This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.3006276, IEEE Transactions on Industrial Informatics

An Efficient Clustering Framework for Massive Sensor Networking in Industrial IoT

Shiva Raj Pokhrel, *Member, IEEE*, Sandeep Verma, Sahil Garg, Ajay K. Sharma and Jinho Choi, *Senior Member, IEEE*

***Abstract*—Massive Machine Type IoT Communication (mMTIC) has the potential for high impact in anticipated future industry 4.0 sensor networking applications. However, the energy limitation and battery life of the IoT nodes have always been one of the long-standing problems. Clustering Routing Protocol (CRP) being the most efficient existing approach often suffers when nodes closer to the sink depletes their energy, thereby producing an unwanted energy hole, where packets in flight towards the sink often get interrupted. Considering mMTIC covering a large geographical area, such as monitoring bush fires, the multi-hop communication among the nodes often causes such an energy hole problem. In this paper, we develop an AI-based CRP (CIRP) framework for incorporating a small periphery of a fixed shaped area to ameliorate such energy holes. Our proposed framework is not only energy-optimized but also acts as a robust approach for massive communication and informed data collection.**

***Index Terms*—Industrial IoT, Industrial sensor networks, cluster-based routing; energy hole problem; Clustering.**

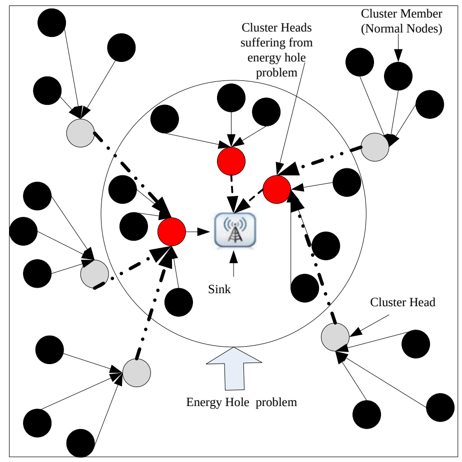


Figure 1: An example of energy hole in mMTIC setting.

I. Introduction

Trillions of Industrial Internet of Things (IIoT), devices and industrial machines will be connected by the year 2025, which will eventually lead to massive machine-type-IoT communication (mMTIC), a well-known use-case for the anticipated sixth generation (6G) and beyond wireless technology. In mMTIC, sensor-based industrial IoT networks have the potential to support modern real-time industry applications such as surveillance and monitoring of critical areas and autonomous mobile robots [1].

In this paper, we consider mMTIC for a typical industry 4.0 sensor networking scenario, where a large amount of static sensor is deployed and report sporadically to an application server in the cloud (e.g., to enable environment monitoring and object condition tracking). Over the past few years, due to the advances in AI (Artificial Intelligence) and wireless networks, mMTIC involving sensor networks have accounted progressive strides in different sectors of industrial applications [2]. One of the inescapable constraints of limited battery resources have engrossed the attention of researchers to solve the challenge and satisfy multifaceted requirements for industry 4.0 applications [3], [4], [5]. To resolve this concern comprehensively, a plethora of energy efficient cluster-based routing techniques has been existing in the literature [6], [7]. The clustering involves different grouping of the industry sensors into clusters. The data transmission from every cluster is via by the Head of Cluster (HoC). HoC of a cluster not only performs the tasks of data collection and aggregation but is also supposed to relay the data from the remotely located HoCs of the sink in a multi hop networking scenario [8], [9]. The process of data forwarding from HoC acts as a catalyst for the energy depletion [10]. Such an area occupied by the relaying HoC nodes often get overloaded with the task of data forwarding. Over the time, these nodes exhaust their energies and eventually no connection zone will be created. This phenomenon as shown in Fig. 1 is refer to as *energy hole* in the mMTIC setting [11] [12].

To resolve the concerns raised by such energy holes, several attempts have been reported in the literature [12] [13]. However, it worth noting that in all proposed solutions, the centralized placement of sink in the network is the primary approach. But, an alternative solution to the hole problem for the applications such as smart factory

S. R. Pokhrel and J. Choi are with the School of IT, Deakin University, Geelong, Australia. S. Verma is with Dr BR Ambedkar National Institute of Technology, India. S. Garg is with Ecole de technologie superieure, Electrical Engineering Department, Montreal, QC, Canada. A. K Sharma is with National Institute of Technology Jalandhar, NIT Campus, India Email: shiva.pokhrel@deakin.edu.au, sandeepv.ec.13@nitj.ac.in, sahil.garg@ieee.org, sharmaajayk@nitj.ac.in, jinho.choi@deakin.edu.au

where the sink placement is not feasible in the middle of the network remains poorly understood in literature. Furthermore, other related work proposing solutions in the direction of ameliorating energy holes, consider to avoid multi hop communication in such network, which is not practical.

Overall, our objective in this paper is to develop an efficient sensing framework so that communication and networking take place based on needs [14]. Our idea differs from existing approaches where first data sets are collected and stored, and then processed depending on needs, which will be very expensive for the required mMTIC [15]. We proposed an approach, which performs data transmission of only the essential data sets (from the end nodes based on the requirements), thus minimizing the network load substantially. Our main contributions in this paper are outlined as follows.

1. We design a novel Clustering with AI-based Routing Protocol (CIRP) that enhances industrial sensor network performance substantially with improved stability period, throughput, network lifetime, etc.
To mitigate the energy hole problem, in CIRP, Cognitive Multiple Gateway Nodes (CMGNs) are placed outside the network vis-à-vis single sink used conventionally in the existing protocols. With the adoption of AI and cognition in multiple gateways, the distance of the IoT nodes from their respective sink can be significantly minimized, which optimizes the energy consumption extensively.

2. The proposed CIRP is evaluated using two orthogonal industry 4.0 use cases; viz., Case A considers smaller network area (100 x 100 square meter deployed with 100 nodes), whereas in Case B exploits larger area (500 x 500 square meter consisting of 300 nodes). Both Case A and Case B has been examined thoroughly.

II. Background Overview

The industrial sensors in mMTIC need to play a vital role as they render a communicating link between physical world and industrial control systems[16]. The dynamic characteristics of industrial sensors that include self-configuring, self-healing and self-organizing make them highly anticipated for harsh environments common in industries [17]. Furthermore, to acquire an efficient massive data transmission anticipated by the industrial sensors, the mechanisms of data compression and sleep scheduling have already been proposed [18], [19]. However, they suffered from the pitfalls of energy hole and inefficient data dissemination in the network.

Of particular relevance to this work is the cluster-based routing for the HoC selection as explained in [6]. While dealing with static network, Akila *et al.* [20] proposed Zone based Clustering algorithm (ZCA) that divides the whole network into different geographical areas. The selection of HoC in ZCA is done only on the basis of remaining energy. The sole consideration of energy factor and neglecting the distance and node proximity factors makes it non promising to acquire optimal network performance. Mohemed *et al.* [21] have implemented the proposed scheme on the paradigm of few nodes only, however it is contemplated that in large area network, the protocol OHCR will incur a high magnitude of delay due to chain based clustering adopted in intra-cluster communication. Mittal *et al.* in [22] proposed TEDRP protocol that efficiently utilized the network for energy efficient routing. However, due to dual hop communication, the energy hole problem still sustained in the network.

Based on the aforementioned studies, the energy hole problem has to be combated in a way that the amount of data transmission should be substantially minimized. Further, if the HoC selection can be improved with the consideration of energy efficient parameters, the overall performance of mMTIC pertaining to hostile applications can be upgraded significantly. In the context of above discussions, with the adoption of AI and cognition in multiple gateways, the distance of the IoT nodes from their respective sink can be significantly minimized. Consequently, energy consumption of the nodes can be decreased considerably.

III. Proposed AI-based Framework for CIRP

In this section, we adopt the fundamental radio energy model known in wireless communication principles and discuss the operation of CIRP as follows. The acronyms and parameters used in this work are provided in Table 1.

A. Working operation of CIRP

The set up phase of CIRP includes the network formation and proposed HoC selection stages, which are discussed as follows.

Total energy, E_T, of the mMTIC network can be computed along the lines of [11]. With relevant insights from [11], other parameters considered for the HoC selection are discussed below.

Energy factor: The energy factor includes the ratio of residual/remaining energy (E_{res}) to the average energy i.e., E_{avg} (computed by following equation (1)) of nodes, is given by factor i.e., $\frac{E_{res}}{E_{avg}}$ for every single candidate node in the network. The average energy of the node is given by:

$$E_{avg} = \frac{1}{n} * E_T \tag{1}$$

Distance factor: The inclusion of distance factor prompts the selection of those nodes as HoC which are located nearer to the sink. The distance computation of i^{th} (where i ranges from 1 to total number of cluster nodes i.e., N_C node from the nearest CMGNs) is determined

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.3006276, IEEE Transactions on Industrial Informatics

Table I: Acronyms and network parameters

Acronym	Full Form
CIRP	Clustering with AI-based Routing Protocol
MEEC	Multiple data sink-based Energy Efficient Cluster-based routing
ZCA	Zone Clustering Algorithm
OHCR	On-Hole Children Reconnection
TEDRP	Threshold-sensitive Energy-efficient Delay-aware Routing Protocol
Parameter	**Definition**
E_{CR}	Energy Consumption Rate
α	Energy fraction of advanced nodes
β	Energy fraction of intermediate nodes
f	Quantity fraction of intermediate nodes
f_o	Quantity fraction of advanced nodes
E_o	Normal node's initial energy
E_{avg}	Average energy of node
E_T	Total energy of nodes
n	Total number of nodes in the network
HoC	Head of Cluster
E_{res}	Residual energy of node
HoC_{prob}	Head of Cluster probability value
P_{opt}	Optimum probability
N_{prx}	Node proximity
N_C	Number of cluster nodes
C_R	Current value of round
T_R	Total value of rounds
E_{Tran}	Energy consumed in transmission
E_{Recp}	Energy consumed in data reception
E_{Agg}	Energy consumed in data aggregation
$D_{N_S(i)}$	Energy consumed in data aggregation
ARN_o	Adaptive Random Number
$Total_N$	Total nodes in the network
D_N	Dead nodes
R_N	Generated random number
$D_{(avg)(i-j)}$	Average distance between i^th and j^th node
$Nprx_{norm}$	Normalized value of node proximity
$Nprx_i$	Node proximity of i^th node
$Nprx_{min}$	Minimum value of node proximity
$Nprx_{max}$	Maximum value of node proximity
T_Ni	Threshold value of i^th node

through the Euclidean distance formula. Further, D_{avg} is computed through the following equation (2).

$$D_{avg} = \frac{1}{n} * \sum_{i=1}^{N} D_{N\_S}(i) \tag{2}$$

Node proximity: This is one of the essentials parameter which proliferates the energy saving of the cluster member nodes as it tends to select that node as HoC which has higher number of neighboring nodes. Here, node proximity or node density is denoted by N_{prx}. The empirical computation of node proximity is explained further in this section.

HoC count: It is imperative to keep the count on every cluster node so that the penalization for the frequent selection as a HoC to any node could be avoided. It is decided based on the epoch for each candidate node. An epoch is the set of rounds that ensures that any particular node will be selected as HoC at least once. Initially, the probability value is assigned as 0.1 and it keeps on incrementing until it reaches value 1 and at this value it is declared as HoC. As soon as the node is declared as HoC, the probability value is re-assigned as value 0.1. This really helps in taking the control on the frequency of a node being selected as HoC. Here, we denote HoC count as HoC_{prob}.

Energy Consumption Rate: It is obligatory to investigate the energy consumption rate of each node to select and appropriate HoC. For the first round it is initialized to value 1 (due to multiplication with the main expression used in equation (3), it does not make any difference for the first round) but afterwards it is updated according to the energy consumed by the candidate node in each round.

$$E_{CR} = \frac{(E_{Tran} + E_{Recp} + E_{Agg})}{C_R / T_R} \tag{3}$$

As per the process, the above mentioned parameters are included for the computation of probability value i.e., $P(i)$ for each node to be selected as HoC as given in following equation (4).

$$P(i) = \begin{cases} \frac{P_{opt}*E_{res}*D_{N\_S}(i)*N_{prx}*HoC_{prob}}{((1+f*\alpha+f_0*\beta)*D_{avg}*E_{avg}*E_{CR})}, \\ \frac{P_{opt}*E_{res}*(1+\beta)*D_{N\_S}(i)*N_{prx}*HoC_{prob}}{((1+f*\alpha+f_0*\beta)*D_{avg}*E_{avg}*E_{CR})}, \\ \frac{P_{opt}*E_{res}*(1+\alpha)*D_{N\_S}(i)*N_{prx}*HoC_{prob}}{((1+f*\alpha+f_0*\beta)*D_{avg}*E_{avg}*E_{CR})}, \end{cases} \tag{4}$$

In above equation (4), the three rows in RHS are for normal, intermediate and advanced, respectively. Further, the symbols used have following meaning. The optimum probability i.e., the pre-defined percentage of number of HoCs is represented by P_{opt}, to understand it better, if the value of P_{opt} is 0.1 it means the percentage number of HoCs are 10%. P_{opt} is the probability which accounts to define the quantity of HoCs in the network. The rest of the parameters are defined in the Table 1. The detailed computation of N_{prx} is done based on following steps.

The distance computation of i^{th} node from the j^{th} node of cluster (where i and j ranges from 1 to total number of cluster nodes i.e., N_C), is determined through the Euclidean distance.

D_{i-j} is the distance of each i^{th} node from the j^{th} node of cluster. The $node(i).x$ and $node(i).y$ are the

Cartesian coordinate for i^{th} node, similarly, $node(j).x$ and $node(j).y$ are the Cartesian coordinates for j^{th} node in the same cluster.

$$\begin{aligned} N_{prx}(i) &= D_{(avg)(i-j)} \\ &= \frac{1}{N_C} * \sum_{i=1}^{N_C} * \sum_{j=1}^{N_C-1} D_{i-j} \end{aligned} \tag{5}$$

The average distance of each node from the other nodes in the cluster represented by $D_{(avg)(i-j)}$ is computed through the equation (5). It is essential to note that in this work, $D_{(avg)(i-j)}$ is said to be a node proximity of the nodes as least the distance among the nodes, more will be the node proximity. The node with the minimum average distance from the other nodes is said to have highest value of node proximity i.e., $N_{prx_{max}}$ and vice versa. The minimum value of node density of node is denoted by $N_{prx_{min}}$.

Normalization of node density is done to keep the final value of node density of the node within the range of $[0-1]$ by using the equation (6).

$$N_{prx_{norm}}(i) = \frac{N_{prx_{(i)}} - N_{prx_{min}}}{N_{prx_{max}} - N_{prx_{min}}} \tag{6}$$

The node with reported higher value of node proximity is prompted to be the HoC.

The threshold formula i.e., $T_N(i)$ for each i^{th} node for the CIRP follows the equation (7).

$$T_N(i) = \begin{cases} \frac{P(i)}{1-P(i)\left(C_R mod\left(\frac{1}{P(i)}\right)\right)}, & if\ S(i)\,\epsilon\, G \\ 0, & otherwise \end{cases} \tag{7}$$

In above equation (7), C_R is the current round, $S(i)$ is the current i^{th} node (where i ranges from 1 to n) under threshold computation and G denotes the group of those nodes which are still to become HoC. Meanwhile, random number i.e., R_N is also generated for each candidate node and the corresponding threshold value for the particular node is compared with that R_N. If R_N for the i^{th} node is less than the threshold value $T_N(i)$ then that node is designated as HoC, otherwise it is called as cluster member node.

To put a control on the random number dependency, an adaptive random number is computed in equation (8).

$$ARN_o = \frac{Total_N - D_N}{Total_N} * R_N \tag{8}$$

Once the set up phase is established, the steady state phase comes into operation. It takes the charge for handling the data transmission in the network. The two genres of communication are involved in this phase; intra-cluster and inter-cluster communication.

Intra-cluster communication: This form of communication is commenced within the cluster as selected HoCs collect data from the nodes and performs data aggregation before forwarding it to the sink. The data collection from each node is performed based on the TDMA scheduling in which each node is assigned a particular slot for data forwarding.
Inter-cluster communication: Although, in this proposed routing strategy, there is no multi-hop communication, but this definition is included to avoid any ambiguity. All HoCs forward their data directly to the sink and inhibit any communication within the different HoCs i.e., inter-cluster routing.

B. Intelligent Clustering Algorithm Design

The AI empowered methodology for CIRP is delineated as follows. It consists of network operation, setup , startup and steady state and energy monitoring phases. The Algorithm 1 presents the different steps considered for the operation of CIRP protocol.

1) Explicative remarks for Algorithm 1: The Algorithm 1 presents the different steps considered for the operation of CIRP protocol. The algorithm is acquainted with some input factors; n (i.e., the total number of nodes deployed), r_{max} defines the total number of rounds encountered during its operation, computed ARN_o (Adaptive Random Number) and located CMGNs around the network. The output sought from this algorithm includes the selected CH, status of alive and dead nodes. Algorithm 1 is explained as follows.

Step 1 in Algorithm 1 uses the random generator for the coordinates to locate each node in the network and further, the CMGNs placement is done around the network area. Step 2 computes E_T from [11], which is further exploited in probability expression. Step 3 initializes the HoC parameter to store number of selected CHs in it. Steps 4- 30 execute *for* loop to cover the whole operational framework of CIRP i.e., set up and steady state phase. Step 5 and Step 6 initialize the variables for storing the updated number of alive and dead nodes during the network run. Steps 7-15 monitor the number of dead and alive nodes and Steps 16-26 execute *for* loop to commence the CH selection process. Step 18 takes the charge for computing the different parameters for the CH selection. Step 19 explores the comparison between the ARN_o and $T_N(i)$, and decides for the node to be CH or the cluster member node. Step 20 updates the number of HoC with respect to above stated conditions. Step 22 assigns the slot for the cluster members to perform data transmission to the selected HoC. Step 23 updates the current value of energy according to fundamental radio energy model. Steps 27-29 state a condition to terminate the whole loop of iterations based on the condition if all nodes are dead. Step 30 ends the *for* loop.

C. Computational complexity analysis of CIRP Algorithm

The applicability potential of any algorithm is decided by its complexity analysis. The computational complexity

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.3006276, IEEE Transactions on Industrial Informatics

Algorithm 1 CIRP Algorithm

```
Input: n, r_max, ARN, snk1 = (50, 105); snk2 = (50, −5);
snk3 = (105, 50); snk4 = (−5, 50)
Output: A = HoC_S, deadnodes, alivenodes
Heterogeneous nodes being deployed randomly and
  CMGNs placed at given location.
Calculate E_T
HoC=0
for r= 1 to r_max do
    alive_nodes =n
    dead=0
    for i= 1 to n do
        if E(i) == 0 then
            dead=dead+1
            if dead==n then
                all_dead=C_R
            end if
            alive = alive - dead
        end if
    end for
    for i= 1 to n do
        if E(i) > 0 then
            Compute D_{N_S}, D_avg, N_prx, E_avg, E_CR,
  P(i), T_N(i), ARN_o
            if T_N(i) > ARN_o then
                HoC_S=HoC_S + 1;  /*transmission to
  nearest CMGNs*/
            else
                i^th node ← slot for data tx to HoC
                Update E_(i) by using radio model
            end if
        end if
    end for
    if dead==n then
        break
    end if
end for
```

of CIRP algorithm depends upon some essential factors, namely r_{max} (maximum number of rounds iterated) and n (total nodes in the network). The following lemma is devised to render a detailed insight into the complexity analysis of proposed algorithm.

Lemma 3.1: Starting from an initial point the trajectory generated by the CIRP Algorithm converges within a fixed number of iterations r_{max}=$O(1)$; therefore, the computational complexity of CIRP is $O(r_{max} * n)$

proof. Once the data transmission is initiated, the energy consumption of the nodes occurs in the proceeding rounds. The whole operation is reported until the total number of nodes are exhausted of their energies in the network. It is concluded that the r_{max} iterations are in inverse proportion to the number of dead nodes. Furthermore, it is observed that the number of nodes is fixed and all nodes tend to consume energy with respect to their positions in the network. Therefore, at a particular value of r_{max}, all nodes become dead. Hence, CIRP converges to a fixed point. The overall computational complexity of CIRP is $O(r_{max} * n)$.

Observe in Algorithm 1 that our idea to determine the nearest sink for the process of HoC selection, tends to have negligible computational complexity. The structure of the computational complexity remains same to the other state of art algorithms as well.

IV. Results and Discussion

In this section, the simulation results of CIRP are discussed against the state-of-the-art protocols namely, MEEC, ZCA, OHCR, IDHR and TEDRP.

A. Simulation analysis of CIRP for performance metrics

We set initial stock of the energy of the normal nodes is 1 Joule, and it is two and three folds to the that for normal nodes for intermediate and advanced nodes, respectively. We set the number of intermediate and advanced nodes as 20% and 10% of the whole nodes quantity in the network. To address the scalability for CIRP, two cases; Case A and Case B are presented wherein former deals with the network dimensions of $100X100$ $meter^2$ deployed with 100 nodes, and latter considers the network area of $500X500$ $meter^2$ deployed with 300 nodes, respectively.

The status of performance metrics is investigated for CIRP protocol against the other protocols.

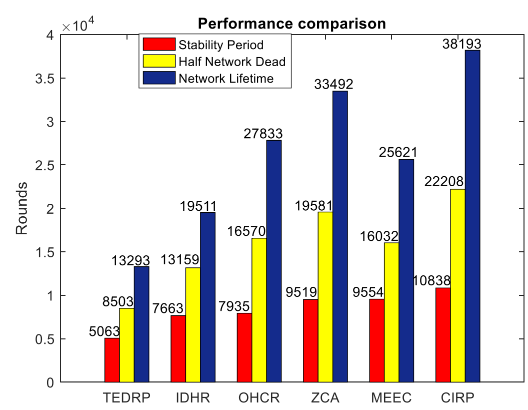


Figure 2: Case A: Performance comparison of CIRP for different metrics

Stability Period: As the name suggests, it renders stability to the network which is determined by the number of rounds successfully completed before the first node is dead.

For Case A: it is evident from Figure 2 CIRP contributes to 13.43%, 13.85%, 36.58%, 41.43%, and 114% gigantic

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.3006276, IEEE Transactions on Industrial Informatics

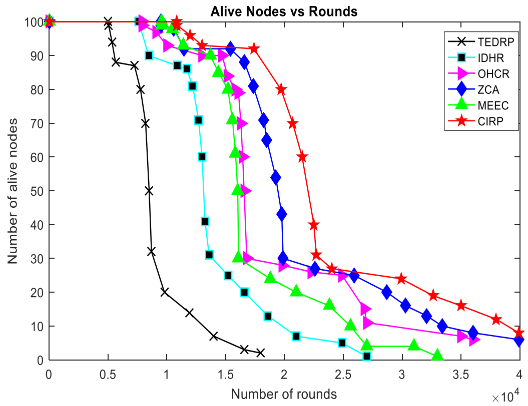


Figure 3: Case A: Comparative analysis of alive nodes of CIRP

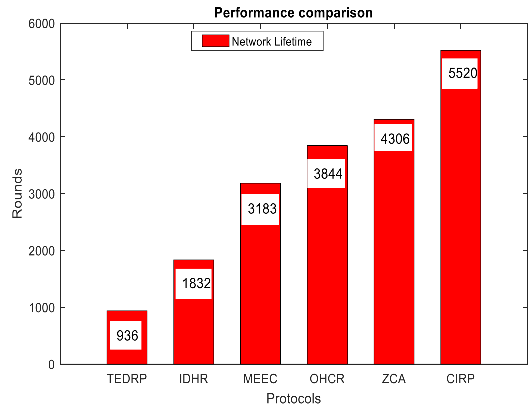


Figure 4: Case B: Performance comparison of CIRP for different metrics

enhancement in stability period vis-à-vis MEEC, ZCA, OHCR, IDHR and TEDRP protocols, respectively. In CIRP, for the given number of rounds, the number of alive nodes are more for the whole network run as shown in Figure 3 . CIRP improves stability period of MEEC by 13.43% which is accounted to the usage of CMGNs that eventually results in reducing the geometric distance among the sink and the respective HoC.

For Case B: As the network area is enlarged, the number of data transmissions are ameliorated enormously, therefore, the stability period is very less for this case. Hence it is not taken into consideration for the proposed scenario. However, the performance of CIRP is found still better than the competitive protocols as shown in Figure 4 and 5.

Network lifetime: It can be defined as the survival period of all nodes in the network, i.e., the rounds covered till all nodes are dead. Although, this parameter is mainly explored for attended applications but in other applications as well, the information availability is assured by this metric.

For Case A: It is observed that network lifetime of CIRP is improved by 49%, 14%, 37.22%, 95.75%, and 187.31% vis-à-vis MEEC, ZCA, OHCR, IDHR and TEDRP protocols, respectively.

For Case B: the network lifetime is improved by 28.19%, 43.6%, 152.86%, 201.3%, 489.7% vis-à-vis ZCA, OHCR, MEEC, IDHR and TEDRP protocols, respectively. The number of rounds completed by the above said protocols are shown in Figure 3. The improvement in network lifetime is attributed to the mitigation of hotspot problem vis-à-vis ZCA, OHCR, IDHR and TEDRP protocols. CIRP deals with single hop communication that too at reduced distance, whereas due to only single sink approach, the other protocols namely, ZCA, OHCR and TEDRP follow multi hop communication.

Network's remaining energy: The one another significant factor is the network's remaining energy which demonstrates how efficiently the energy is consumed.

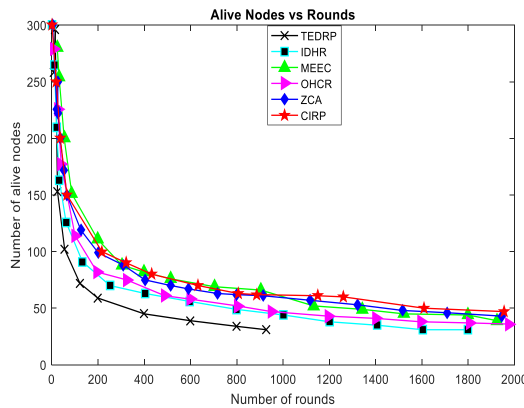


Figure 5: Case B: Comparative analysis of alive nodes of CIRP

For Case A and Case B: As evident from the Figure 6 the remaining energy of the nodes is saved in a significant proportion in CIRP vis-à-vis other protocols due to the number of communication hops which are reduced to only 1. The graph of remaining energy for Case B, follows the same pattern as that of Figure 5 and hence, it is not considered here. It can be analyzed that at the given number of rounds, the remaining energy is at higher value for CIRP vis-à-vis MEEC, ZCA, OHCR, IDHR and TEDRP protocols.

Throughput: The throughput is defined as the quantity of data packets sent successfully to the sink per unit round. It is one of the recurring parameter for proliferating the QoS (Quality of Service) that further ensures the reliability of the network.

For Case A: It is observed through the simulation analysis that the number of packets transmitted to the nearest sink in CIRP is 1440946, 1027384, 1269466, 959246,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.3006276, IEEE Transactions on Industrial Informatics

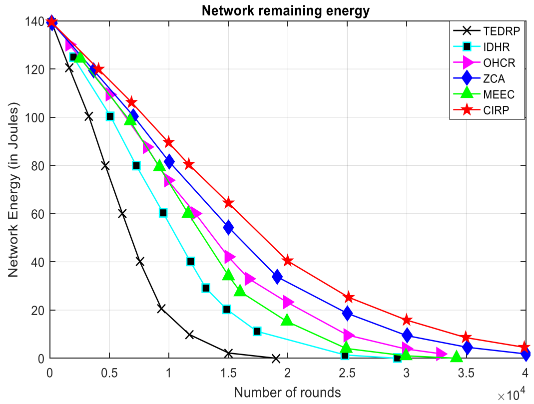


Figure 6: Case A: Comparative analysis of network's remaining energy of CIRP

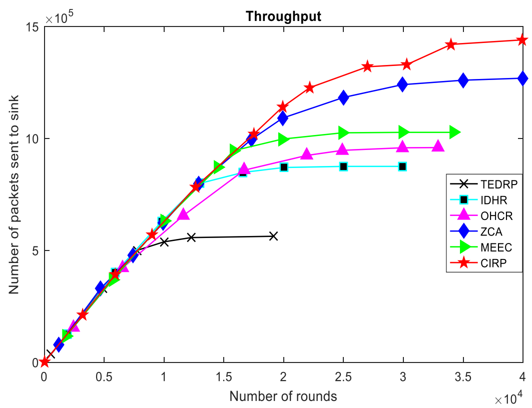


Figure 7: Case A: Comparative analysis of throughput of CIRP

875230, and 562750 data packets in case of MEEC, ZCA, OHCR, IDHR and TEDRP protocols, respectively. in the network. In addition to this, CIRP also outperform MEEC in throughput metric as illustrated in Figure 7. It is observed that CIRP transmits 413562 data packets more than that sent by MEEC i.e., improving throughput by 40.25%.

For Case B: Similar analysis of throughput is observed for Case B as shown in Figure 8 wherein CIRP outperforms the other protocols in the transmissions of more number of data packets and that too for the longer interval of time. The improvement in throughput is obtained due to the network longevity achieved by CIRP which helps in data collection and hence, transmissions of enhanced number of packets to the sink are performed.

V. Conclusion

We propose and evaluate a novel Clustering with AI-based Routing Protocol, CIRP, for massive industry 4.0 sensor networking. In our settings, the new design incorporates multiple gateway nodes outside the network for data dissemination to deal with energy hole problem. The energy efficient HoC selection is performed considering five main factors, viz. ratio of 'remaining energy to the initial energy', distance factor, node proximity, HoC count, and 'energy consumption rate'. CIRP has demonstrated promising performance improvements when compared to the state of the art protocols.

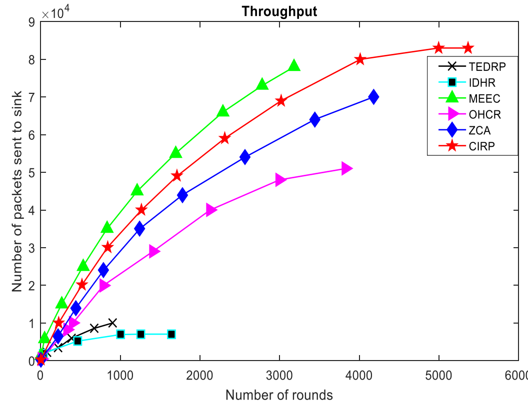


Figure 8: Case B: Comparative analysis of throughput of CIRP

References

[1] Ren C Luo and Tung Jung Hsiao. Dynamic wireless indoor localization incorporating with an autonomous mobile robot based on an adaptive signal model fingerprinting approach. *IEEE Transactions on Industrial Electronics*, 66(3):1940–1951, 2019.

[2] Faezeh Farivar et al. Artificial intelligence for detection, estimation, and compensation of malicious attacks in nonlinear cyber physical systems and industrial IoT. *IEEE transactions on industrial informatics*, 2019.

[3] Shree Krishna Sharma and Xianbin Wang. Towards massive machine type communications in ultra-dense cellular IoT networks: Current issues and machine learning-assisted solutions. *IEEE Communications Surveys & Tutorials*, 2019.

[4] Tian Wang, Hao Luo, Weijia Jia, Anfeng Liu, and Mande Xie. MTES: An intelligent trust evaluation scheme in sensor-cloud enabled industrial internet of things. *IEEE Transactions on Industrial Informatics*, 2019.

[5] Ozgur B Akan, Sergey Andreev, and Ciprian Dobre. Internet of things and sensor networks. *IEEE Communications Magazine*, 57(2):40–40, 2019.

[6] Fakhrosadat Fanian and Marjan Kuchaki Rafsanjani. Cluster-based routing protocols in wireless sensor networks: A survey based on methodology. *Journal of Network and Computer Applications*, 2019.

[7] Jingjing Zhong and *et al.* An intelligent and improved density and distance-based clustering approach for industrial survey data classification. *Expert Systems with Applications*, 68:21–28, 2017.

[8] Toan Van Nguyen, Tri-Nhu Do, Vo Nguyen Quoc Bao, Daniel Benevides da Costa, and Beongku An. On the performance of multihop cognitive wireless powered d2d communications in wsns. *IEEE Transactions on Vehicular Technology*, 69(3):2684–2699, 2020.

[9] Xintao Huan, Kyeong Soo Kim, Sanghyuk Lee, Eng Gee Lim, and Alan Marshall. A beaconless asymmetric energy-efficient time synchronization scheme for resource-constrained multi-hop wireless sensor networks. *IEEE Transactions on Communications*, 2019.

[10] Houlian Wang, Gongbo Zhou, Laksh Bhatia, Zhencai Zhu, Wei Li, and Julie A McCann. Energy-neutral and QoS-aware protocol in wireless sensor networks for health monitoring of hoisting systems. *IEEE Transactions on Industrial Informatics*, 2020.

[11] Sandeep Verma, Neetu Sood, and Ajay Kumar Sharma. Genetic algorithm-based optimized cluster head selection for single and multiple data sinks in heterogeneous wireless sensor network. *Applied Soft Computing*, 85:105788, 2019.

[12] Jian Li and Prasant Mohapatra. Analytical modeling and mitigation techniques for the energy hole problem in sensor networks. *Pervasive and mobile Computing*, 3(3):233–254, 2007.

[13] Hadi Asharioun, Hassan Asadollahi, Tat-Chee Wan, and Niyayesh Gharaei. A survey on analytical modeling and mitigation techniques for the energy hole problem in corona-based wireless sensor network. *Wireless Personal Communications*, 81(1):161–187, 2015.

[14] Jinho Choi. A cross-layer approach to data-aided sensing using compressive random access. *IEEE Internet of Things Journal*, 6(4):7093–7102, 2019.

[15] You-Chiun Wang and Kuan-Chung Chen. Efficient path planning for a mobile sink to reliably gather data from sensors with diverse sensing rates and limited buffers. *IEEE Transactions on Mobile Computing*, 18(7):1527–1540, 2018.

[16] Vehbi C Gungor and Gerhard P Hancke. Industrial wireless sensor networks: Challenges, design principles, and technical approaches. *IEEE Transactions on industrial electronics*, 56(10):4258–4265, 2009.

[17] Ying Zhang, Guangyuan Yang, and Bin Zhang. Fw-pso algorithm to enhance the invulnerability of industrial wireless sensor networks topology. *Sensors*, 20(4):1114, 2020.

[18] Siguang Chen, Shujun Zhang, Xiaoyao Zheng, and Xiukai Ruan. Layered adaptive compression design for efficient data collection in industrial wireless sensor networks. *Journal of Network and Computer Applications*, 129:37–45, 2019.

[19] Mithun Mukherjee, Lei Shu, R Venkatesha Prasad, Di Wang, and Gerhard P Hancke. Sleep scheduling for unbalanced energy harvesting in industrial wireless sensor networks. *IEEE Communications Magazine*, 57(2):108–115, 2019.

[20] IS Akila and R Venkatesan. An energy balanced geo-cluster head set based multi-hop routing for wireless sensor networks. *Cluster Computing*, 22(4):9865–9874, 2019.

[21] Reem E Mohemed, Ahmed I Saleh, Maher Abdelrazzak, and Ahmed S Samra. Energy-efficient routing protocols for solving energy hole problem in wireless sensor networks. *Computer Networks*, 114:51–66, 2017.

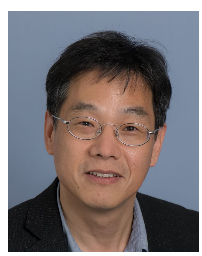
[22] Nitin Mittal, Urvinder Singh, and Balwinder Singh Sohi. A novel energy efficient stable clustering approach for wireless sensor networks. *Wireless Personal Communications*, 95(3):2947–2971, 2017.

Shiva Raj Pokhrel ((M'19)) is a Lecturer of Mobile Computing at Deakin University, Geelong, Australia. He received the B.E./M.E. degree in 2007/2013 from Pokhara University, Nepal and PhD degree in 2017 from Swinburne University of Technology, Australia. He was a research fellow at the University of Melbourne (2017-2018) and a network engineer at Nepal Telecom (2007-2014). His research interests include federated learning, industry 4.0, blockchain modeling, optimization, recommender systems, beyond 5G, cloud computing, dynamics control, Internet of Things and cyber-physical systems as well as their applications in smart manufacturing, autonomous vehicles and cities. He serves/served as the Workshop Chair/Publicity Co-Chair for several IEEE/ACM conferences including IEEE INFOCOM, IEEE GLOBECOM, IEEE ICC, ACM MobiCom, and more. He was a recipient of the prestigious Marie Skłodowska-Curie grant Fellowship in 2017.

Sandeep Verma is pursuing PhD in Electronics and Communication Engineering Department in Dr BR Ambedkar National Institute of Technology, Jalandhar. He completed his Masters of Engineering in Electronics and Communication from NITTTR Chandigarh in 2013 and Bachelor of Technology degree in Electronics and Communication from LCET Katani Kalan, Ludhiana in 2010. His research interest includes energy efficient Wireless Sensor Networks, IoT based architectures and Intelligent Transportation System.

Sahil Garg (M'18) is currently a Postdoctoral Research Fellow at École de technologie supérieure (ÉTS), Montreal, Canada. He got his Ph.D. degree from Thapar Institute of Engineering and Technology, India, in 2018. His research interests are mainly in the areas of Machine Learning, Big Data Analytics, Knowledge Discovery, Cloud Computing, Internet of Things, Software Defined Networking, and Vehicular Ad-hoc Networks. He is also an Associate Editor of IEEE Network Magazine, IEEE Systems Journal, Elsevier's Future Generation Computer Systems (FGCS), and Wiley's International Journal of Communication Systems (IJCS). In addition, he guest edited a number of Special Issues in top-cited journals including IEEE T-ITS, IEEE TII, the IEEE IoT Journal, IEEE Network Magazine, Future Generation Computer Systems (Elsevier), Neural Computing Applications (Springer), etc.

Ajay K. Sharma joined IKG Punjab Technical University, Jalandhar as 13th Vice Chancellor. Earlier, Prof. Sharma was Director, National Institute of Technology, Delhi and during this period he also served as Director, Hamirpur H.P (Additional Charge) and Mentor Director (Additional Charge), IIIT Una. He received his Ph.D. in Electronics Communication and Computer Engineering in the year 1999 from NIT Kurukshetra. His major areas of interest are broadband optical wireless communication systems and networks, dispersion compensation, fiber nonlinearities, optical soliton transmission, WDM systems and networks, Radio-over-Fiber (RoF) and wireless sensor networks and computer communication.



Jinho Choi (SM'02) was born in Seoul, Korea. He received B.E. (magna cum laude) degree in electronics engineering in 1989 from Sogang University, Seoul, and M.S.E. and Ph.D. degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST) in 1991 and 1994, respectively. He is with the School of Information Technology, Burwood, Deakin University, Australia, as a Professor. Prior to joining Deakin in 2018, he was with Swansea University, United Kingdom, as a Professor/Chair in Wireless, and Gwangju Institute of Science and Technology (GIST), Korea, as a Professor. His research interests include the Internet of Things (IoT), wireless communications, and statistical signal processing. He authored two books published by Cambridge University Press in 2006 and 2010. Prof. Choi received the 1999 Best Paper Award for Signal Processing from EURASIP, 2009 Best Paper Award from WPMC (Conference), and is Senior Member of IEEE. Currently, he is an Editor of IEEE Trans. Communications and IEEE Wireless Communications Letters and a Division Editor of Journal of Communications and Networks (JCN). We also had served as an Associate Editor or Editor of other journals including IEEE Communications Letters, JCN, IEEE Trans. Vehicular Technology, and ETRI journal.