Efficient Localized Data Aggregation of IoT Devices and its Application in Agriculture

C. Sathish, K. Srinivasan

Abstract: The Internet of Things (IoT) has a likely future for all technologies that are associated to the life of humans. Any communication between the social environments along with the user contexts will be made through the smart interfaces. The IoT will have to link to different diverse devices found in the Wireless Sensor Network (WSN). Thus, the routing optimization which is energy efficient has become the primary factor in the performance of the network in the IoT. The widely popular routing used in WSN, Multi-hop Low Energy Adaptive Clustering Hierarchy (LEACH) protocol, is found to be energy inefficient. The work will deal with the choice of finding the optimal path in routing through meta heuristic techniques to improve the lifespan of the network and the energy efficiency of the network. There are different techniques of metaheuristics such as Teacher Learning Based Optimization (TLBO) and Particle Swarm Optimization (PSO) were used effectively for finding optimal solutions for problems. In this work, the PSO-based algorithms were used for locating an optimal sink position to their nodes that make the network efficient in terms of energy. The TLBO metaheuristic is population-based and is based on the concept of teaching and the learning procedure observed in a classroom, which is adapted to the routing problem. The results of the experiment prove that the proposed technique was achieved better levels of performance compared to the other methods.

Keywords: Internet of Things (IoT), Low Energy Adaptive Clustering Hierarchy (LEACH), Particle Swarm Optimization (PSO) and Teaching Learning Based Optimization (TLBO), Wireless Sensor Network (WSN).

I. INTRODUCTION

Internet of Things (IoT) reflects a connected set of anything at any place and for any service in any network. It is a popular trend seen in the next-generation technologies which may affect the entire business field which is an interconnection of recognizable smart devices in the infrastructure of the internet in today’s situation that has increased advantages. The advantages includes improved connectivity of the devices, services, and systems that go beyond situations of Machine-to-Machine (M2M). The IoT also offers suitable solutions for a range of operations and applications like health care, industrial control, retail, logistics, emergency services, security, structural health, waste management, traffic congestion and smart cities [1].

There are some real-life objects like actuators or sensors that are connected to one another for transferring sensed data to the centralized servers wherein data was stored together and kept accessible for the users having correct credentials to access. The characteristics of the IoT that makes use of the wireless technologies were found to be different from the traditional wireless or wired networks owing to the devices that participate in the communication. Additionally, the traffic for each IoT device that collects data and further transmits smaller amounts of data to the IoT servers even though the generated data from large numbers of objects can have an impact on the performance of the network. Also, the IoT network will operate in a stable and sustainable manner for longer periods of time without a need for intervention [2]. The devices in IoT operate on the basis of the power sources of batteries and thus it is very important to keep them efficient in terms of energy and device management. While looking at a domain of the Wireless Sensor Network (WSN), the efficiency of energy for the sensor nodes and the prolonging of the network lifetime are issues of research for a long time. In the same way, as the IoT devices that operate in the network paradigm of the IoT were also battery operated with the consumption of battery that has to be kept in mind at the time of developing the IoT deployment. But, another major challenge that is faced to be a stringent constraint today in terms of bandwidth of communication, storage of memory and energy consumption. But, there still is a supporting technique in the WSN which is a problem of localization with a major concern in data markers, data aggregation and geographic routing for which data becomes meaningless in case of no information of location being available. The technique of traditional locations like the GPS may not be used by the WSNs since there is a costly need for sophisticated equipment with high consumption of energy that has constrained applications of scale. For the conservation of the WSN energy, there is a consideration of this being a core issue and the costs and node size that is small enough that can be applied to applications of a large scale [3]. For addressing the issue, there are several algorithms of localization that were developed and they do not make use of the technique of GPS directly but will employ it to be of assistance with further efforts made to excavate the WSN. Until now, the localization algorithms that were proposed were classified into two which were the range-based localization and the range-free localization.

Agriculture has been the backbone of the economic growth of India. The primary barrier arising in conventional farming is climate change. Its effects are: there may be heavy rainfall, heatwaves, and intense storms. Owing to this, there is a decrease in productivity. The climatic changes can raise other environmental consequences like seasonal changes in the plant life cycle. For the purpose of boosting productivity and minimizing barriers in the field of agriculture, innovative technology such as IoT will have to be employed. The IoT has been effectively been adapted in the agriculture industry thus allowing farmers to face the challenges.
Efficient Localized Data Aggregation of IoT Devices and its Application in Agriculture

The farmers can now have easy access to various data related to their agriculture, as well as knowledge with IoT [4]. The market of smart agriculture can touch $18.45 Billion by the year 2022, at a CAGR which is about 13.8%. The BI evaluates that about 75 million IoT devices may be used in agriculture by 2020 at a CAGR of about 20%. The devices of the IoT can be of major help while aiming at enhancing production. Furthermore, smart agriculture helps in monitoring the productivity of livestock and health. The IoT sensors can also provide the farmers with some information on the yields of crops, soil nutrition, pest infestation, and rainfall. The IoT has real-time, exact and shared features bringing about major changes to the supply chain of agriculture providing a serious knowledge to establish agricultural logistics smoothly [5].

Aggregation of data will be responsible for bringing about an increase in the lifetime of the network and bringing down consumption of energy. Every node will have the capability of keeping the received data from the subsequent nodes or the data created for a particular timespan and combine them to send out its aggregated result. Data will have to be aggregated in order to maintain energy and resources. Aggregation of data will be the actual process of certain nodes to assemble the results that are collected from the other nodes. The data that is collected has to be handled by the same node for reducing transmission and this will be the base station for a user with permission to interact with the same network. Aggregated data has been transferred into the sink through an effective route. Normally, the methods of data aggregation have offered many other advantages such as: helps in improving the accuracy of information and efficiency achieved through the entire network. There is a specific level of redundancy that is present in the data which is collected from the nodes and so this procedure will be needed for decreasing unwanted information. Further, it can decrease the traffic load and save node energy [6].

Metaheuristic algorithms are powerful tools that solve problems of optimization. They are widely used in dealing with problems that are complicated in various areas such as engineering and science. There are different methods used to solve problems of optimization and most of them are based on mathematical programming. The methods will need various conditions that are not met by other real-world problems of optimization. Further, there are plenty of real-world problems that may be challenging to solve with limited resources of time. Even though the traditional methods of mathematical programming are ideal to deal with problems of a small-scale and is also posed to solve instances of a large-scale for the Non-Deterministic Polynomial (NP)-hard problems [7]. There have been plenty of metaheuristic algorithms working with several mechanisms inspired from the physical, mechanical, cultural, political, economic, social and natural concepts. From among them, the Ant Colony Optimization (ACO), Tabu Search (TS), Artificial Bee Colony (ABC), Gravitational Search Algorithm (GSA), Imperialistic Competition Algorithm (ICA), Genetic Algorithm (GA), and so on are applied successfully to several problems of management and engineering. In this work, the PSO and the TLBO algorithm is proposed for optimizing the routing. Section 2 presents reviews of works available in the literature. Section three details various procedures used in the work. Section four elaborates an experimental results and section five concludes the work.

II. RELATED WORKS

Rauniyar et al., [8] had proposed a localization algorithm that was newly distributed. This was by using the Social Learning-based PSO (SL-PSO) for the IoT. With the SL-PSO algorithm, the algorithm works at precise localization of the deployed sensor nodes for reducing the complexity of computation to improve the lifetime of such resource-constrained sensor nodes in the IoT. The results of the experiment will represent the SL-PSO algorithm will not be able to improve the rate of convergence but can also reduce the average localization error in comparison to the traditional PSO and its variants. Hasan and Al-Turjman [9] presented a bio-inspired Particle Multi-Swarm Optimization (PMSO) based on the routing algorithm for constructing, recovering and selecting the k-service parameters. The strategy of multiple swarms will enable the exchange of messages from various positions. The proposed algorithm’s validity had been assessed and their results had demonstrated solutions of high quality in comparison with Canonical PSO (CPSO). Results had indicated the superiority of this multi-swarm PMSO that had an average improvement over that of the CPSO equal to 88.45% which was in terms of the sensors and about 89.15% and 86.51% which was under the topologies of ring and mesh respectively. A secure communication will be the primary issue for sensors as they will concentrate on the services as opposed to the security. For resolving this, Nagavalli and Ramachandran [10] had focused on the development of an effectual scheme of routing that made use of the Cuckoo Search Algorithm (CSA). The CSA was verified experimentally with the Node-Red simulation and had been tested in different cases. The outcomes had indicated that the algorithm proposed was effectual in modern algorithms.

Dhumane and Prasad [11] had proposed a new Multi-objective Fractional GSA (MOFGSA) for identifying optimal cluster heads for the routing protocol that is energy efficient. For extending the node lifetime, a Fractional GSA (FGSA) was proposed to find out an optimal cluster head node in an iterative manner for the IoT network mode. A CH node is chosen in the FGSA evaluated by a fitness function that uses multiple objectives like energy, link lifetime, delay and distance called the MOFGSA. The results of the simulation analysed the performance with MATLAB. The performance was compared to the other algorithms like Multi-Particle Swarm Immune Cooperation Algorithm, GSA and ABC. So, the MOFGSA algorithm proposed will ensure an extension of IoT node lifetime.

Said et al., [12] had proposed a new routing algorithm based on ACO for optimizing the choice of the path that is the best for the transmitting data in the IoT system.
The IoT situation was characterized based on the type of network. After this, it is applied to an appropriate ACO technique onto the network. Also, the algorithm considers the problem of routing in the intersected areas arising in the IoT system. Lastly, a Network Simulator 2 will be used for the assessment of the performance of the proposed method. The results of simulation demonstrated the method’s effectiveness which was in terms of the ratio of energy consumption, control bit overheads, throughput, bandwidth consumption, packet loss ratio, and end-to-end delay.

An algorithm for the indexing of cluster resources in the IoTs was based on an enhanced ACO was presented by Hong et al., [14]. There were directed graph models used for the construction of a model of distribution structure for the cluster resource indexing nodes in the IoTs that conduct a semantic association extraction of feature in a storage information of cluster resources and the flow of the IoTs. The proposed ACO was employed to crawl the network and obtain all cluster info. Based on the route data of the ACO, both velocity and position of a cluster resource indexing in the IoTs were updated. This was used to conduct both global search and a global search for initializing their clustering centre. The target function for cluster resource indexing of the IoTs was constructed and a parameter of the optimization was solved using a limitation for the minimum variance for the entire fitness. There was also a very strong capability of the global optimization found in the ACO that was used to understand the optimization of resource indexing.

III. METHODOLOGY

The Multihop-LEACH protocol, the proposed PSO and TLBO methods are discussed in this section.

A. Multihop-LEACH protocol

In WSN, the LEACH will be the primary protocol that employs hierarchical routing for the expansion of the lifetime of the network. All the network nodes will organize itself to be part of local clusters among the node which performs as the CH. Every non-CH node will transmit data to the CH, which in turn processes the data and forwards the aggregated data to the base station. The LEACH further incorporates a random cycle of a CH position of high energy for rotating amidst the sensors. The energy load of a CH is shared consistently among nodes [15].

The actual distance between the CH and its base station will be enhanced as soon as the diameter of the network is far from a certain stage. The condition is not suitable for the LEACH, as CHs are single-hop from base station. The multi-hop LEACH, an extension of LEACH was introduced to solve the issue. The Multi-hop LEACH is also more energy efficient. The Multihop-LEACH was a new routing algorithm based on the cluster that is chosen as a CH that gathers data from every sensor node in the cluster. This combined all data obtained by means of fusing techniques and broadcasting data into an optimal path which was among the CH and base station with the remaining intermediate CH by retaining the CHs as a relay station and broadcast it through them.

A Multihop-LEACH function [16] was split up into two different stages: a set-up stage and a steady-state transfer stage. The former includes the systematizing of clusters that were followed by selecting the CH. In the entire phase, based on the probability percentage network clustering is chosen. The CHs as re-selected based on the number of times a node is chosen. The choice had been made by each node n choosing a number in a random manner between 0 and 1. If the number is lower in comparison to the threshold T (n), the node will turn the CH for the current round as in equation (1):

\[
T(n) = \begin{cases} 
\frac{p}{1-p(r \mod 1/p)} & \text{if } n \in G, \\
0 & \text{otherwise.}
\end{cases}
\]

Wherein, P is the necessary probability of the CH, \( r \) denotes the number of the present round and G the set of nodes that were not the CHs in the 1/P rounds.

A Multi-hop LEACH has a communication process that includes inter-cluster and intra-cluster communication. In the former, the entire network is classified into clusters and each one in the CH will be in charge of communicating among the cluster nodes. The CHs will achieve information from the nodes at a single-hop and this will be combined and forwarded in a straight line to the intermediate CH. For the multi-hop inter-cluster communication, and for the huge distance was among the CH and the base station where the CH employed a new intermediate CH to ensure the transmission to base station was achieved.

B. Particle Swarm Optimization (PSO) Algorithm

A PSO algorithm deals with problems in optimization in non-linear functions. By employing an iterative process, the algorithm had performed a search to find an optimal solution for a particular problem within a d-dimensional space. In PSO, all particles had social behaviour by interacting among themselves using the swarm. For every step, the particles had been evaluated using a fitness function with their speed and position updated. This algorithm stopped at the time a quasi-optimal solution was identified or maximum number of iterations [17].

The equations (2) and (3) have defined both speed and position of updating every particle found in the t-th step in the algorithm.

In the case the determined particle is i within the d-dimensional search space, it will be: \( x_i = [x_{i1}, x_{i2}, x_{i3}, \ldots, x_{id}] \) which is a particle position vector;

\( v_i = [v_{i1}, v_{i2}, v_{i3}, \ldots, v_{id}] \) which is a particle velocity vector; the pBest denotes the individual position of a particle i which has the best of fitness value; the gBest denotes a global position having the best of swarm fitness value; \( c_1 \) and \( c_2 \) denote both social, as well as cognitive parameters, and; \( r_1 \) and \( r_2 \) denote the values that are generated randomly within an interval [0, 1].
A term \( \omega \) is the inertia factor for achieving better levels of control of particle velocity within the search space. This is initialized by using maximum value \( \omega_{\text{Max}} \) and is also linearly decreased to a minimum value \( \omega_{\text{Min}} \), and this is subject to the actual maximum number of steps in the algorithm.

\[
v_i(t+1) = \omega v_i(t) + c_1 r_1 (p_{\text{Best},i} - x_i(t)) + c_2 r_2 (g_{\text{Best}} - x_i(t))
\]

(2)

\[
x_i(t+1) = x_i(t) + v_i(t+1)
\]

(3)

Both the position and speed of every particle has been limited by constraints \([x_{\text{Min}}; x_{\text{Max}}]\) and \([v_{\text{Min}}; v_{\text{Max}}]\) for preventing the paths outside of search space from getting explored.

C. Teaching-Learning-Based Optimization (TLBO) Algorithm

The important process is teaching-learning wherein an individual attempts to learn from the other individuals for improving themselves. The algorithm contains two modes of learning: (i) teacher phase – learning through the teacher and (ii) learner phase – learning through interaction with the other learners. The TLBO is based on the population in which a student set (learners) is measured to be a population of different subjects that have been offered to analogous learners. The results of the learner will also be analogous to the value of fitness for the problem of optimization. The solution is the best for the whole population is the teacher. The TLBO operation has been explained with both the teacher and the learner phase [18].

The Teacher phase: This particular phase will simulate student learning through the teacher. In this phase, the teacher will convey knowledge to learners and will make an effort to bring about an increase in the mean result. If there are ‘\( m \)’ number of subjects that are offered to the ‘\( m \)’ number of the learners. For every sequential cycle of teaching-learning, \( i, M_j \) will be the mean result of learners in a certain subject ‘\( j \)’ (\( j = 1 \) to \( m \)). As a teacher is knowledgeable and experienced on a particular subject, the best learner in the population was taken to be a teacher for the algorithm. If \( X_{\text{Best}} \) is the actual result of that of the best learner taking the subjects identified to be a teacher for the cycle. The teacher further puts further effort into the level of knowledge for the entire class and the learners gain more knowledge in accordance with the teaching quality of the teacher and the learning quality of the learners in the class. Taking this fact into consideration, there is a difference between the result of a teacher and also the mean result of learners for every subject as in (4):

\[
\text{Difference}_\text{Mean}_{ji} = r_i (X_{j,\text{kbest,}i} - M_{ji})
\]

(4)

Wherein, \( X_{j,\text{kbest,}i} \) will be the teacher’s results (the best learner) for subject \( j \). The \( T_F \) denotes a teaching factor that decides the mean value that needs change and \( r_i \) denotes the random number within the range \([0, 1]\). The actual value of \( T_F \) will either be 1 or 2. The \( T_F \) value is randomly decided with equal probability (5):

\[
T_F = \text{round}(1 + \text{rand}(0,1)[2−1])
\]

(5)

Wherein, \( \text{rand} \) denotes a random number falling in the range \([0, 1]\). The \( T_F \) will not be a TLBO parameter. The \( T_F \) was not given to be an input to an algorithm and the value will be decided randomly using equation (5).

On the basis of the Difference_Mean, its existing solution will be updated in the phase of the teacher in accordance with the expression below (6):

\[
X'_{j,k,i} = X_{j,k,i} + \text{Difference}_\text{Mean}_{ji}
\]

(6)

Wherein, \( X'_{j,k,i} \) denotes \( X_{j,k,i} \) and its updated value.

Accept \( X'_{j,k,i} \) can give a function value that is better. The accepted function values in the end of the phase will be retained and the values will be the input of the learner phase.

It has to be noted that values of both \( r_i \) and \( T_F \) will affect the TLBO algorithm’s performance. \( r_i \) denotes a random number within the range \([0, 1]\) and \( T_F \) denotes its teaching factor. But the values of both \( r_i \) and \( T_F \) were generated in a random manner and the parameters were not supplied to be the input algorithm (unlike the supply of probabilities of both crossover and mutation in the Genetic Algorithm (GA), the inertia weight, the cognitive and the social parameters of the PSO, the colony size and the limit of the Artificial Bee Colony (ABC)). The TLBO will need tuning only for common parameters of control such as the algorithms based on population. So, the TLBO is known as the algorithm-specific parameter-less algorithm [20].

The Learner phase: This is the phase where the algorithm simulates learning by means of interaction. Students will gain knowledge by means of discussing and further interacting with the other students. The learner will further learn new data have better knowledge. The phase has a learning phenomenon that is expressed as shown below:

Choose two learners, \( P \) and \( Q \) so that \( X'_{\text{Total-P,j}} \neq X'_{\text{Total-Q,j}} \), in which, \( X'_{\text{Total-P,j}} \) and \( X'_{\text{Total-Q,j}} \) are updated values of \( X_{\text{Total-P,j}} \) and \( X_{\text{Total-Q,j}} \) in the teacher phase (7 and 8).

\[
X'_{j,P,i} = X_{j,P,i} + r_i (X'_{j,P,i} - X'_{j,Q,i})
\]

(7)

\[
X'_{j,Q,i} = X_{j,Q,i} + r_i (X'_{j,Q,i} - X'_{j,P,i})
\]

(8)

(The equations above are for the problems of maximization and the reverse will be true for the problems of minimization).

Accept \( X'_{j,P,i} \) in case it provides better function value according to its species count.

(6) Perform mutation probabilistically for modifying every non-elite habitat. Not re-compute every HSI.
(7) Now go to step (3) for the subsequent iteration and repeat until such time a predefined number of generations are met or this is done after a certain acceptable solution is found.

IV. RESULTS AND DISCUSSION

The multi-hop LEACH, PSO and TLBO methods are used. Experiments are carried out using 200 to 1000 number of nodes and 0 to 800 number of rounds. The average end to end delay, average Packet Delivery Ratio (PDR) and lifetime computation as shown in Tables I to III and Figures 1 to 3.

**Table I Average End to End Delay for TLBO**

<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>Multi Hop LEACH</th>
<th>PSO</th>
<th>TLBO</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.0019</td>
<td>0.0018</td>
<td>0.0019</td>
</tr>
<tr>
<td>400</td>
<td>0.0018</td>
<td>0.0023</td>
<td>0.0019</td>
</tr>
<tr>
<td>600</td>
<td>0.0186</td>
<td>0.0221</td>
<td>0.0194</td>
</tr>
<tr>
<td>800</td>
<td>0.032</td>
<td>0.0264</td>
<td>0.0305</td>
</tr>
<tr>
<td>1000</td>
<td>0.0686</td>
<td>0.0602</td>
<td>0.0688</td>
</tr>
</tbody>
</table>

**Fig 1 Average End to End Delay for TLBO**

Fig 1 shows that the TLBO has lower average end to end delay by same value & 5.4% for 200 number of nodes, by 5.4% & 19.04% for 400 number of nodes, by 4.21% & 13.01% for 600 number of nodes, by 4.8% & 13.33% for 800 number of nodes and by 0.29% & 13.33% for 1000 number of nodes when compared with multi-hop LEACH and PSO respectively.

**Table II Average Packet Delivery Ratio for TLBO**

<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>Multi Hop LEACH</th>
<th>PSO</th>
<th>TLBO</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.6381</td>
<td>0.7619</td>
<td>0.8307</td>
</tr>
<tr>
<td>400</td>
<td>0.6418</td>
<td>0.7346</td>
<td>0.825</td>
</tr>
<tr>
<td>600</td>
<td>0.6518</td>
<td>0.6816</td>
<td>0.872</td>
</tr>
<tr>
<td>800</td>
<td>0.5752</td>
<td>0.6504</td>
<td>0.7775</td>
</tr>
<tr>
<td>1000</td>
<td>0.547</td>
<td>0.6553</td>
<td>0.7446</td>
</tr>
</tbody>
</table>

**Fig 2 Average Packet Delivery Ratio for TLBO**

Fig 2 shows the TLBO has higher average PDR by 26.22% & 8.64% for 200 number of nodes, by 25.04% & 11.59% for 400 number of nodes, by 28.9% & 24.51% for 600 number of nodes, by 29.91% & 17.8% for 800 number of nodes and by 30.59% & 12.75% for 1000 number of nodes when compared with multi-hop LEACH and PSO respectively.

**Table III Lifetime Computation for TLBO**

<table>
<thead>
<tr>
<th>Number of rounds</th>
<th>Multi Hop LEACH</th>
<th>PSO</th>
<th>TLBO</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>84</td>
<td>91</td>
<td>95</td>
</tr>
<tr>
<td>200</td>
<td>72</td>
<td>76</td>
<td>90</td>
</tr>
<tr>
<td>300</td>
<td>59</td>
<td>67</td>
<td>81</td>
</tr>
<tr>
<td>400</td>
<td>13</td>
<td>19</td>
<td>45</td>
</tr>
<tr>
<td>500</td>
<td>9</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>600</td>
<td>5</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>700</td>
<td>2</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>800</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

**Fig 3 Lifetime Computation for TLBO**

Fig 3 shows that the TLBO has higher lifetime computation by 12.29% & 4.3% for 100 number of rounds, by 22.22% & 16.86% for 200 number of rounds, by 31.42% & 18.91% for 300 number of rounds, by 110.34% & 81.25% for 400 number of rounds, by 138.46% & 83.87% for 500 number of rounds and by 171.42% & 88.89% for 600 number of rounds when compared with multi-hop LEACH and PSO respectively.
V. CONCLUSION

There are more sensors that get interconnected each data for making the IoTs real. But as the WSNs contain devices that are limited in terms of energy, the techniques will save energy. This will have consequences that have to be considered thoroughly for maintaining effective communication. The technique of clustering was used in the multi-hop LEACH protocol for improving network lifetime. For this work, the PSO was originated from the swarm intelligence and this is based on all types of research on the fish flock and the bird flock movement and behaviour. The PSO is very robust and can achieve solutions to real-world problems of environmental monitoring easily. The TLBO makes use of two productive operators namely, the teacher and the learning phase for ideal solutions. Owing to its attractive traits, the TLBO was able to capture attention and has also been extended for handling constrained, large-scale, dynamic and multi-objective problems.

REFERENCES