sabrine Kheriji^{2,3}, Kamel Besbes¹ and Olfa Kanoun¹

¹Chair for Measurement and Sensor Technology, Technische Universität, 09111 Chemnitz, Germany

²Microelectronics and Instrumentation Laboratory, Faculty of Sciences of Monastir, 1002, Tunisia

³Centre for Research on Microelectronics and Nanotechnology, Technopark of Sousse, 4000, Tunisia

Keywords: WSNs, Localization, Range-based, Multiple Linear Regression, Guideline, IoT.

Abstract: Localization is an essential feature in numerous Wireless Sensor Network (WSN) applications, including tracking, health monitoring, and military supervision. Analytical modeling and analysis of the localization system remain challenging and infeasible since it offers oversimplified results with limited reliability to the evaluated cases. Likewise, disseminating test-beds involves a lot of effort, making the simulation phase indispensable to study the WSN localization. The defined localization model needs to ensure solid and pragmatic network assumptions during the simulation. However, most network simulators don't meet specific criteria related to network definition, such as scalability and heterogeneity. As part of this endeavor, a guideline for evaluating and analyzing technical methods of range-based localization is developed. Multiple linear regression is used to generate the different localization instances, which enables to support different and non-dependent parameters. The developed guideline for range-based localization is tested and validated for existing localization solutions.

1 INTRODUCTION

In Wireless Sensor Networks (WSNs), sensor nodes are installed in the field of interest to monitor and provide certain physical and environmental information, such as the temperature, humidity, and activity of monitored quantities (Kanoun et al., 2021). The Localization of these installed nodes is critical and required for different applications, like object tracking, process supervision, and monitoring (El Houssaini et al., 2020; Naguib et al.,). Besides, following the urgent trend to include the Internet of Things (IoT) concept, the location information remains critical and necessary for remote control and supervision activities (Ahmed Mansoor and Irtaza, 2019). The knowledge of the position of installed nodes serves during the network activities such as the routing, data transmission, and network topology (Khrijji et al., 2018; Paul and Sato, 2017; Egea-Lopez et al., 2005; Abdelhabib and Brahim, 2008). Indeed, the node's location is critical to ensure better network functionality and enhance the lifetime of the node itself and the complete network. It is essential to effectively decide upon the localization technique for the network assumptions and characteristics, such as the net-

work size, environment (indoor/outdoor, the existence of obstacles, mobile system, etc.), hardware specification (sensing and communication range and power module) and energy constraints. Thus, the realization of a specific test-bed is extremely expensive and challenging. Furthermore, the implementation of real experiments always takes more time than the simulation. Therefore, simulation phase is crucial for the development of WSNs.

The use of WSNs simulators allows users to separate numerous factors to evaluate and test their approaches, like flexible network size, different sensing and communication ranges, and predefined communication modules. Various free and open-source simulators are commercialized in the market, with some advantages. NS2 is a discrete event-driven simulator, mainly used for academic research in the areas of computer networks, MANETs, and WSNs (Hogie et al., 2006). It is built based on C++ and supports a wide range of protocols. It also provides complete support of communication protocol. The main limitation of NS2 is its graphical interface, as it only provides a simple reflection of the network. It does not allow the scalability and extension of the network. Furthermore, NS3 was developed based on the

main concepts of NS2 (Nsam,). It integrates NetAnim module for the graphical simulation of a network model. Additionally, different simulators integrate GUI interfaces. One example is OMNET++ (Omnet,), which is an example of a modular discrete event simulator written in C++ and provides a powerful GUI library for animation and tracing and debugging support. However, their libraries remain limited compared to other simulators, as they don't support enough protocols for communication and transmission. For instance, J-sim (Java Simulator) (JSim,) is a compositional simulation environment. It is built according to the component-based software paradigm called autonomous component architecture (ACA). J-Sim has the advantage of supporting many protocols, including a highly detailed simulation of WSNs plus Localization features. However, the JSim development is closed, and its scale is medium.

This paper aims to develop a generalized guideline, which identifies the appropriate localization technique for the test network. It considers the definition of the deployment environment (Network size and path loss exponent) and the system assumptions, such as the sending power and the communication range. The guideline implements a multiple regression module to build the decision upon the choice of the localization technique. The comparison is based on the localization accuracy and measurement error.

The remainder of the paper is organized as follows: Section 2 illustrates the related works presenting an overview of range-based localization techniques. In section 3, the proposed evaluation guideline is provided. Section 4 presents some test and evaluation results. Finally, a conclusion is provided in Section 5.

2 RELATED WORKS

The localization techniques can be classified into two main categories: Range-based and range-free. The range-based technique uses the distance or range information to determine the position of a node, such as Global Positioning System (GPS), Angle of Arrival (AoA), Time of Arrival (ToA), Time Difference of Arrival (TDoA), and Received Signal Strength Indicator (RSSI) (Alsheikh et al., 2014; Bekcibasi and Tenruh, 2014). As for the range-free techniques, the position estimation relies on connectivity information between two nodes (Singh and Sharma, 2015). They use radio connectivity to communicate among the nodes to estimate their locations, so no coordinates system is used. The position of the target node can be obtained by applying geometrics relations. Most range-

free techniques are based on the number of surrounding neighbors known as DV-Hop, Approximate Point-In-Triangulation test (APIT), or centroid system. DV hop (Hu and Li, 2013; Liu et al., 2016) estimates the range between nodes using the hop count. Indeed, by comparing and evaluating the performance and limitations of these methods, it is necessary to find a trade-off between centralized and distributed topologies combining low energy consumption and localization accuracy. Considering the application requirements and constraints, such as low energy consumption, secure data transmission, and efficiency, the localization technique can differ from one application to another. For this purpose, it is necessary to efficiently simulate and evaluate the localization technique before the realization and implementation to ensure better network performance. Various research efforts were performed to qualitatively classify and review the range-based localization techniques. Using the Web of Science database, research for English articles including the words "localization WSN* survey" or "Localization WSN* review" or "localisation WSN* survey" or "Localisation WSN* review" (* to include the plural) in their title, abstracts or keywords. The first-round results in 53 review papers related to localization techniques in WSN. After reading their abstracts, 42 papers are excluded because they are unrelated to reviews on Localization in WSN. The 11 remaining papers are analyzed here.

Authors in (Saad et al., 2018) classified localization techniques into five categories, but they focus only on the ranging technique, which is divided into range-based and range-free methods. An overview of the AoA, TDoA, and RSSI in the range-based methods is provided. Four range-free techniques are illustrated: Centroid, APIT, DV-Hop, and Amorphous. To compare the performance of the described techniques, some metrics are considered, including accuracy, scalability, cost, and power consumption. In (Sneha and Nagarajan, 2020), the localization techniques are classified as proximity-based, range-based, and range-free Localization. The factors influencing the localization measurement are studied. A general comparative analysis of only the range-based and range-free methods is discussed in terms of accuracy, cost, power consumption, and additional hardware requirements. Then, a comparison between RSSI, AoA, TDoA/ToA is performed by describing their advantages and disadvantages. In (Khelifi et al., 2015), the advantages and drawbacks of the range-based and range-free localization technique are described. Authors proved that it is hard to find the best algorithm, which dominates by all criteria. Authors in (Ismail et al., 2021) discussed the localization techniques us-

ing GPS, range-based methods including RSSI, ToA, AoA, and TDoA and range-free methods such as centroid, DV-Hop, Amorphous. The discussed reviews are summarized in Table 1.

Table 1: List of survey papers studying range-based localization technique.

Ref.	Comparison parameters
(Khelifi et al., 2015)	Energy, accuracy, complexity, hardware requirement, communication traffic, coverage
(Saad et al., 2018)	Accuracy, scalability, cost, power consumption
(Sneha and Nagarajan, 2020)	Accuracy, cost, power consumption, requirement for additional hardware
(Ismail et al., 2021)	Principal of operation, special hardware, attenuation problem, cost

Existing reviews perform the comparison with certain configuration parameter, such as the network size, number of nodes, deployment strategy, energetic model. Thereby, obtained results remain valid only for some instances, and there is no generalized evaluation combining all performance metrics. Indeed, there is no quantitative comparison aiming to improve the existing techniques. Thus, which algorithm is the best for a specific scenario remains open. These shortcomings motivate us to develop a methodology for a generalized comparison. The paper aims to develop a guideline that englobes the most relevant range-based localization techniques.

3 DESIGN OF THE EVALUATION GUIDELINE

In the localization method, the estimation of the coordinates of nodes varies from one technique to the other, where the network definition and characteristics describe the input parameters of the system. Since, the performance evaluation of these techniques, presents a multidimensionality problem, machine learning-based model offers a promising solution to support the use of various non-dependent input parameters. The localization evaluation guideline offers a generic comparison module of the basic localization techniques. The system is based on a multiple linear regression module, which enables an overview of the complete localization scenario, where a predictive module is attributed. In this guideline, the path loss exponent, the communication range of nodes, and the sending power all contribute to determining accuracy and position error. The choice of a suitable localization technique depends on the network specification and the application requirement and is highly influenced by these definitions and initializations. Obtained results are identified for the four main range-based localization techniques. It offers an understanding of suitable localization techniques in terms

of localization accuracy. The system includes five main blocks: Network initialization, localization instance, data preparation, multiple regression module, and evaluation (see Figure 1). In the network initialization, parameters related to the network definition are initialized.

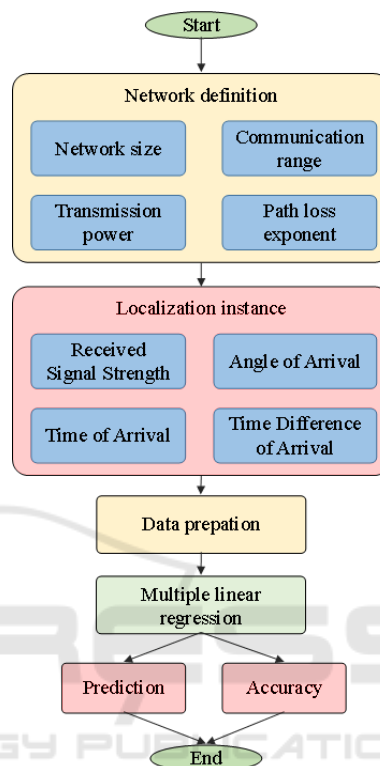


Figure 1: Block diagram of the evaluation guideline for range-based localization algorithms.

It contains information about the network size expressed as Width and Length, the path loss exponent to highlight the environment characterization (indoor/outdoor, free space/ interferences), the communication range of nodes, and the sending power. The input parameters are as follows:

- Transmission power of the sensor node;
- Network size of the area of interest, $Width \times Length$;
- Environmental coefficient: Path loss exponent η ;
- Communication range of sensor nodes R_c ;

Considering these inputs, the localization instance is created for each localization technique concerning its defined equation. Later, the calculated data of each localization instance are prepared as input to the regression module to estimate the position of nodes. After that, a prediction module and accuracy evaluation are carried out. Once all instances are created

and evaluated, the most suitable localization technique corresponding to the network initialization is proposed in terms of localization accuracy and position estimation error.

3.1 Localization Module

The first phase of the guideline revolves around four basic range-based wireless sensor localization algorithms. The algorithms are implemented individually using the algorithms themselves. Four different functions are responsible for the Localization, which run the following sequence.

- fun_{rssi} takes input and runs RSSI algorithm;
- fun_{toa} takes input and runs ToA algorithm;
- fun_{tdoa} takes input and runs TDoA algorithm;
- fun_{aoa} takes input and runs AoA algorithm.

All these algorithms take the same set of inputs, enabling respective functions to create the files and save the data for several iterations as a .csv file. This .csv is then later read by machine learning.

3.2 Regression Module

The performance and selection of machine learning algorithms are solely based on the available datasets. Although no algorithm has superiority over another algorithm, there are some traits and properties of each algorithm, which can be considered by tuning the hyper-parameters. The regression module enables the analysis of multi-factor data, which helps decide over the most suitable localization techniques based on the input parameters. The regression module permits to translate the localization choice into a mathematical model describing the relationship between the dependent and independent variables. More precisely, the regression module allows mapping the results based on numeric inputs. The general mathematical module of multiple linear regression is illustrated in Equation 1.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad (1)$$

where, for $i = n$ observations, x_i , y_i are the dependent and explanatory variables, respectively. β_0 is the y-intercept (constant term), β_p is the slope coefficient for each explanatory variable and is the residual value.

The input parameters of the regression module are the localization instances of each localization technique. Distances between nodes are estimated based on each range-based technique. Once the distances are estimated, a simple trilateration method is carried out to estimate the coordinates of nodes. 10000 samples for each localization technique are created based

on the experiment. 80 % of the total number of samples is defined to train the regression module to identify each localization technique. The remaining 20 % of samples are used later to evaluate and test the regression module.

4 TEST AND EVALUATION

The design of the proposed guideline is based on built, reusable modules, which use the Python programming language. Python provides a powerful and easy-to-use programming tool, which supports extensive machine learning computations. Also, it allows access to complete library support that integrates machine learning and mathematical models. A user-friendly, easy-to-use, and flexible guideline is proposed in this work. It allows the user to define its network parameters, like the network dimensions, the path loss exponent, and sensor nodes characteristics. Figure 2 illustrates the designed interface. The interface allows the user to define all network assumptions that characterize the test field. As output and based on the localization aspect, the decision of the suitable localization technique is made based on the accuracy.

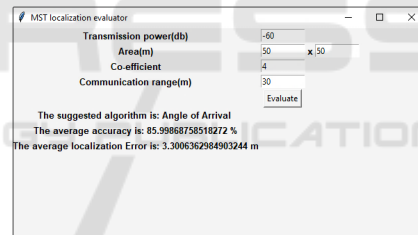


Figure 2: Illustration of the developed guideline; data inputs and network definition.

In the background, the input variables are collected to define the localization instance for each range-based technique. The most suitable localization technique is defined according to the used nodes, localization error, and inaccuracy. Indeed, the proposed system offers different graphical results to understand the comparison results better. The first graphical comparison is a histogram modeling, as presented in Figure 3. The illustrated results are for the localization error over various possible nodes collections. It presents a comparative analysis of the nodes deployed in the system. This is an accuracy graph, which shows the cumulative accuracy of the system. The x-axis shows the average Localization by plotting the error of each node, and the y-axis shows the total number of nodes. It gives an impression of the density of nodes and their respective localization errors.

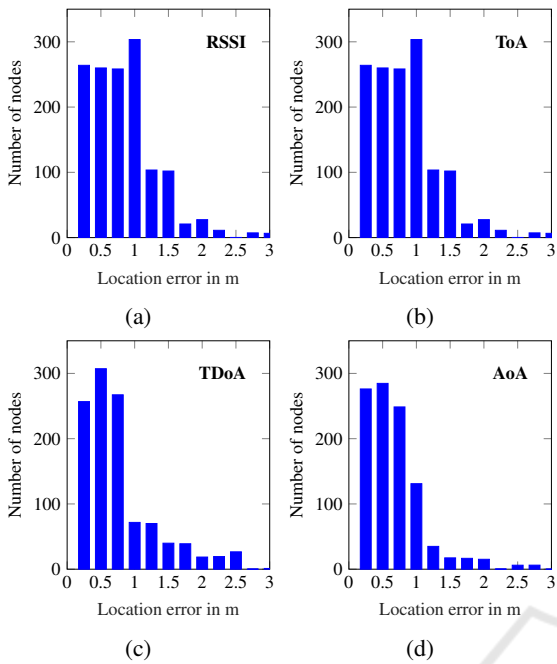


Figure 3: Error distribution of range-based localization technique for a network size of $15 \times 20 m^2$, communication range $R_C =$, transmission power of $-70 dBm$ and path loss exponent of $\eta = 4$.

In this work, the influence of the network definition is studied. In fact, during the localization phase, nodes are placed in a specific network with specific hardware characterization. Each parameter may directly impact the choice of the localization technique itself. In the following, the guideline is tested for different scenarios: (1) influence of network size, (2) influence of path loss exponent, and (3) influence of communication range of the installed nodes. Figure 4 summarize the obtained results for the three test scenarios. It proves that the choice of the suitable localization technique is strongly dependent on the network definition. For example, for large-scale networks, the outperforms other localization techniques (Figure 4a), whereas the works better for small-scale networks (Figure 4a), whereas the works better for small-scale networks (Figure 4a), whereas the works better for small-scale networks (Figure 4a). The path exponent presents the effect of the network’s scale impact, as it defines the environmental characteristic. It helps to identify if there exists some interference and shadowing in the network and if the signal is propagating in the line of sight. Figure 4b illustrates the effects of the path loss exponent on the choice of the Localization.

It is clear that for the free space environment ($\eta = 2$) (Figure 4c), the RSSI outperforms other techniques, whereas, in the case of industrial environments and buildings ($\eta = 4$), the AoA technique gives better results. In the last scenario, the communication range of the instated nodes is studied. The com-

munication range enables deciding if the network is covered or not, which helps reduce energy consumption and maintain communication between installed nodes. If the communication range increases, the localization capabilities increase, which helps reduce the measurement error.

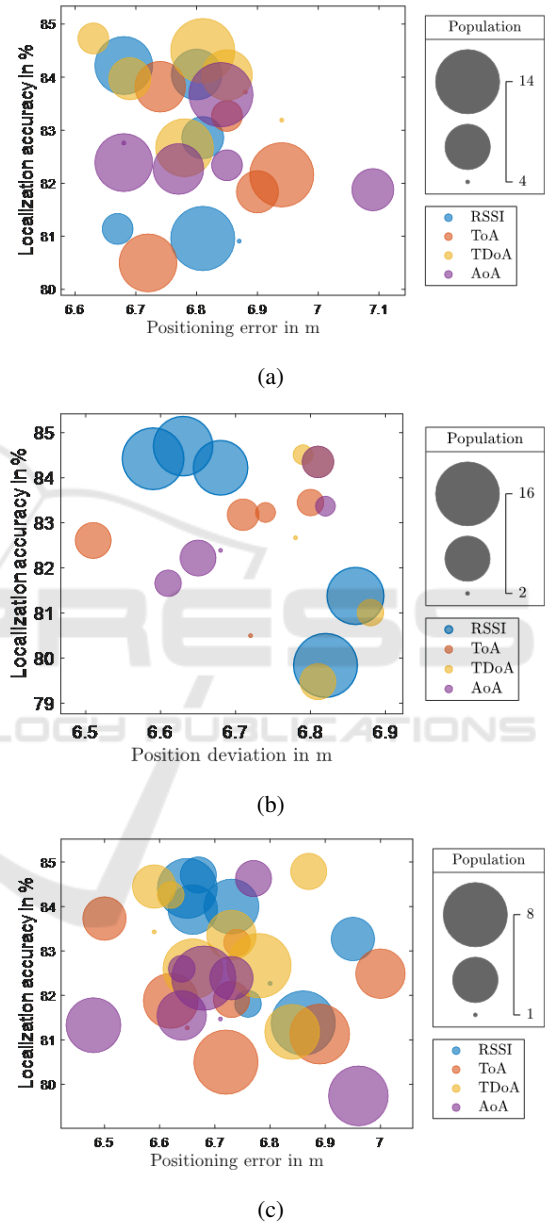


Figure 4: Performance evaluation of range-based localization techniques: (a) Influence of network size, (b) influence of path loss exponent and (c) influence of communication range.

Finally, the proposed guideline is validated using different types of localization algorithms, and each one is implemented using different methods. Many

Table 2: Comparison of the developed guideline to the result of some localization algorithms.

Ref.	Test field (m^2)	Accuracy (%)	Localization error (m)	Computation time (s)
(Weingartner et al., 2009)	20	98	0.5	2.01
(Weingartner et al., 2009) *	4 A + 25 NN	98.32	0.256	4.07
(Rahman et al., 2012)	100	81	4	10.54
(Rahman et al., 2012) *	15 A + 35 NN	89.23	5.25	3.2
(Ahmadi and Bouallegue, 2015)	20	95	0.409	3.2
(Ahmadi and Bouallegue, 2015) *	3 A + 15 NN	97.49	0.34	4.79
(Jin et al., 2010)	100	65	9	5
(Jin et al., 2010) *	30 A + 100 NN	93	4.32	5.84
(Wang et al., 2009)		75	5	3.33
(Wang et al., 2009) *	3 A + 1 NN	93.77	0.95	5.7

scenarios validate our guideline to study the effects of different parameters on the design. As seen in Table 2, obtained results are compatible with the original results reported by the selected localization algorithms in their papers. Different localization algorithms have been tested using the proposed guideline. The selected localization algorithm are chosen based on their achieved localization accuracy. Besides, the selected works implements some artificial intelligence technique for the determination of the nodes' location. We selected different works based on RSSI techniques only since its strongly dependent of the deployment environment and sending power of the nodes. As illustrated in Table 2, obtained results remain compatible with those reported by selected localization algorithms in their papers. In (Weingartner et al., 2009) authors propose an RSSI-based localization system. Their simulation achieved 98 % accuracy, compared to our guideline, which shows 98.32 % with a reduced error. In (Rahman et al., 2012), authors propose an RSSI-based system, which uses the regression Tree by comparing its performance with Least Squares Support Vector Regression and Multi Layers Perceptron Neural Network. The evaluation considers the localization error and the complexity of the algorithm. Using the regression tree, they reached 81 % of localization accuracy. Considering the same network definition, the proposed work in (Rahman et al., 2012) is tested using the developed guideline, where the localization accuracy reached 89 %. Similarly, authors in (Ahmadi and Bouallegue, 2015) propose a fingerprint-based localization scheme that considers the channel impulse response to computing the location of nodes. The distance and position estimation is carried out using non-parametric kernel regression. The localization accuracy reached 95 %, whereas, in the developed guideline, the total localization accuracy is around 97 %, with a minimal localization error of 0.34 m. In (Jin et al., 2010), the RSS-based lateration method is used to compute the location of nodes. In their work, the authors provided two approaches, regression-based and correlation-

based. The regression-based approach uses linear regression to discover a better fit of the signal propagation model between RSS and the distance. In contrast, the correlation-based approach utilizes RSS correlation in the local area to obtain more accurate signal propagation. As a result, they obtained 65 % for localization accuracy. In paper (Wang et al., 2009), RSSI information is used to estimate the position of nodes. The estimated localization accuracy is around 75 %, compared with the accuracy of the guideline of 93 %.

A preliminary study on the evaluation of range-based localization techniques is proposed in this work. An evaluation guideline is presented. It enables to build a better knowledge of the performance of the localization techniques with the definition of the network assumptions, mainly, the network size, communication range, and sending power. This preliminary study provides a good insight for evaluating the localization range-based technique, where the developed guideline is based on the different localization instances. However, a more detailed study and evaluation are required to ensure better functionality of the developed guideline in terms of dependency to the real-world scenario (e.g., Existence of obstacles, hardware failure) and dependency on the localization constraints (e.g., Energy consumption and lifetime). As a future perspective, a study on design constraints will be considered along with different localization techniques. Moreover, the multiple linear regression module will be compared with another possible alternative, which can improve the flexibility and robustness of the developed system.

5 CONCLUSIONS

A generic evaluation guideline based on multiple linear regression is developed for range-based localization techniques. The results enable a better choice of which standard localization technique to implement. The localization techniques are evaluated based on their accuracy and localization error. Having com-

pared the guideline results with the results of the selected localization algorithms, it was found that they had been consistent with the original results of these algorithms in their original papers. In the future, other localization algorithms can be incorporated into the guideline, which can be used by other researchers. For better assumptions considerations, it is necessary to include the energetic module and deployment strategies in the design. Moreover, the evaluation metrics (accuracy) can be extended to network lifetime, energy consumption, and communication.

REFERENCES

- Abdelhabib, B. and Brahim, B. (2008). Japrosim: a java framework for process interaction discrete event simulation. *Journal of Object Technology*, 7(1):103–119.
- Ahmadi, H. and Bouallegue, R. (2015). Rssi-based localization in wireless sensor networks using regression tree. In *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 1548–1553. IEEE.
- Ahmed Mansoor, S. and Irtaza, M. (2019). *IOT localization in Smart Buildings*. PhD thesis, Department of Computer and Electrical Engineering, COMSATS University.
- Alsheikh, M. A., Lin, S., Niyato, D., and Tan, H.-P. (2014). Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Communications Surveys & Tutorials*, 16(4):1996–2018.
- Bekcibasi, U. and Tenruh, M. (2014). Increasing rssi localization accuracy with distance reference anchor in wireless sensor networks. *Acta Polytechnica Hungarica*, 11(8):103–120.
- Egea-Lopez, E., Vales-Alonso, J., Martinez-Sala, A. S., Pavon-Marino, P., and Garcia-Haro, J. (2005). Simulation tools for wireless sensor networks. In *Summer Simulation Multiconference-SPECTS*, volume 2005.
- El Houssaini, D., Khrijji, S., Besbes, K., and Kanoun, O. (2020). Iot based tracking of wireless sensor nodes with rssi offset compensation. In *2020 17th International Multi-Conference on Systems, Signals & Devices (SSD)*, pages 897–902. IEEE.
- Hogie, L., Bouvry, P., and Guinand, F. (2006). An overview of manets simulation. *Electronic notes in theoretical computer science*, 150(1):81–101.
- Hu, Y. and Li, X. (2013). An improvement of dv-hop localization algorithm for wireless sensor networks. *Telecommunication Systems*, 53(1):13–18.
- Ismail, M. I. M., Dziauddin, R. A., Ahmad, R., Ahmad, N., Ahmad, N. A., and Hamid, A. M. A. (2021). A review of energy harvesting in localisation for wireless sensor node tracking. *IEEE Access*.
- Jin, Y., Soh, W.-S., and Wong, W.-C. (2010). Indoor localization with channel impulse response based fingerprint and nonparametric regression. *IEEE Transactions on Wireless Communications*, 9(3):1120–1127.
- JSim. Jsim home page. <https://www.physiome.org/jsim/>. Accessed: 06-12-2021.
- Kanoun, O., Bradai, S., Khrijji, S., Bouattour, G., El Hous-saini, D., Ben Ammar, M., Naifar, S., Bouhamed, A., Derbel, F., and Viehweger, C. (2021). Energy-aware system design for autonomous wireless sensor nodes: A comprehensive review. *Sensors*, 21(2):548.
- Khelifi, M., Moussaoui, S., Silmi, S., and Benyahia, I. (2015). Localisation algorithms for wireless sensor networks: A review. *International Journal of Sensor Networks*, 19(2):114–129.
- Khrijji, S., El Houssaini, D., Kammoun, I., and Kanoun, O. (2018). A fuzzy based energy aware unequal clustering for wireless sensor networks. In *International Conference on Ad-Hoc Networks and Wireless*, pages 126–131. Springer.
- Liu, J., Wang, Z., Yao, M., and Qiu, Z. (2016). Vn-apit: Virtual nodes-based range-free apit localization scheme for wsn. *Wireless Networks*, 22(3):867–878.
- Naguib, A., Hamouda, A., Abdel-Mageid, S., and Marie, M. Development and validation of a localization framework for wireless sensor networks.
- Nsam. Ns3, network simulator. <https://www.nsnam.org/>. Accessed: 06-12-2021.
- Omnet. Omnet++, discrete event simulator. [omnet++ discrete event simulator. https://omnetpp.org/](https://omnetpp.org/). Accessed: 06-12-2021.
- Paul, A. K. and Sato, T. (2017). Localization in wireless sensor networks: A survey on algorithms, measurement techniques, applications and challenges. *Journal of sensor and actuator networks*, 6(4):24.
- Rahman, M. S., Park, Y., and Kim, K.-D. (2012). Rss-based indoor localization algorithm for wireless sensor network using generalized regression neural network. *Arabian journal for science and engineering*, 37(4):1043–1053.
- Saad, E., Elhosseini, M., and Haikal, A. Y. (2018). Recent achievements in sensor localization algorithms. *Alexandria engineering journal*, 57(4):4219–4228.
- Singh, S. P. and Sharma, S. (2015). Range free localization techniques in wireless sensor networks: A review. *Procedia Computer Science*, 57:7–16.
- Sneha, V. and Nagarajan, M. (2020). Localization in wireless sensor networks: A review. *Cybernetics and Information Technologies*, 20(4):3–26.
- Wang, Y., Xu, X., and Tao, X. (2009). Localization in wireless sensor networks via support vector regression. In *2009 Third International Conference on Genetic and Evolutionary Computing*, pages 549–552. IEEE.
- Weingartner, E., Vom Lehn, H., and Wehrle, K. (2009). A performance comparison of recent network simulators. In *2009 IEEE International Conference on Communications*, pages 1–5. IEEE.