A Self-organized Sensor Data Path Selection Method in Internet of Things

Researcher: Milad Khademi Nori

Supervisors: Saeed Sharifian & Amirhossein Rezaei

May 18, 2019

Electrical Engineering Department at Amirkabir University of Technology

Agenda

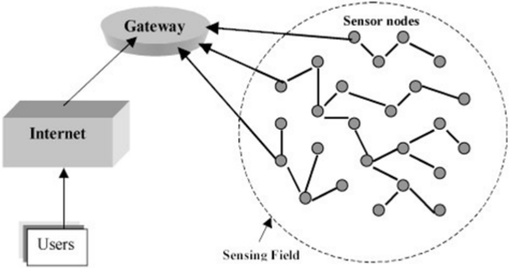
- Internet of Things (IoT)
  - Why IoT? & What is IoT?
  - Layers of IoT & Our Intended (Studied) Layer
  - Sensor Data Path Selection Problem
- Current Protocols for Sensor Data Path Selection Problem
  - Non-classic Protocols & Their Pros and Cons
  - Classics Protocols & Their Pros and Cons
  - Literature Review
- Our Proposed Protocol
  - Euclidean Distance Matrix (EDM) & Low-rank Matrix Reconstruction with Noise
- Simulations
- Conclusion

Internet of Things

- Why IoT?
  - Information is a great way to reduce waste and increase efficiency, and that's really what the internet of things provides.
- What is IoT?
  - The (IoT) is the extension of internet connectivity into physical devices and everyday objects.

- Layers of IoT
  - Things (Sensors)
  - Connectivity
  - Applications

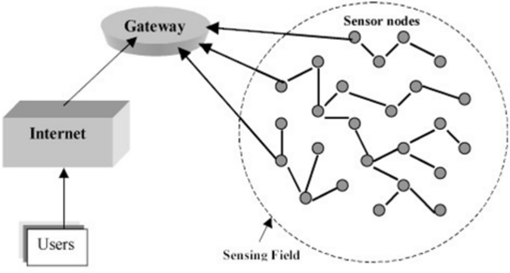
Things (Sensors) Layer

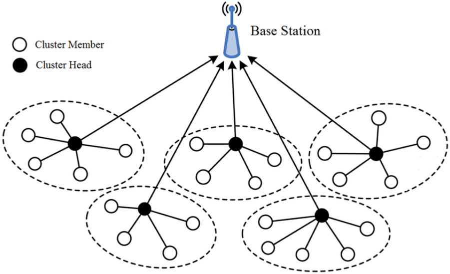


- Things are our worlds digital nervous system, hence the major task of things is to gather all of the environmental parameters, aggregate and forward them to the Base Station (BS) which is the gateway to the connectivity layer or internet backbone using a variety of routing protocols [1].
- Obstacles
  - Occasional energy shortage
  - Stochastic nature of Energy Harvesting Rate.
  - A need for an efficient protocol [2]

Sensor Data Path Selection Protocol Types

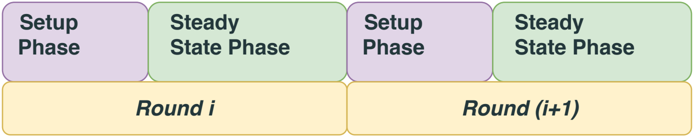
- Flat protocols
  - In the flat protocols, network works as a whole and all things play the same role [2].
- Hierarchical protocols
  - In the hierarchical protocols, things are partitioned into a number of clusters and some things are used as Cluster Heads abbreviated as CHs [2].
  - Hierarchical protocols outperform flat protocols in term of energy efficiency [2].





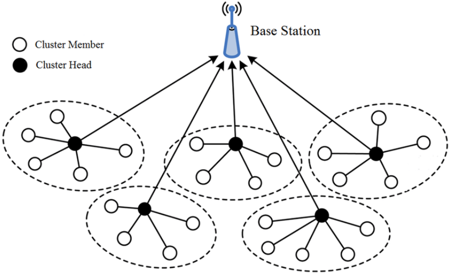
Hierarchical Protocols

- In hierarchical protocols, two decisions mainly determine the energy-efficiency of the protocols [1].
  - Clustering
  - Cluster Head Selection
- In hierarchical protocols, networks work in consecutive rounds. Each round consists of:
  - Setup phase
  - Steady-state phase



Non-classic Protocols & Classic Protocols

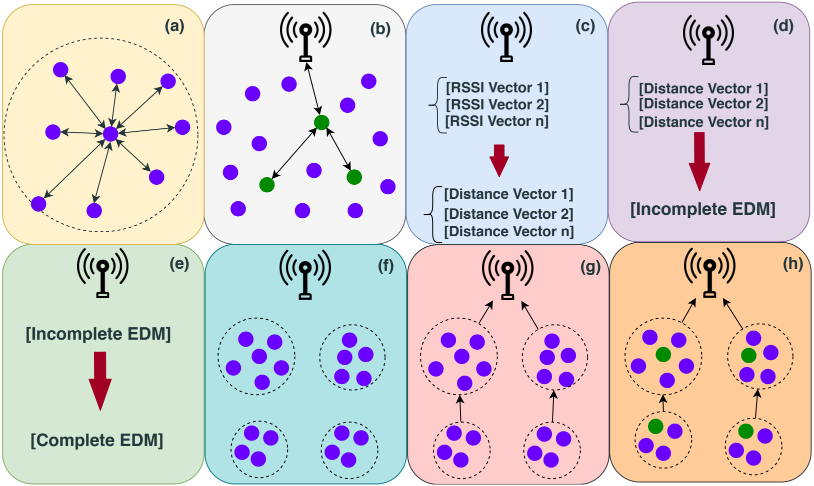
- Non-classic Protocols (LEACH [3], DEARER [1], EPSO-CEO [4], and WOA-C [5])
  - They do not rely on GPS.
  - They form inefficient clusters.
  - High energy-consumption.
  - High energy-shortage.
- Classic protocols (KCA [6] and Mk-means [7])
  - They can form clusters efficiently.
  - They rely on GPS module [8].
    - GPS increases implementation cost [8].
    - GPS increases energy consumption [8].
    - GPS works poorly in indoor applications [8].



Protocols	Inter	NGPS	ResEn	HarEn	Cov	Clas	Dis
LEACH 2000 [9]	×	√	×	×	×	×	√
LEACH 2017 [10]	√	√	×	×	×	×	×
LEACH 2016 [11]	×	√	√	×	×	×	×
LEACH 2016 [12]	×	√	×	×	√	×	√
DEARER 2016 [13]	×	√	×	√	×	×	×
GEEC 2015 [14]	×	√	√	√	×	×	×
KASSAN 2018 [15]	×	×	√	×	×	×	×
K-means 2016 [16]	×	×	√	×	√	√	×
EECPK 2016 [17]	×	×	√	×	√	√	×
Fuzzy 2018 [18]	×	×	√	×	√	√	×
K-medoids 2018 [19]	×	×	×	×	√	√	×
EMBEDMENT [20]	√	√	√	√	√	√	√

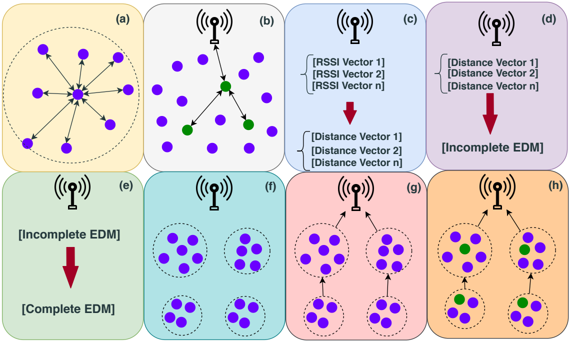
Inter: Inter-cluster communication, NGPS: No-GPS, ResEn: Residual energy, HarEn: Harvesting energy, Cov: Coverage, Clas: Classic Clustering, Dis: Distributed

Our Proposed Protocol



First Step

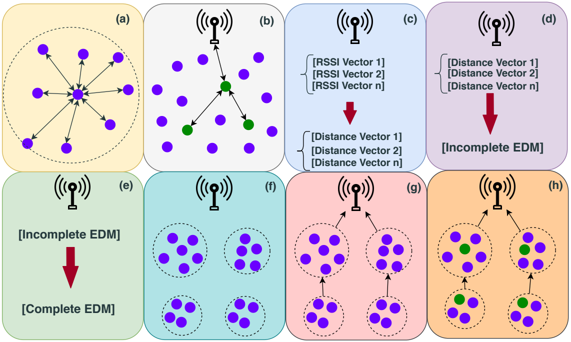
- RSSI measurement
- RSSI collection
- Storing RSSI in a vector
- Send and receive must happen more than 20 times



$$\begin{cases} s_1 = [\cdot]_{(N+1)\times 1} & \text{thing number } 1 \\ \vdots & \quad\vdots \\ s_i = [\cdot]_{(N+1)\times 1} & \text{thing number } i \\ \vdots & \quad\vdots \\ s_{(N+1)} = [\cdot]_{(N+1)\times 1} & \text{thing number } (N+1) \end{cases} \tag{1}$$

Second Step

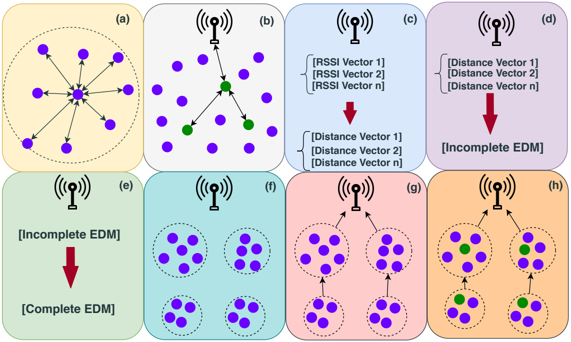
- RSSI transmission
- One problem!
- Low-range
- An elementary clustering should be performed. This could be LEACH with 10 % CHs.



- Flooding
- BS, in that end, receives all RSSI measurement vectors
- RSSI helps BS to estimate distances between things

Third Step

- Different Path Loss (PL) models for each environment.
- Free Space Model
- Two-Ray Ground Model
- Log-Normal Shadowing Model



- These PL models help us to estimate the distance based on the PL which is the difference between TSSI and RSSI [21].

Third Step

- Theoretical Models for PL

$$Free\ Space\ Model:\ \ PL = 20\log(d) + 20\log(f) - 27.55 \tag{2}$$

$$Two - Ray\ Ground\ Model:\ \ PL = 40\log(d) - 20\log(h_t) - 20\log(h_r) \tag{3}$$

$$Log - Normal\ Shadowing\ Model:\ \ PL = PL(d_0) + 10n\log(\frac{d}{d_0}) + X_\sigma \tag{4}$$

- Empirical Models for PL

$$Weissberger:\ \ PL = \begin{cases} 1.33 \times f^{0.284} d^{0.588} & 14m < d \leq 400m \\ 0.45 \times f^{0.284} d & 0m \leq d < 14m \end{cases} \tag{5}$$

$$ITU - R:\ \ PL = 0.2 \times f^{0.3} d^{0.6} \tag{6}$$

Third Step

$$COST: \ PL = 15.6 \times f^{-0.009} d^{0.26} \tag{7}$$

$$Proposed: \ PL = X f^{Y} d^{Z} \tag{8}$$

- The empirical method in equation (8) is employed to estimate the distances [21]. More precisely,

$$d = \left(\frac{PL}{X \times f^{Y}} \right)^{1/Z} \tag{9}$$

- Now, we have distances though it maybe inaccurate for long distances which multi-path fading affects them. This inaccuracy could be overcome.

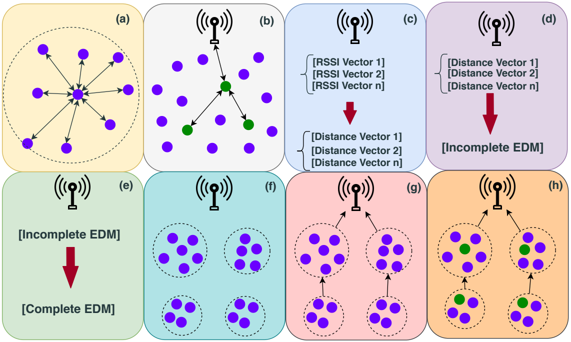
Third Step

$$
\begin{cases}
d_1 = [\cdot]_{(N+1)\times 1} & \text{thing number } 1 \\
\vdots & \quad\vdots \\
d_i = [\cdot]_{(N+1)\times 1} & \text{thing number } i \\
\vdots & \quad\vdots \\
d_{(N+1)} = [\cdot]_{(N+1)\times 1} & \text{thing number } (N+1)
\end{cases}
\tag{10}
$$

- Now, the BS knows the distance of each thing to its neighbors
- The above vectors are sparse because each thing can receive packet from its neighbors.
- But it needs the distance of each thing to all of the things.

Fourth Step

- Time to obtain the Euclidean distance matrix.
- A semi-complete matrix.
- A lot of unknown entries!
- Unfortunately, known entries are also noisy.



$$\hat{D} = [d_1, d_2, \cdots, d_{(N+1)}]_{(N+1)\times(N+1)} \tag{11}$$

Fifth Step

- GOOD NEWS!
- "Matrix completion with noise" can reconstruct our matrix [22].
- It also can denoises our matrix.
- But how?
- By solving the following SDP optimization problem:

$$\begin{aligned} \textit{minimize} \quad & \mathit{rank}(X) \\ \textit{subject to} \quad & \|X - Y\|_F \leq \delta \end{aligned} \tag{12}$$

- But, this is NP-hard, instead, it is possible to solve the relaxed form of the SDP problem:

Fifth Step

$$\begin{aligned} \textit{minimize} \quad & ||X||_* \\ \textit{subject to} \quad & ||X - Y||_F \leq \delta \end{aligned} \tag{13}$$

- Which $||X||_*$ the is nuclear norm of X, and can be calculated as follows:

$$||X||_* := \sigma_1 + \sigma_2 + \sigma_3 + \cdots + \sigma_i + \cdots + \sigma_N \tag{14}$$

- We're done. BS can this way reconstruct the semi-complete matrix.
- But hold on!! It's not that easy.

Fifth Step

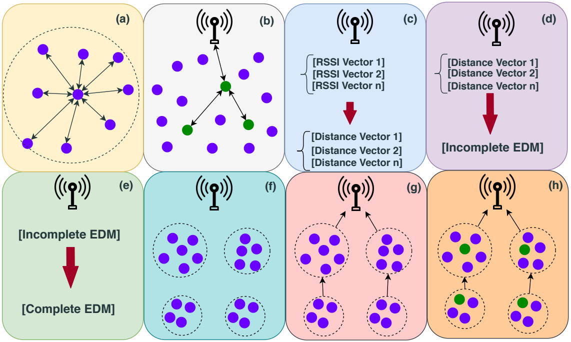
- Still in the 5th step ...
- Conditions need to be hold to able to reconstruct accurately and even exactly:
  - Exact case: The semi-complete matrix should have $N \times R \times (log\ N)^2$ entries coherently distributed in all of the rows (N: dimension of matrix, R: rank of matrix). Take $N = 1000$ and $R = 4$ (2D EDM),then having at least 36000 entries is enough to reconstruct other 964000 entries [22]. A miracle!!
  - Accurate case ($\pm 3\%$ error reported): Having $N \times R$ entries coherently distributed is enough. Therefore, assuming again $N = 1000$ and $R = 4$, having at least 4000 entries enough to reconstruct other 996000 entries [22].

$$Complexity\ of\ Reconstruction:\ O(3 \times N \times R^2 + R^3) \quad (15)$$

- Now the BS knows the peer-to-peer distances.

Sixth Step

- Clustering takes place.
- K-medoids handles outliers much more resiliently than k-means.
- Complexity [23]:
- J_{opt} clusters [24].
- They are isolated.
- Let's connect them.

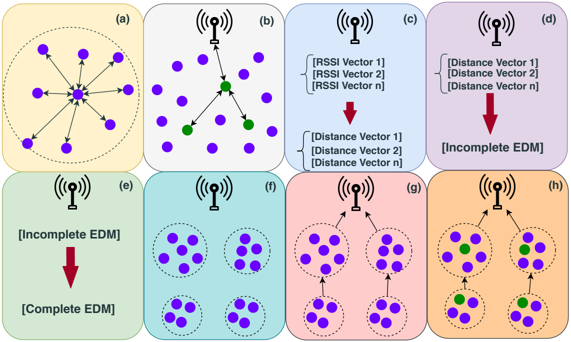


$$PAM : O(J(N-J)^2) \tag{16}$$

$$J_{opt} = \sqrt{\frac{N\varepsilon_{fs}}{2\pi\varepsilon_{mp}}\frac{L}{d_{toBS}^2}} \tag{17}$$

Seventh Step

- Minimum spanning tree could connect the clusters [24].
- Allows far clusters to reach the BS.
- In the next step, the BS has to select a CH for each cluster.



- An ILP optimization problem is formulated to select.
- Lowest energy-consumption in a round.

Eighth Step

$$\textit{minimize} \quad R(\ell)\sum_{i=1}^{C_j} x_i(\Lambda_{ij}+\Delta_{ij}) \tag{18}$$

$$\textit{subject to} \quad \sum_{i=1}^{C_j} x_i\Big(E_i(\ell)+H_i(\ell)-R(\ell)\Delta_{ij}\Big) \geq 0 \tag{19}$$

$$\sum_{i=1}^{C_j} x_i = 1, \quad x_i \in \{0,1\} \tag{20}$$

- After all of these steps, BS sends a packet to all things, including:
  - Euclidean distance matrix.
  - Cluster membership.
  - The minimum spanning tree.
  - CHs.

Simulations

- Comparison criteria:
  - Energy efficiency [2]

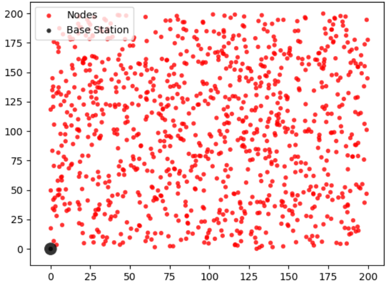
$$\eta = \lim_{L\to\infty} \frac{\sum_{\ell=1}^{L}\sum_{i=1}^{N} P_i(\ell)}{\sum_{\ell=1}^{L}\sum_{i=1}^{N} H_i(\ell)} \quad (Packet/Joule) \tag{21}$$

  - Packet loss [2]

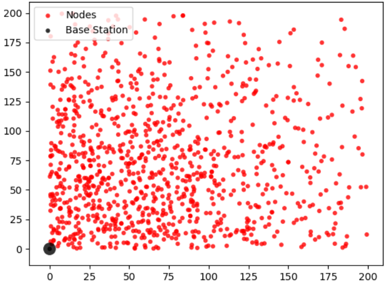
$$\zeta = \lim_{L\to\infty} \frac{\sum_{\ell=1}^{L}\sum_{i=1}^{N} P_i(\ell)}{\sum_{\ell=1}^{L}\sum_{i=1}^{N} A_i(\ell)} \tag{22}$$

  - Lifetime [3]
    - The passed time until one thing dies.
- Opponents
  - DEARER [2]
    - The same assumptions: No GPS
  - K-means [3]
    - An additional assumption: Things equipped with GPS

Scenarios

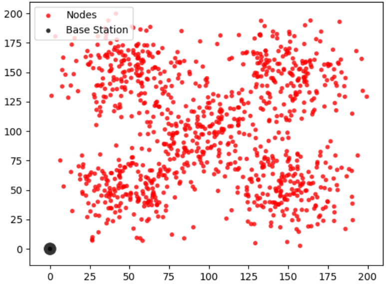


Uniform distribution



Gaussian distribution

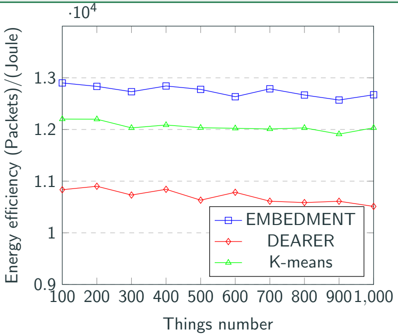
$N(\mu = 0, \sigma = 100)$

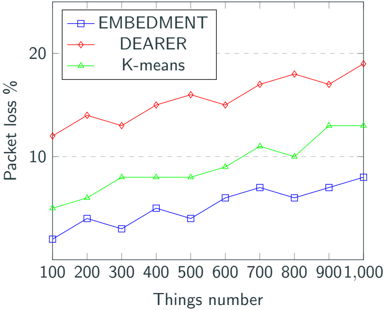


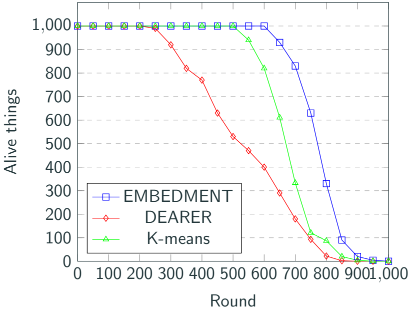
Gaussian mixture

$N(\mu, \sigma = 30)$

First Scenario

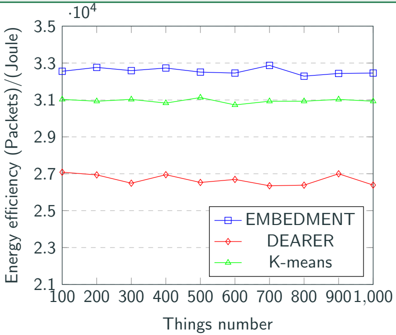


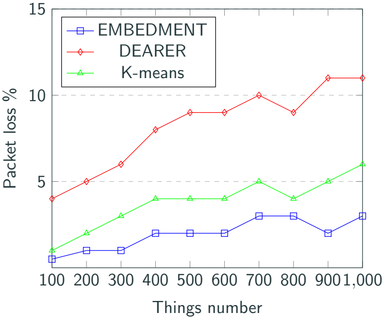


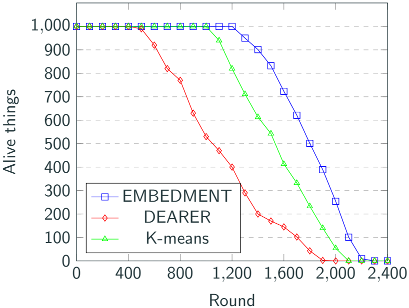


- More energy efficiency, lower packet loss, longer lifetime.
- Using a robust clustering method.
- An optimization problem leading to the optimal CH selection.
- The objective is the lowest energy consumption of a cluster.

Second Scenario

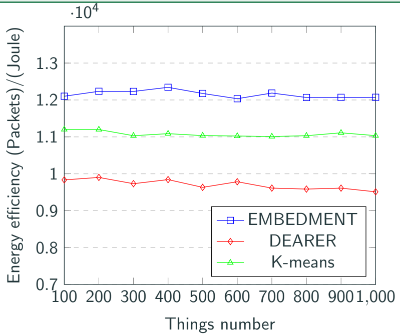


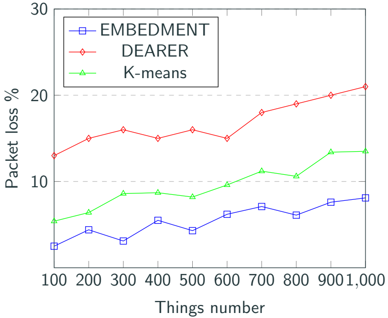


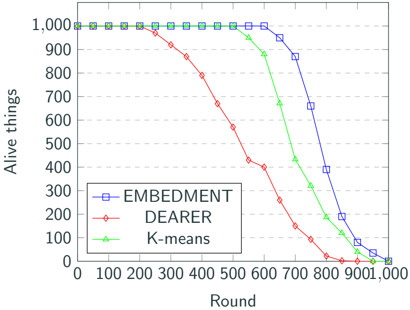


- All factors have been improved 2.5× in Gaussian distribution compared to uniform distribution.
- Why?
- Average distance of things to the BS reduced app. twice.
- Things have to send to closer BS.

Third Scenario







- The results for the MG dis. are similar to uniform distribution.
- The same average distance of things to the BS.
- Average distance is an important factor.
- It plays an important role in energy consumption.

Conclusion

- We considered sensor data path selection problem with respect to energy consumption in the sensor layer of the IoT.
- Current approaches fail to support:
    - Intercluster communication, avoiding GPS
    - Residual energy, energy harvesting
    - Coverage, classic clustering
    - Distributed
- Proposed approach support all thanks to two primary contribution:
    - Euclidean matrix reconstruction
    - An ILP optimization problem for CH selection
- Simulations corroborate that proposed protocol is superior considering:
    - Energy efficiency, packet loss, lifetime

Questions?

References

[1] Y. Dong, J. Wang, B. Shim, and D. I. Kim, Dearer: A distance-and-energy-aware routing with energy reservation for energy harvesting wireless sensor networks, IEEE Journal on Selected Areas in Communications, vol. 34, no. 12, pp. 37983813, Dec 2016.

[2] D. Wu, J. He, H. Wang, C. Wang, and R. Wang, A hierarchical packet forwarding mechanism for energy harvesting wireless sensor networks, IEEE Communications Magazine, vol. 53, no. 8, pp. 9298, August 2015.

[3] W. R. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in: System sciences, 2000. Proceedings of the 33rd annual Hawaii international conference on, IEEE, 2000, pp. 10pp.

[4] J. RejinaParvin and C. Vasanthanayaki, Particle swarm optimization-based clustering by preventing residual nodes in wireless sensor networks, IEEE Sensors Journal, vol. 15, no. 8, pp. 42644274, Aug 2015.

[5] A. R. Jadhav and T. Shankar, Whale optimization based energy-efficient cluster head selection algorithm for wireless sensor networks, CoRR, vol. abs/1711.09389, 2017. [Online]. Available: http://arxiv.org/abs/1711.09389

[6] J. Wang, K. Wang, J. Niu, W. Liu, A k-medoids based clustering algorithm for wireless sensor networks, in: 2018 International Workshop on Advanced Image Technology (IWAIT), 2018, pp. 14.

[7] S. Periyasamy, S. Khara, S. Thangavelu, Balanced cluster head selection based on modified k-means in a distributed wireless sensor network 2016 (2016) 111.

[8] S. K. Singh, P. Kumar, J. P. Singh, A survey on successors of leach protocol, IEEE Access 5 (2017) 42984328.

[9] Y. Miao, H. Wu, L. Zhang, The accurate location estimation of sensor node using received signal strength measurements in large-scale farmland, Journal of Sensors 2018.

[10] A. Tasissa and R. Lai, Exact reconstruction of euclidean distance geometry problem using low-rank matrix completion, CoRR, vol. abs/1804.04310, 2018.

References

[11] Wang, J., Wang, K., Niu, J., Liu, W. (2018). A k-medoids based clustering algorithm for wireless sensor networks. In: 2018 International Workshop on Advanced Image Technology (IWAIT), vol. 2018, pp. 14

[12] Periyasamy, S., Khara, S., Thangavelu, S. (2016). Balanced cluster head selection based on modified k-means in a distributed wireless sensor network. In: International Journal of Distributed Sensor Networks, vol. 2016, pp. 111

[13] Wang, Q., Guo, S., Hu, J., Yang, Y. (2018). Spectral partitioning and fuzzy c-means based clustering algorithm for big data wireless sensor networks. EURASIP Journal on Wireless Communications and Networking 2018(1), 54 65

[14] Miao, Y., Wu, H., Zhang, L. (2018). The accurate location estimation of sensor node using received signal strength measurements in large-scale farmland. Journal of Sensors 2018, 1 10

[15] Rasool, I., Salman, N., Kemp, A.H. (2012). Rssi-based positioning in unknown path-loss model for wsn. In: Sensor Signal Processing for Defence (SSPD 2012), pp. 15

[16] Yoon, Y.J., Park, W.T., Li, K.H.H., Ng, Y.Q., Song, Y. (2013). A study of piezoelectric harvesters for low-level vibrations in wireless sensor networks. International Journal of Precision Engineering and Manufacturing 14(7), 12571262

[17] Yang, H., Zhang, Y. (2015). Estimation of supercapacitor energy using a linear capacitance for applications in wireless sensor networks. Journal of Power Sources 275, 498 505

[18] Yang, H., Zhang, Y. (2011). Self-discharge analysis and characterization of superca- pacitors for environmentally powered wireless sensor network applications. Journal of Power Sources 196(20), 8866 8873

[19] Quaranta, G., Trentadue, F., Maruccio, C., Marano, G.C. (2018). Analysis of piezoelec- tric energy harvester under modulated and filtered white gaussian noise. Mechanical Systems and Signal Processing 104, 134 144

[20] A. Tasissa and R. Lai, Exact reconstruction of euclidean distance geometry problem using low-rank matrix completion, CoRR, vol. abs/1804.04310, 2018.

References

[21] Wang, J., Wang, K., Niu, J., Liu, W. (2018). A k-medoids based clustering algorithm for wireless sensor networks. In: 2018 International Workshop on Advanced Image Technology (IWAIT), vol. 2018, pp. 14

[22] Periyasamy, S., Khara, S., Thangavelu, S. (2016). Balanced cluster head selection based on modified k-means in a distributed wireless sensor network. In: International Journal of Distributed Sensor Networks, vol. 2016, pp. 111

[23] Wang, Q., Guo, S., Hu, J., Yang, Y. (2018). Spectral partitioning and fuzzy c-means based clustering algorithm for big data wireless sensor networks. EURASIP Journal on Wireless Communications and Networking 2018(1), 54 65

[24] Miao, Y., Wu, H., Zhang, L. (2018). The accurate location estimation of sensor node using received signal strength measurements in large-scale farmland. Journal of Sensors 2018, 1 10