Article

An Efficient Algorithm for Retrieving the Updated Copy of Data Sited on Least Hops in Wireless Sensor Network

Kardan Journal of Engineering and Technology
1 (1) 23–36

Kardan Publications
Kabul, Afghanistan
DOI: 10.31841/KJET.2021.3
https://kardan.edu.af/Research/Kardan_journal_of_engineering_and_technology.aspx
#

Muhammad Haleem
Amjad Khan
Farrukh Sair

Abstract

Sensors as data producers, record and forward the observations to base stations for further processing within the resource constraints. Over a time period of almost decade, lot of work has been carried out on retrieval of data from wireless sensor network with different aspects. In this paper a technique is provided for retrieving Updated Copy of info (Data) situated on Least Hops (UCDSLH) in wireless sensor network. The technique works well for querying the global maxima (UCDSLH) as well from the specific data centric nodes. This retrieval is at searching cost of order one O (1). Based on time complexity and energy consumption factors, for retrieval of global maxima as compared to the existing techniques, our technique is significantly efficient. As it retrieves the interested UCDSLH without flooding in existence of metadata nodes. Parameter Refresh Protocol (PRP) populates the Metadata nodes and notifies the base station about priorities of the metadata nodes. Metadata nodes have data about the sensed data stored on the Data Centric Storage (DCS) nodes. DCS – a logical collection of multiple sensors; sited on the least hops having the UCDSLH get the high priority. Base station links to the nearest available Metadata node based on signal strength. This metadata node act as intermediate channel as it points out the DCS to retrieve the UCDSLH to satisfy the query. DCSs and metadata nodes are replicated to ensure robustness. Our algorithm shows complete procedure of our technique. We simulate and evaluate our approach with necessary query analyzer implemented in C++.

Keywords: *Information & Communication Technology. UCDSLH, PRP, DCS metadata, C++, Sensors, Base Station, Wireless Sensor Network*

Mr. Muhammad Haleem, Mr. Amjad Khan and Mr. Farrukh Sair *are Assistant Professors at Faculty of Engineering and Technology, Kardan University - Kabul, Afghanistan*

Introduction

Motes Network or wireless sensors networks (WSN) consist self-manageable and self-configurable sensor nodes. "These Sensor nodes are mostly referred as motes or mote computers. A mote has the ability to perform some processing, aggregation of information and communication with the rest of nodes in the WSN for provision of data in the domain they are deployed for" [1]

Sensor motes can deploy large amount of number in some remote areas that form a wireless ad-hoc networks that can supportively televise somatic or environmental circumstances such as sound, temperature, pressure, vibration, motion and pollution at various places over the area under observation [6]. WSNs have the capacity to succor in a variety of applications [2]. WSNs are suitable for military and civil applications object tracking infrastructure monitoring [5], territory sensing, battlefield reconnaissance, health nursing and natural misadventure assistance [3].

Sensor nodes transmit the gathered multi-dimensional observations of the environment to a central station known as base station - where the observed data can be analyzed to extract some decision support information.

Research has been done over the decades to ensure that the sensor network can perform better within its resource constraints [7]. Optimally synchronized non overlapping slot allocation and routing algorithms are proposed for contribution to the issues in the field of WSN [2].

This paper concentrates on extraction of updated data copy situated on the shortest path, from Base Station. To the best of our knowledge, this research work carried out regarding efficient usage of resources and the way of query processing in WSN include: exterior storage [9], local based storage, Central storage of data [10], multiple data copies of DCS [11] and Cooperative Signal and Data Processing [13], "Data Storage and Retrieval with protocol for minimum power networks [17].

2 Background and Related Work

In WSN data storage and retrieval are tightly coupled to each other. The sensed data is stored somewhere on storage in the domain of WSNs for the purpose of latter use, while the data to be retrieved is necessary to be stored first. The literature review in this area returns some related work that set the stage for our technique is summarized here as:

A. Extraction of Info from External Data Store

In this storage mechanism the internal nodes send the sensed info to the storage located outside of the sensor network [15]. There is no cost for

user queries as the information are available exterior. But the cost of transferring the resultant dataset to the base station on each time sense is of huge cost, thus it proved inefficient [9].

B. *Extracting Data from Native (Local) Data Store*

Here the detected event information is stored at local data stores at the detecting node. The data storage process initiates upon detection of an event which sustains no communication cost. In this technique all the queries are swamped to the nodes in its flooding domain at a cost of $O(n)$.

C. *Extracting Data from Central Data Store or Data Centric Storage*

Tracking the interested patterns by use of flooding is proved incompetent due to more energy consumption [10]. DCS, a shared reference system is pointed for storing and retrieval of data. Using this approach, a shared referencing technique has been presented. Data items are tied with a named key and located on a particular spot. For a given piece of data, hashing techniques produce the hash code for the data name; this hash code refers to the particular single DCS node. Queries as well as data items are both sent to this common referencing point. This approach reduces number of transmissions which cause flooding. But no replication of DCS nodes leads to the problem of robustness. The robustness issue ultimately costs the data lose in special cases of single point failure.

D. *Replication of DCS for Robustness*

Replication of the DCS nodes provides better data centric method for data storage in WSN with improved availability of essential info. It gives us the edge of robustness [11]. In this method, authors vouch for multiple resiliency levels based on sensitivity of nature of events.

The messaging cost to store the event is $O\sqrt{n}$. Here reliability of facts and figures is ensured by having multiple copies of data. The issue with these multiple data copies is that the query producing flooding for searching interested information in multiple DCSs, as the hash function is not applicable over here [7], hash function can return same value for multiple inputs, so it can't uniquely point out the specific DCS.

E. *Cooperative Signal and Data Processing*

Cooperative Signal and Data Processing presents a goal tracing technique [13]. This work completely relies on cluster head (CH). The detecting node triggers a data report to cluster head. Then the CH will choose the required subset of sensor nodes that can be queried for further queried data set. Base station points to CH for querying. But again the

robustness problem exists as this technique suffers from the single point of failure because there is only one cluster head.

F. *Data Storage and Retrieval with Routing Protocol.*

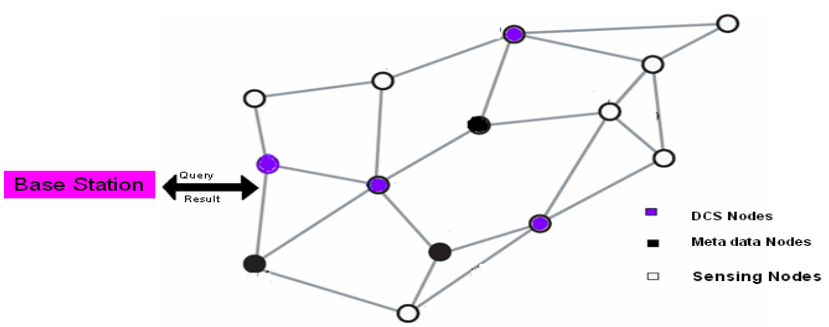
This technique is based on IPV6 and its main propos (purpose) is the retrieval of data with Protocol for Low power networks (RPL) [17]. The paper presents the impression of distributed storage in WSNs which Focuses on reducing the possibility of data loss in case of critical level of power by using the rank approach where "the nodes having the critical level of power would assign high rank to be more attention for the data they have".

The above listed related work in the area of data retrieval in WSN does not cover the issue of retrieving the global maxima with robust scenario and low energy using the least hops to the destination. Our technique contributes in retrieval of the specified global maxima UCDSLH, with low energy at $O(1)$ in presence of metadata nodes.

3 Proposed Data Retrieval Technique

In this technique the sensor nodes deployed in its wireless domain are grouped as *Data Sensors, Data Centric Stores and nodes for Metadata*. The nodes for sensing get the data items and store them to the replicated DCS nodes in its access domain. Sensors send captured data items to the accessible DCS(s). where data can be recorded to multiple DCS nodes.

Figure 1: Data Retrieval Technique



Source: Capture of Simulation Result

This replication strengthened the data robustness that overcome the problem of CSIP. DCS nodes are periodically called by parameter refresh protocol PRP [18] and they keep the metadata nodes updated. Our technique is composed of two parts; preparing the metadata and retrieval of UCDSLH. In metadata preparation phase, the information consists of *time stamp of data event and the hops taken to them from BS* are returned to the metadata node. The phase also covers assignment of prioritized alpha-

numeric logical names to the DCS nodes in ascending order, starting from nearest node based on their dispersion range from the BS. The address slot inside metadata node is populated in such a way that the first nearest DCS gets the 1st address space; the second gets the second and continues till no other exists. PRP triggers the searching procedure to find the DCS having greatest logical time stamp - the most updated, and marks this as the node of interest to all the metadata nodes. The metadata nodes prepare and send their own priority detail to BS to specify which of them is available at commendable time and energy cost for receiving the query.

In the second phase of our technique, the base station using the received priority list is reported by the PRP. This list act as guide for the base station to send query to the available highest priority holder metadata node. When BS queries for global maxima the UCDSLH, the query is sent to the interested DCS directly in existence of metadata information which collected in the first phase. This direct access is of *O (1)*. The other cases of retrieving the overall and specific nominated DCS's data are also covered. To response the query for data from all DCS nodes, if required, they are traversed using iteration starting from first logical name till the last in a sequence. For responding the query for only DCS is quite simple in our technique as the desired DCS is resolved to its logical name from the metadata node's address slot and returns the data to the base station.

4 Problem is Existing Approaches

In existence of multiple data copies with WSNs, the info in sensing domain needs to be navigated for probing the interested info. The existing solutions from current approaches to cover all the search space is flooding because the current technique use hashing, it can point out one node, the data centric storage [11]. Thus, this hashing technique is not handy to search for global maxima among multiple nodes of WSN. Flooding can't be applied because it proved to be inefficient in terms of WSN constraints as it consumes more energy, the active issue [17]

As a result of literature review cited in this paper, no solution available to access the UCDSLH – the Updated data copy situated on least hope as no metadata is used in current WSNs.

CSIP suffers from single point of failure, the CH-cluster head. Thus this approach has the problem of robustness the likely data loss [14].

Thus to extract the rationalized (updated) copy of data situated on minimum hops in the WSN is still an active challenge.

5 Proposed Algorithm

The literature review pointe that a solution is required to address the issue of extracting the UCDSLH in WSN.

This research organizes the sensor network motes by grouping the sensor node into three categories. The categorization is mainly based on behavior of sensor nodes. First category motes are committed to have data about other sensor, their data and base station. These nodes are known as metadata nodes.

This algorithm covers both aspects the query and event data collection. The query based approach returns the dataset in response of the user query. While in event base approach, the data retrieval is based on the event when it happens.

This algorithm groups the nodes as:

a. *Motes for Sensing data*
b. *Sensors for Metadata*
c. *Nodes for data centric storage*

A. Motes for Sensing Data

the motes are dedicated for sensing the data and storing infor to the relevant DCS after its identification. Then these nodes store the hashed information to DCS nodes using the key. Data stores multiple DCS nodes to avoid the single point of failure.

B. Nodes for Data Centric Storage

WSNs store the collected data after sensing operation. WSN works within resource limit, the low battery power, low memory, slow processor and insufficient bandwidth, due to its implementation. Periodically DCS nodes are accessed from base station via queries for their data. Unlike the traditional network, handling data storage in WSNs is quite difficult and challenging due to the resource constraints of wireless sensor networks.

The Storage motes, DCS, capture events with data names. The data stored in DCS is accessable by a unique name declare for data set, for examle temp for all temprature data. Actually, DCS is set of multiple data sensors, while logically it is treated as a one info file. To have enough storage with resouse constraint, multiple motes get collects to form a single DCS.

C. Sensor Nodes for Metadata

The technique introduced by this research work uses the metadata sensor nodes. These nodes are responsible for having the metadata nodes. In order to co-ordinate the sensor nodes with metadata nodes for sharing the data status and values. The parameter refresh protocol works well. The PRP [9] periodically activates and refresh the desired status.

The happened events time stamp and the count of hops for reaching to DCSs marked in metadata nodes are properly stored in metadata nodes. The

nodes having UCDSLH in the installed WSN will be marked as interested node in metadata. This protocol – PRP returns the hops count to base station as well. These hops count values then helps in direct retrieval of data from the WSN. The minimum count of hops to the DCS that have the UCDSLH gets the highest priority. Priority values are there for metadata nodes as well. These values are based on the hops count distance of metadata to DCS and base station. This is a cognitive approach where the minimum count is given the higher priority. The metadata priority values are sent back to base station (BS) as additional info to be used for query direction. This directed query approach insures the *O (1)* access to global maxima from base station.

As a result, the highest value of priority is given to the DCS which holds the UCDSLH. The priority list populated on base station helps in posing query over the WSN after viewing it. Thus query from BS will be forwarded to the priority DCS via metadata node to return the data set. As flooding is proved in-efficient. So our solution is designed with

consideration of accessing global maxima without letting it. This direct access takes constant time which is known as O(c), access at the order (O) of constant (c) time. In turn it can be written as O(1), based on asymptotic rules for analysis of Algorithms.

6 Algorithm of The Proposed Technique

Algorithm: Part 1 Preparing Metadata

1. [Traverse the installed WSN, Using PRP for accessing *the time stamp and the number of hops count to them from BS, and* the information is returned to the reserved Metadata nodes]

Traversing the DCS in WSN

2. [Assignment of prioritized Names to DCS in ascending order based on their dispersion range from BS]

Assign logical names to each DCS

3. [Construct the ordered address slot of DCS on Metadata nodes such that the nearest get the 1st address space, the second nearest get the 2nd and so on]

Mark addresses of DCS in Metadata nodes

4. [Search by starting from the first DCS nodes and mark the node in meta data node for extracting the updated copy of data stored on least hops count from base station]

Search until you find the copy of data with greatest logical time stamp.

5. [Metadata nodes prepare and PRP report their detail to BS that which of them is 1st,2nd … nearest to base station for responding the query]

Prepare and send Metadata priority detail to BS

6. Exit

Algorithm: Part 2 Retrieval of data

1. [Available highest priority holder Metadata node receives the query]

Receive the Query from Base Station

2. [Case1: Response to query for all data from all the DCS]

If (Query for retrieval of all the data of all DCS)

i. For i =1 to N using Metadata [Retrieve data] If(DCS(i) exists) Retrieve the Data of DCS(i)

Else

Skip and prompt the absence

ii. Go to step 5

3. [To response the query for all data of specific node may be DCS or Metadata]

If (Query for retrieval of data from a single node)

[Retrieve data]

If (node exists)

i. Retrieve the Data

ii. Go to step 5

Else

Prompt the absence

4. [Response to Query for UCDSLH]

i. Retrieve data from DCS the pointed DCS

ii. Go to Step 5.

5. Exit

7 Implementation

The technique is implemented in C++. The simulation provides best illusion of the overall scenario to evaluate the approach with necessary query analyzer. The file handling features of C++ has been used to simulate the storages at DCSs. The random routines are included from header file <math.h> with necessary configurations. The system clock is provided as seed value to generate the random data sample assumed as temperature

from sensor nodes. A light query analyzer is developed for parsing the query according to its clauses and criteria.

Our query analyzer supports the exact and fuzzy match queries. A query example for fuzzy match is *SELECT UCDSLH FROM DCS%.*

As shown in the given figure 1. This query results are the desired outputs that returned from the DCS (0). The recorded value for temperature is 85. It is assumed to capture the fire event observations. As sensors can be applied for such type scenarios, where environments are observed in unattended situations.

In the given figure 1 the UCDSLH retrieved successfully. In comparison section, this extraction is marked as access of *O(1)*, because it is a direct access of UCDSLH in existence of metadata.

Figure 2: Extraction

Source: Capture of Simulation Result

The query analyzer works properly to simulate the scenario for our deployed WSN. The deployed sensor nodes detect and copy the data value to various DCS nodes in their access scopes along with the current time stamp. Time stamp is a given time unit used to assist in identification process of updated copy of the data. These multiple sets of data and time stamp ensure the robustness. The Figure 2 shows that data item 131 copied 6 times to DCS11, DCS5, DCS6, DCS2, DCS3 and DCS4 on the time stamp 8. It is possible for a data item to be copied in duplication for the same time stamp to the same DCS as our assumption that synchronization issues are there in WSN and are not concerned to this research. The scenario test screen shot,

Figure 3: Scenario Test Screen Shot

```
*************************************
Time stamp: 16
Data item (93) copied 10 time(s)
*************************************

Data        DCS
_____________________

93          11
93          5
93          0
93          7
93          5
93          9
93          10
93          7
93          8
93          3
```

Source: *Capture of Simulation Result*

Figure 3, shows that data item 93 copied 10 times for time stamp 16 but actually it is stored to eight DCSs.

The technique is evaluated with simulation by assuming 100 sensor nodes, 12 duplicated DCSs are given having logical names DCS (0), DCS (1), DCS (2) …. DCS (11) based on their distance from base station for time stamps 1 to 20. The deployed metadata nodes, duplicated three times, are named metadata1, metadata2, metadata. The technique allows the variable number of sensor nodes, dynamic times of duplication level for the DCS and metadata nodes based on criticality and sensitivity of information and deployment scenarios.

The query analyzer, located on base station, parse the query clauses accurately. The simulation shots given bellow show the syntax of queries supported by our technique along with the responded data sets. This query returns the global maxima, the updated copy of data sited on least hops, UCDSLH has shown

The query *SELECT * FROM DCS (0)* returns DCS (0)'s contents as shown here in the given screen shot Figure 4 of the simulation result.

Figure 4: Result of Query

```
        Result of Query:  SELECT UCDSLH FROM DCS%

Retrieval of UCDSLH
_____________________________

Retrived from DCS No = 0
Stamped time= 20
Last temprature = 85
Average temprature =97
_____________________________
```

Source: Capture of Simulation Result

Figure 5: Result of Query

```
Result of Query:  SELECT * FROM DCS(0)

Retrieval of data from:DCS(0)

Dat,Stmp
111, 5
104, 7
110, 13
121, 14
93, 16
84, 17
84, 17
81, 19
85, 20
```

Source: *Capture of Simulation Result*

This technique allows us to query over a single specified DCS. Also, base station user can run a query over all the DCSs for retrieving all the data. For all data retrieval the query to be executed from base station is *SELECT * FROM DCS%.* The given figure 5 and figure 6 shows this case of our simulation results.

Figure 6: Result of Query

```
Result of Query:  SELECT * FROM DCS(0)

Retrieval of data from:DCS(0)

Dat,Stmp
111, 5
104, 7
110, 13
121, 14
93, 16
84, 17
84, 17
81, 19
85, 20
```

Source: Capture of Simulation Result

Figure 7: Retrieving Data

```
Retrieving data from DCS(1)
Dat,Stmp
105, 4
111, 5
118, 6
118, 6
104, 7
92, 12
128, 15
123, 18
81, 19
85, 20
```

Source: Capture of Simulation Result

Similarly, this retrieval is continued and will return the data sets from DCS (0) - DCS (11).

8 Comparison Table

The given comparison shows the results in term of energy consumption and data retrieval complexity of data retrieval with different storage

schemes of WSN. WSNs works within resource constraints. Here cost means Energy consummation. For *n* - the number of installed WSN nodes, $O(n)$ is the cost for floods [14] while the cost of $O(n^{1/2})$ for point-to-point routing. [8].

Table 1: Comparison Table Data Storage in WSN

	External	Local	Data Centric	Multiple DCS	Multiple DCS with Metadata Nodes
Cost of Storage	$O(n^{1/2})$	*o*	$O(n^{1/2})$ *Not robust*	$O(n^{1/2})$ *Robust*	$O(n^{1/2})$ *Robust*
Cost of Query	*o*	$O(n)$	$O(n^{1/2})$	$O(n)$ *No P2P*	$O(c)=O(1)$ *Direct Access*
Cost of Response	*o*	$O(n^{1/2})$	$O(n^{1/2})$	$O(n^{1/2})$	$O(c)=O(1)$

Source: *Asymptotic complexity*

The above table shows the cost in term of energy consumption by given three criteria. These criteria are:

a. Cost of storing the data to given *data stores*.
b. Cost of query for stored data.
c. Cost of response against query to base station.

Our technique focus on efficiency in term of energy consumption and time complexity in data extraction, and it shows significant difference in term of energy consumption and time complexity for querying over global maxima in WSN.

9 Future Work

To know that which sensor has generated the updated copy of data.

10 Conclusion

WSN has variety of application for overviewing the unattended environment. It is valuable info to get the latest updates of the sensing field in observation and maintenance process.

The information extraction technique for global maxima presented by this paper improve the information retrieval from regarding complexity from $O(n^{1/2})$ *to* $O(1)$. This technique focus on satisfying the query without flooding. This literature review criticizes the flooding approach as it is incompetent in term of the essential resource in WSN, the energy. More to the point, this research issues the never addressed issue of retrieving UCDSLH. The fresh and valuable fresh data from the WSN. Here the robustness has been considered at metadata as well as DCS level. PRP has been applied for co-ordinating the data state. This feature helps in guiding the query to already know DCS for direct retrieval of data.

References

[1] Ahmed, Khandakar., et al. Wireless *"Sensor Network Data Centric Storage routing using Castalia"* IEEE 9 Nov. 2012,Telecommunication Networks and Applications Conference (ATNAC)

[2] Castillo-Effen, M., et al., "*Wireless sensor networks for flash-flood alerting*" Proceedings of the Fifth IEEE International Caracas Conference on Devices, Circuits, and Systems, Dominican Republic, 2004: p. 142-146.

[3] Akyildiz, I.F., et al., "*Wireless Sensor Networks: A Survey*" Computer Networks, 2002. 38(4): p. 393-422.

[4] Lorincz, K., et al., "*Sensor networks for emergency response: challenges and opportunities.* Pervasive Computing for First Response" IEEE Pervasive Computing, 2004.

[5] Gao, T., et al., "*Vital signs monitoring and patient tracking over a wireless network*" Proceedings of the 27th IEEE Engineering in Medicine and Biology Society, Annual International Conference, 2005: pp. 102-105.

[6] Romer, Kay; Friedemann Mattern. "*The Design Space of Wireless Sensor Networks*". IEEE Wireless Communications (December 2004) 11 (6): 5461. doi:10.1109/MWC. 2004.1368897.

[7] Choi, J., D. Shin, and D. Shin, "*Research and Implementation of the Context-Aware Middleware for Controlling Home Appliances*" IEEE Transactions on Consumer Electronics, 2005. 51(1): pp. 301-306.

[8] Sylvia Ratnasamy., et al., "*Data-Centric Storage in Sensornets*" WSNA - Atlanta, Georgia, USA. 2002. ACM 1581135890/02/0009

[9] Sylvia Ratnasamy, Brad Karp et al. "*GHT: A Geographic Hash Table for Data Centric Storage*" Proceedings of the Sixth IEEE International Caracas Conference on Devices, Circuits, and Systems, Dominican Republic, 2005: p. 158-162.

[10] Md. Zakirul Alam Bhuiyan, Guojun Wang and Jie Wu "*Target Tracking with Monitor and Backup Sensors in Wireless Sensor Networks*" IEEE Wireless Communications and Networking Conference (WCNC 2004), pp. 1377–1383, 2004.

[11] Ravinder Tamishetty, Lek Heng Ngoh "*An efficient resiliency scheme for data centric storage in wireless sensor networks" In Proc. of the IEEE 60th Vehicular Technology. Conf. 2006, vol. 4, pp 53-58*

[12] *Omotayo, A. Hammad, M.A. Barker, K.* A Cost "Model for Storing and Retrieving data in Wireless Sensor Networks" *Data Engineering Workshop, December 2007 IEEE 23rd International Conference. pp 29-38.*

[13] *ZhijunYu, Jianming Wei, and Haitao Liu.* "An Energy Efficient Target Tracking Framework in Wireless Sensor Networks" *Article ID 524145, EURASIP Journal on Advances in Signal Processing, May 2009.*

[14] *Ratnasamy, Sylvia & Karp., et al., "Data-Centric Storage in Sensornets with GHT, a Geographic Hash Table" Mobile Networks and Applications (2003). 8. 10.1023/A:1024591915518.*

[15] Mallick, Chandrakant & Satpathy, Suneeta. *"Challenges and Design Goals of Wireless Sensor Networks: A State-of-the-art Review"* International Journal of Computer Applications (2018). 179.42- 47. 10.5120.

[16] Pandithurai, O & Suresh Kumar, C. CPEGASIS: "*Efficient data retrieval in wireless sensor networks using cloud computing environment*" (2016). doi:15.3725-3729. 10.3923/ajit.2016.3725.3729.

[17] B. Prathiba, K. J. Sankar and V. Sumalatha, "*Enhancing the data quality in wireless sensor networks — A review*," International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, 2016, pp.48-454. doi: 10.1109/ICACDOT.2016.7877626

[18] J. He, G. Sun, Y. Zhang and Z. Wang, "*Data Recovery in Wireless Sensor Networks with Joint Matrix Completion and Sparsity Constraints,*" in IEEE Communications. 2015. vol. 19, no. 12, pp.2230-2233, doi:10.1109/LCOMM.2015.2489212