A Big Data Analysis Framework for Localization in Cloud based Sensor and Robot Networks

Zuleyha Akusta Dagdeviren[1]*, Orhan Dagdeviren[2]

[1]* Ege University, International Computer Institute, Izmir, Turkey (ORCID: 0000-0001-9365-326X), zuleyhaakusta@gmail.com
[2] Ege University, International Computer Institute, Izmir, Turkey (ORCID: 0000-0001-8789-5086), orhan.dagdeviren@ege.edu.tr

Abstract

Internet of Things (IoT) is constructed from billions of sensor devices connected over Internet. Wireless sensor networks (WSNs) are very important communication technologies of IoT for providing large scale data from the environment. Since the data collected grows exponentially, the design of cloud based big data analysis techniques is crucial. Localization in sensor networks is finding the location of a node based on the reference nodes's coordinates. In many applications such as target tracking and military surveillance, providing localization is necessary. Manually entering the locations of nodes during the deployment phase is not applicable especially for large scale sensor networks. Besides, integrating a GPS receiver to each sensor node is a very costly solution and may not be affordable for large scale networks. Since sensor nodes are mostly battery-powered, design of an energy-efficient localization method is highly desirable to prolong the lifetime of the network. Existing localization techniques may require many message transmission which causes high energy consumption. To tackle with this problem, we propose an energy-efficient localization framework in this paper. A mobile robot is traversed along the sensing area and communicate with sensor nodes to localize these nodes. Different than the previous approaches, our proposed approach requires only 3 messages per node. Besides, most of the execution needed for the localization is not carried by ordinary nodes. We simulate our proposed approach in ns2 simulator. We measure the localization quality and energy consumptions of our proposed approach with its counterparts. Also, we measure the localization quality and energy consumption against varying node counts and degrees. From extensive simulation results, we obtain that the localization qualities of our proposed approach is significantly better than its competitors. Besides, the energy consumption of our proposed algorithm is just 0.06 J per node and far more better than the distributed algorithm. Conclusively, our proposed framework is a significant candidate for IoT and big data applications requiring energy-efficient localization.

Keywords: Internet of Things, Big Data, Cloud Computing, Wireless Sensor Networks, Localization, Robots.

Bulut Tabanlı Sensör ve Robot Ağlarında Konumlandırma için Büyük Veri Analizi Çerçevesi

Öz

Nesnelerin İnterneti (IoT), İnternet üzerinden bağlı milyarlarca sensör cihazdan oluşturulmaktadır. Kablosuz sensör ağları (KSA'lar), ortamdan büyük ölçekli veri sağlamak için IoT'nin çok önemli iletişim teknolojileridir. Toplanan veriler katlanarak büyüdüğü için bulut tabanlı büyük veri analizi tekniklerinin tasarımı çok önemlidir. Sensör ağlarda konumlandırma, referans düğümlerin koordinatlarına göre bir düğümün konumunu bulmaktır. Hedef takibi ve askeri gözetim gibi birçok uygulamada konumlandırmanın sağlanması gerekmektedir. Dağıtım aşamasında düğüm konumlarının manuel olarak girilmesi, özellikle büyük ölçekli sensör ağları için geçerli değildir. Ayrıca, her sensör düğümü için bir GPS alıcısı entegre etmek çok maliyetli bir çözümdür ve büyük ölçekli ağlar için uygun olmayabilir. Sensör düğümleri çoğunlukla pille çalıştığından, enerji açısından verimli bir konumlandırma yönteminin tasarımı, ağın ömrünü uzatmak için oldukça önemlidir. Mevcut konumlandurma teknikleri, yüksek enerji tüketimine neden olan birçok mesaj iletimini gerektirebilir. Bu sorunu çözmek için, bu makalede enerji açısından verimli bir konumlandırma çerçevesi

---

* Sorumlu Yazar: zuleyhaakusta@gmail.com

önermekteyiz. Bir mobil robot, algılama alanı boyunca hareket ettirilmekte ve bu düğümleri yerelleştirmek için sensör düğümleriyle iletişim kurmaktadır. Önceki yaklaşımlardan farklı olarak, önerilen yaklaşımımız düğüm başına yalnızca 3 mesaj gerektirmektedir. Ayrıca, konumlandırma için gereken işlem çoğu sıradan düğümler tarafından gerçekleştirilmemektedir. Önerdiğimiz yaklaşımı ns2 benzetim ortamında simüle etmekteyiz. Önerdiğimiz yaklaşımın konumlandırma kalitesini ve enerji tüketimini rakipleriyle karşılaştırmaktayız. Ayrıca, konumlandırma kalitesini ve enerji tüketimini değişen düğüm sayılarına ve derecelerine göre ölçmekteyiz. Kapsamlı simülasyon sonuçlarından, önerdiğimiz yaklaşımın konumladırma kalitesinin rakiplerinden önemli ölçüde daha iyi olduğunu elde etmekteyiz. Ayrıca, önerdiğimiz algoritmanın enerji tüketimi düğüm başına sadece 0,06 J'dir ve bu sonuç dağıtık algoritmadan çok daha iyidir. Sonuç olarak, bu makalede önerilen çerçevemiz, enerji verimli konumlandırma gerektiren IoT ve büyük veri uygulamaları için önemli bir adaydır.

Anahtar Kelimeler Nesnelerin İnterneti, Büyük Veri, Bulut Bilişim, Konumlandırma, Robotlar.

1. Introduction

Internet of Things (IoT) is a network of billions of sensor nodes connected through Internet. Wireless sensor networks (WSNs) are indispensible IoT technologies used in the communication layer. Since the data collected by sensor nodes grow exponentially, design of big data analysis techniques located in cloud systems are of utmost importance (Harb et al., 2017)(Kim et al., 2019)(Tripathi et al., 2021). Localization in WSN involves determining the location of the sensor node based on the other nodes' coordinates (Alhmeidat and Yang, 2007) (Xiao and Ouksel, 2006). In many circumstances, it is necessary for a node in a WSN to be aware of its location. In large scale sensor networks, manually configuring the location into each node during the deployment is not applicable. Besides this, GPS is a very costly solution for large scale WSN; thus localization process is very important to reduce the costs (Karl and Willig, 2005). The localization objectives for sensor networks can be listed as follows:

- The error ratio of the localization should be as small as possible. The error ratio of the localization algorithm for a single node is $|M\text{-}R| / D$ where M is the measured distance from the origin, R is the real distance from the origin and D is the diagonal of the sensing area.
- The number of nodes equipped with GPS (reference nodes) should be as small as possible to avoid costs.
- The localization algorithm should consume low energy and should not dependent of the underlying protocols.

Existing localization techniques can be categorized as centralized and distributed. Centralized localization techniques that involve transmitting data to a central node or a sink node in order to calculate the location for each node are given in (Doherty et al., 2001), (Shang et al., 2003). After the localization computations are handled at the powerful central node, the results are spanned back to every node in the network. On the other hand, Niculescu (Niculescu and Nath, 2003) and Nagpal (Nagpal et al., 2003) proposed distributed localization approaches that rely on flooding technique to calculate the position of the nodes. In distributed localization approaches, generally some nodes are pointed as to be the reference roots in the network and they flood their localization data into the network. Multilateration technique is based on the following parameters from multi reference nodes: reference node locations together with the signal strength or time of arrival of signal. In the concept of iterative multilateration, the new localized ordinary nodes in the network can also be represented as the new reference nodes. Savarese (Savarase et al., 2002) proposed an iterative multilateration approach, in which the ordinary nodes

can be set as the new reference nodes if they are localized with the use of data from at least three reference nodes whose channels are not intersected. A detailed survey on localization approaches can be found in (Patwari et al., 2005). Localization methods mentioned above may require many message exchanges that may cause high energy consumption. Since WSNs generally consist of battery-powered sensor nodes, minimizing energy consumption is crucial to maximize the network lifetime. In this manner, we propose an energy-efficient localization technique in this paper. In contrast to the previous work, our framework requires only 3 messages per node. We implement our proposed approach in ns2 simulator. We explain our design and the results of our implementation in the rest of this paper.

2. Material and Method

2.1. Network Model

In our hybrid network model shown in Fig. 1, there are three types of devices: Ordinary sensor device (node), sink sensor device (sink node) and mobile robot device. The following assumptions are made about the network:

- Each device has distinct device id. The sensor devices are stationary. The robot device is mobile and has a map of the network area. The robot can plan a path.
- Links between all devices are symmetric. Thus if there is a link from u to v, there exists a reverse link from v to u. The distances between nodes determine the quality of the links. The quality of the link reduces with the distance. The model can be represented as an edge weighted graph.
- The sensor devices do not know their positions. They are not equipped with a position tracker like a Global Positioning System (GPS) receiver. The robot device can learn its position via GPS.
- The ordinary sensor devices are equal in terms of processing capabilities, radio, battery and memory. The sink sensor device (IoT gateway) and mobile robot may communicate over Internet to send the aggregated data to the user. The robot device has all the capabilities mentioned above and enough energy, memory, radio and CPU resources to perform localization and clustering. The robot device is not responsible for collecting the sensed data from the ordinary sensor devices and forwarding the sensed data to the user, but it may be configured by the user over Internet.
- Big data analyzer is deployed on a cloud system where users can connect to gather information. Collected data is processed in this system.
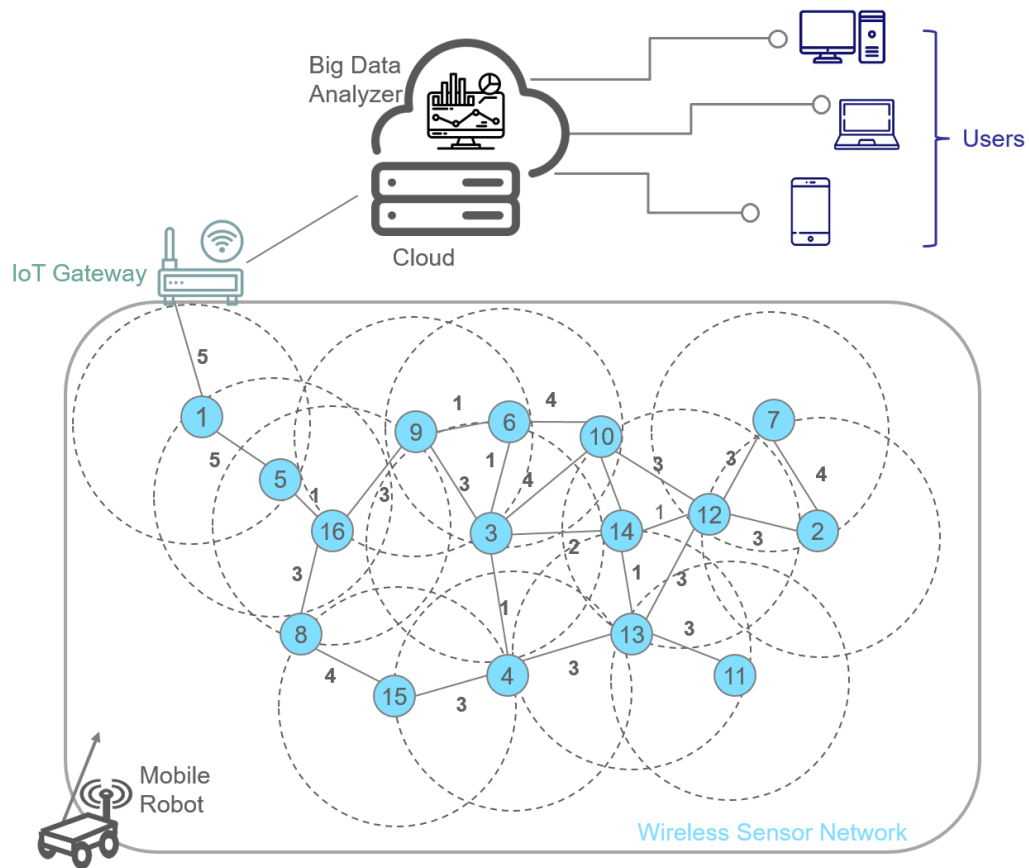
Figure 1. The Network Model

2.2. Proposed Framework

In order to decrease energy consumption caused by localization, we propose to use a robot that computes the position of each node while moving along the predefined path. This path is configured by the robot to reach every node in the network area as shown in Fig. 2. The robot uses trilateration as a localization technique. In this technique, three distance measurements and three positions are needed to calculate the position of the node. The robot's recorded positions should be non-collinear to use trilateration for localization technique. This is one of the reasons why we use a zigzag path, which allows robot to select positions along the path. The other reason is providing robot to reach all nodes in the network area. Any other path than zigzag path, that provides collection of non-collinear positions from nodes, as well as visiting all nodes in the network can be used.

In our localization algorithm, the robot moves along the predefined path, periodically stops and broadcasts a *REQUEST(x,y)* message to the sensor nodes. When a node receives *REQUEST(x,y)* message from the robot, it starts a timer that fires at (*node id* x *message transmission time*) seconds later if it has not received three *REQUEST* messages having collinear *x* and *y* coordinates. When the timer is fired, the node sends a *REPLY* message to the robot. This timer mechanism provides a time division multiple Access (TDMA) for the ordinary nodes thus collisions are prevented. When the robot receives a *REPLY* message, it records the RSS of the *REPLY* message and records its position at the time it received *REPLY* message from the node. These RSS values and the position knowledge are necessary for the robot to calculate each node's position.

In Fig. 3, how the robot uses trilateration technique to calculate a node's position is depicted. In order to calculate node 8's position, robot collects three *REPLY* messages from node 8 at different positions (L1(x,y) to L3(x,y)). By using distances and the robot's positions, robot calculates the node's position with trilateration technique. The equation for the trilateration are given in Eq. 1. For $1 \leq i \leq 3$, (X_{L1}, Y_{L1}), (X_{L2}, Y_{L2}), and (X_{L3}, Y_{L3}) are the selected coordinates of the mobile robot and d_{RSS1}, d_{RSS2}, d_{RSS3} are the derived distance values from the recorded RSS. By using the three derived equations, we find X_{node}. With the aim of decreasing energy consumption, after sending three *REPLY* messages, there is no need for the node to be awake; so it sleeps up to the finishing time of the robot's movement. In Observation 1, we summarize the theoretical basis of our approach. In Theorem 1, we give the message complexity of our framework.

$$(X_{node} - X_{L_i})^2 \text{ - } (Y_{node} - Y_{L_i})^2 = d_{RSS_i}^2 \tag{1}$$

Observation 1. To localize nodes with trilateration, 3 non-collinear coordinates are needed.

Theorem 1. *The message complexity of the framework is O(3) per node.*

Proof. Each node only replies 3 *REQUEST* messages since localization of a node in trilateration requires 3 messages coming from the robot. Thus the message complexity of the framework per node is O(3). □
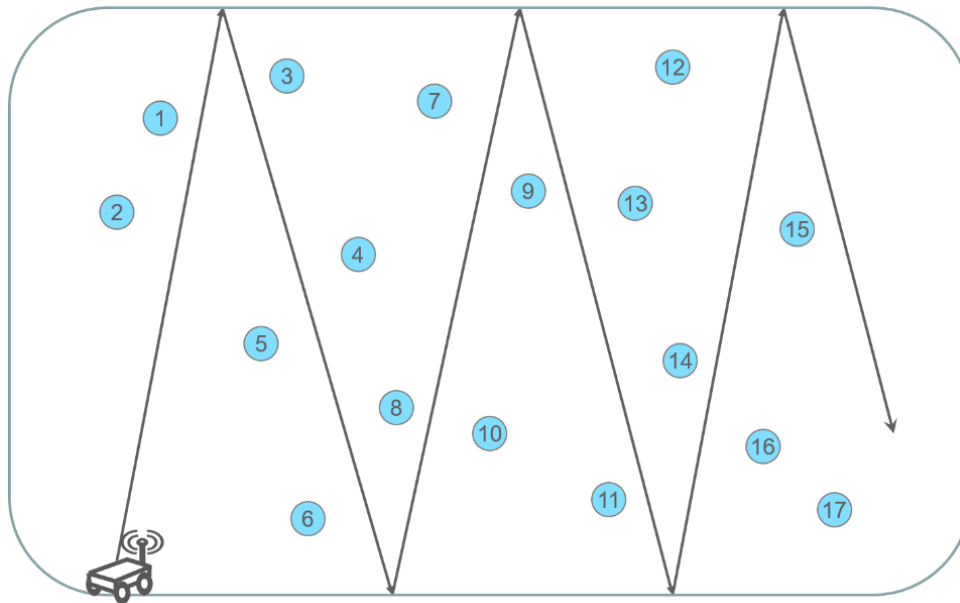
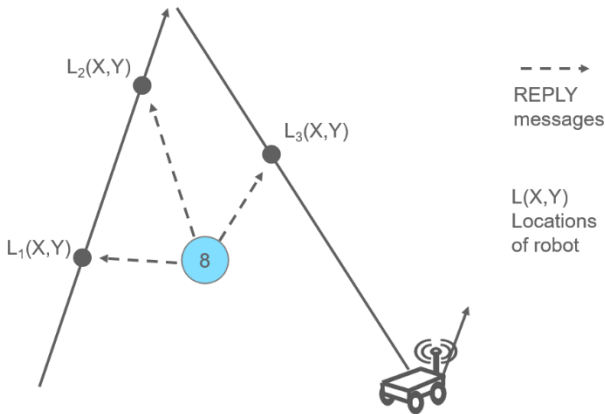*Figure 2. The Robot Moving in Predefined Path*

Figure 3. Localization Method

3. Results and Discussion

We implemented our framework in *ns2* (VINT, 2021). We generated randomly connected networks with 100 to 400 nodes that are uniformly distributed. The nodes are configured with IEEE 802.11 radio and MAC standards readily available in *ns2* simulator. Our algorithms only require *send*, *receive* primitives which are important services required from medium access control and physical layer, thus various types of lower layer standards can be used. Two way ground is used as the propagation model. Each node is equipped with a battery which is randomly assigned an energy value between 1J and 100J. The transmission power is 0.660 mW, the received power is 0.395 mW, and the communication range of a sensor node is 250 m. We measured the performance of the algorithms for average node degrees varying between 4, 5, and 6. To vary degrees, diferent flat surface areas were chosen as shown in Table 1.

To measure the clustering quality of the CDS algorithms, we use the number of clusters as the cluster quality metric since CDS algorithms are generally targeted to reduce the set size to approximate minimum CDS. When degree is increased, the cluster count produced by BICOT-CDS is decreased as shown in

Fig. 4.a since a cluster head may dominate more nodes in dense networks. As seen in Fig. 4.b, the cluster count generated by BICOT-CDS is approximately half of the number of clusters generated by DIST-CDS. As the network size increases, BICOT-CDS performs surely better than DIST-CDS.

TABLE 1. Size of Surface Areas (X x Y (m))

Node Count / Degree	4	5	6
100	2700 x 1200	2520 x 1200	2340 x 1040
200	5100 x 1200	4760 x 1120	4420 x 1040
300	7800 x 1200	7280 x 1120	6760 x 1040
400	10200 x 1200	9520 x 1120	8840 x 1040

All the nodes except the robot device in the simulation area are stationary. Localization accuracy, clustering quality and energy consumptions of the algorithms are measured. Table 2 summarizes the simulation parameters. We implemented an iterative trilateration algorithm to compare its localization performance with our framework. In the iterative trilateration technique (Karl and Willig, 2005) that we implemented, reference nodes send their positions to the other nodes in order to localize them. Once a node finds its position, it becomes a reference node.

TABLE 2. Simulation Parameters

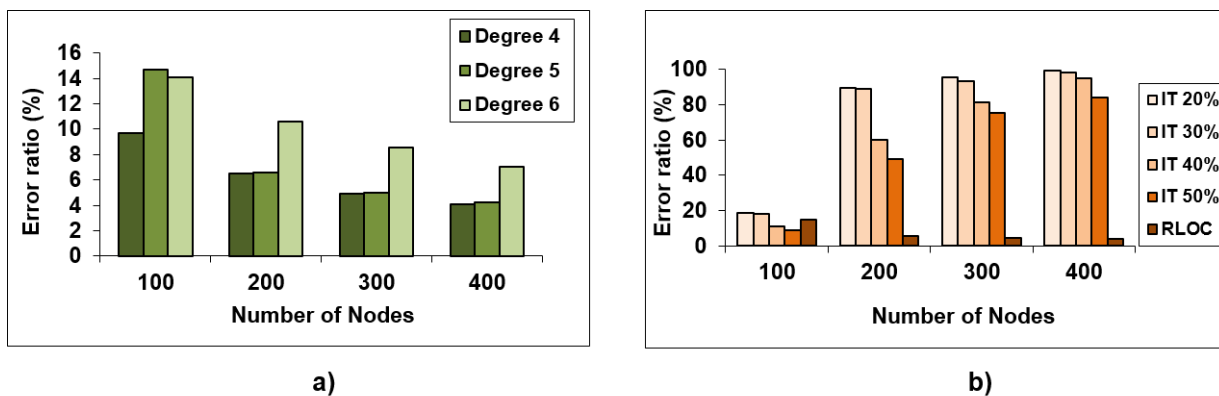| Node Distribution	Random
Node Mobility	Robot is mobile, sensor nodes are stationary
Number of Sensors	100 - 400
Medium Access Control	802.11
Transmission Power	0.660w
Receiving Power	0.395w
Initial Energy	Randomly between 1J and 100J
Transmission Range	250 m
Node degrees	4, 5 and 6

Figure 4. a) Error Ratios of RLOC vs. Node Count and Degree b) Error Ratios of Algorithms vs. Node Count
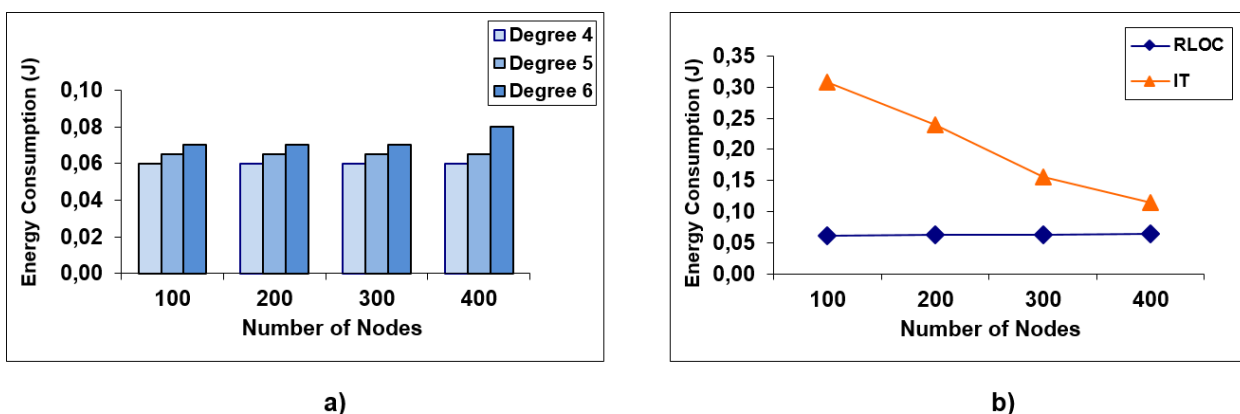

Figure 5. a) Energy Consumption of RLOC vs. Node Count and Degree b) Energy Consumption of Algorithms

3.1. Localization Performance

We firstly measured the localization accuracy of our framework against the varying degrees. Fig. 4.a shows the error ratio of the proposed localization approach. We calculated the overall error ratio of the algorithm as the average of all nodes'. The localization error ratio slightly increases as the network degree increases and localization error ratio decreases as the number of nodes increases as shown in Fig. 4.a. This shows that our localization approach is scalable.

We compared the localization error ratios of the iterative trilateration (IT) and robot based localization (RLOC). We varied the percentage of the reference nodes from 20% to 50%. As shown in Fig. 4.b, although the localization error ratio of IT decreases as the percentage of the reference nodes increases, the localization error ratio of the RLOC is far more smaller than the error ratios produced by IT.

3.2. Energy Consumption

We measured the energy consumptions of the algorithms. As shown in Fig. 5.a, when degree increases, the energy consumption per node slightly increases.

We compared the energy consumption of our framework and distributed approaches in Fig. 5.b. Our framework totally consumes just 0.06 J per node. As also shown in Fig. 5.b, our framework outperforms compared to the distributed approaches.

Conclusively, our framework's localization performance is far more better and energy consumption is significantly lower than its counterparts.

4. Conclusions and Recommendations

IoT is a very large scale network that includes billions of devices. WSNs collect data from the environment and the size of this data grow exponentially. Thus, the design of cloud based big data analysis methods is of utmost importance. Localization in WSNs aims to find the location of nodes based on other nodes' coordinates. Military surveillance and target tracking are example WSN applications in which maintaining localization is crucial.

To deal with this important problem, we propose a localization framework in this paper. In our framework, a mobile robot is traversed along the network area to provide a localization service to ordinary nodes. The robot sends a *REQUEST* message and collects *REPLY* messages from ordinary nodes. Our proposed algorithm requires only 3 messages per node different than the previous approaches.

We implement our proposed algorithms in ns2 simulator and measure the localization quality and energy consumptions. We reveal that our algorithm outperforms its competitors in terms of both localization quality and energy consumptions.

References

Alhmiedat,T. A. and Yang, S. H. (2007) A survey: Localization and tracking mobile targets through wireless sensors network. *PGNet*.

Doherty, L., Pister, K. S. J., and Ghaoui, L. E. (2001) Convex position estimation in wireless sensor networks. *Proc. of the 12th Annual Joint Conf. of the IEEE Computer and Communications Societies*, Anchorage, Alaska, 22-26 April, pp. 1655-1663. IEEE Computer Society, Washington.

Harb, H., Makhoul, A., Idrees, A., Zahwe and O. and Taam, M.. (2017) Wireless Sensor Networks: A Big Data Source in Internet of Things. *International Journal of Sensors, Wireless Communications and Control*.

Karl, H. and Willig, A. (2005) *Protocols and architectures for wireless sensor networks.* John Wiley & Sons.

Kim, B.-.S, Kim, K.-I., Shah, B., Chow, F. and Kim, K. H. (2019) Wireless Sensor Networks for Big Data Systems, *Sensors* 19, no. 7, 1565.

Nagpal, R., Shrobe, H., and Bachrach, J. (2003) Organizing a global coordinate system from local information on an ad hoc sensor network. *Proc. of the 2nd Int. Workshop on Information Processing in Sensor Networks*, Palo Alto, CA, USA, 22-23 April, pp. 333-348. Springer-Verlag, Berlin.

Niculescu, D. and Nath, B. (2003) DV based positioning in ad hoc networks. *Telecommunication Systems*, 22, pp. 267-280.

Patwari, N., Ash, J. N., Kyperountas, S., Hero, A. O., Moses, R. L., and Correal, N. S. (2005) Locating the nodes: Cooperative localization in wireless sensor networks. *IEEE Signal Processing Magazine*, 22(4), 54-69.

Savarese, C., Langendoen, K., and Rabaey, J. (2002) Robust positioning algorithms for distributed ad-hoc wireless sensor networks. *Proc. of USENIX Technical Annual Conf.*, pp. 317-328.

Shang, Y., Fromherz, M. P. J., Ruml, W., and Zhang, Y. (2003) Localization from mere connectivity. *Proc. of the 4th ACM Int. Symp. on Mobile Ad Hoc Networking and Computing*, Annapolis, Maryland, USA, 1-3 June, pp. 201-212. ACM Press, New York.

Tripathi, A. K., Sharma, K., Bala, M., Kumar, A., Menon, V. G. and Bashir, A. K. (2021) A Parallel Military-Dog-Based Algorithm for Clustering Big Data in Cognitive Industrial Internet of Things, *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2134-2142.

Xiao, L. and Ouksel, A. M. (2006) Scalable self-conflguring integration of localization and indexing in wireless ad-hoc sensor networks. *Proc. of the 7th Int. Conf. on Mobile Data Management*, Nara, Japan, 9-13 May. IEEE, Washington.

VINT project. (2021) Network Simulator version 2 (NS-2). *Technical Report*, available from: http://nsnam.sourceforge.net/wiki/index.php/Main_Page.