UNIVERSITY OF THE
WEST *of* SCOTLAND
UWS

UWS Academic Portal

Multi-hop similarity-based-clustering framework for IoT-Oriented Software-Defined wireless sensor networks

Shafique, Ayesha; Asad, Muhammad; Aslam, Muhammad; Shaukat, Saima; Cao, Guo

IET Wireless Sensor Systems

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Multi-hop similarity-based-clustering framework for IoT-Oriented Software-Defined wireless sensor networks

Ayesha Shafique[1] | Muhammad Asad[2] | Muhammad Aslam[3] | Saima Shaukat[2] | Guo Cao[1]

[1]School of Computer Science, Nanjing University of Science and Technology, Nanjing, China

[2]Department of Computer Science, Nagoya Institute of Technology, Nagoya, Japan

[3]School of Computing, Engineering and Physical Sciences, University of the West of Scotland, Paisley, UK

Correspondence

Guo Cao, School of Computer Science, Nanjing University of Science and Technology, Nanjing, China.
Email: caoguo@njust.edu.cn

Abstract

The performance of Internet of Things (IoT)-based Wireless Sensor Networks (WSNs) depends on the routing protocol and the deployment technique in modern applications. In a plethora of IoT-WSNs applications, the IoT nodes are essential equipment to prolong the network lifetime with limited resources. Data similarity-based clustering protocols exploit the temporal correlation among the neighbouring sensor nodes through the subset of data. In bendy supervision, IoT-based Software Defined WSNs provide an optimistic resolution by allowing the control logic to be separated from the sensor nodes. The benefit of this SDN-based IoT architecture, allows the unified control of the entire IoT network, making it easier to implement on-demand network management protocols and applications. To this end, in this paper, we design a Multi-hop Similarity-based Clustering framework for IoT-oriented Software-Defined wireless sensor Networks (MSCSDNs). In particular, we construct data-similar application-aware clusters in order to minimise the communication overhead. Also, we adapt inter-cluster and intra-cluster multi-hop communication using adaptive normalised least mean square and merged them with the proposed MSCSDN framework that helps prolong the network lifespan. The proposed framework is compared with the state-of-the-art approaches in terms of network lifespan, stability period, instability period, report delay, report delivery, and cluster leader nodes generations. The MSCSDN achieves optimal data accuracy concerning the collected data.

KEYWORDS

internet of things, multi-hoping, similarity-based clustering, software defined networking

1 | INTRODUCTION

In recent times, the Internet of Things (IoT) has envisioned low-power and resource constraint IoT sensor nodes, which play a vital role in IoT applications [1]. One of the main characteristics of IoT is that the objects must be context-aware. Many IoT applications require that the "things" know their environmental conditions, and adjust their behaviour to them and to the other objects that are nearby. In this way, Wireless Sensor Networks (WSNs) can be considered the extension of the Internet towards the physical environment in the IoT, and thus they are one of the most valuable parts of any IoT system

[2–4]. The IoT accommodates extensive applications that include industry 4.0, advanced parking, healthcare, animal, and security monitoring applications. Those IoT applications utilise hundreds-to-thousands of IoT nodes to form a wired or wireless network to design an information collection network [5]. The deployed IoT nodes sense various network attributes, such as feeling, hearing, monitoring, and triggering an event while coordinating with the fellow participated nodes [6]. The network's collected information is forwarded to the centralised base Station (BS) through the designated routing protocol [7]. In many applications, the participated IoT nodes are equipped with the limited resource-constraint and non-rechargeable

battery, which is difficult to replace in dense and rugged environments [8]. Therefore, the design of an energy-efficient routing protocol is a critical issue in IoT applications.

In the past, several energy conversation approaches have been proposed for those applications, but most of them are insensitive to the data correlations [9, 10]. Clustering organisation of IoT nodes develops most accurate and suitable collaboration which allow data compression near the data point [11]. One of the major research challenge include the negligence of data similarity during clustering formulation. Such network attributes of data similarity can play a vital role to construct more suitable clustering formulation. The cluster leader nodes (CLNs) aggregation can be more efficient if the data are similar in each cluster [12]. This efficiency can be further improved through the central control in advanced networking settings through Software-Defined controllers. This central control acts as a network operating system that maintains the data correlation and data control functions, including the network topology. This network topology provides the physical location of IoT nodes within the data plane arrangements for deployed nodes and their respective report transmitting tasks to the wireless communication settings [13]. Simultaneously, the operations of these data planes are centrally monitored and managed in the control plane using the central SDN controller. The network management is carried out by these new primary controller interventions, enabling central network reconfiguration via SDN controller global system settings [14]. Existing work includes trend information to enhance the integrity of the cluster that lacks a huge control overhead. To repress the spatial and temporal data redundancy in the inter-cluster and intra-cluster multi-hop communication, many data aggregation schemes at the CLNs through the central controller [15]. This aggregation reduces the data size by selecting data based on its importance and omit the remaining data, which may not be recoverable at the BS [16]. These schemes struggle in terms of expensive latency for data collection, which may not be acceptable for practical implementation in real-time data collection environments.

To this end, in this paper, we propose a Multi-hop Similarity-based Clustering framework for IoT-oriented Software-Defined wireless sensor Networks (MSCSDNs). In detail, the proposed MSCSDN framework works in three folds: (1) a reliable data-aware distributed clustering technique is introduced where the clusters are created based on magnitude and trend-similarity, (2) the SDN controllers select a set of most eligible IoT nodes as the CLNs and multicast the notification to the member nodes, and (3) a multi-hop inter-cluster and intra-cluster communication begin among the selected CLNs and the neighbouring IoT nodes. By exploiting these three folds, the proposed framework collects the data at every instant without any latency. The proposed framework can be mainly effective for continuous data gathering environments and industrial control systems. The major contributions in this paper are summarised below:

- The proposed MSCSDN framework introduces the similarity-based-clustering of IoT nodes, where cluster formulation depends upon the size and trend similarity between the data series of IoT nodes. The constructed clusters are comparatively more reliable in terms of data collection.

- The SDN controllers are deployed in a distributed manner to select the most eligible CLNs for inter-cluster communication. The selection process of CLNs is guaranteed an error tolerance for minimal communication overhead.

- The proposed framework manages the heterogeneity-aware data-similar cluster formation for heterogeneous networks. The application-aware property is also introduced that enables IoT nodes to aggregate the sensed information in multiple applications.

- Extensive simulation experiments are conducted and compared with the state-of-the-art approaches that prove the significant performance gain of MSCSDN in various metrics such as network lifespan, stability period, instability period, report delay, report delivery, and CLNs generation.

The rest of the paper is organised as follows: in Section 2, we briefly explain the related work based on IoT-oriented software defined WSNs and similarity-based clustering. In Section 3, we define the overview of the proposed framework. In Section 4, we present the system models that serve as a foundation of our framework. In Section 5, we proposed the MSCSDN framework and explain the whole execution mechanism. In Section 6, we define the experimental setup and present the performance analysis in Section 7. Finally, we conclude our paper and define the future work in Section 8.

2 | RELATED WORK

In contrast to modern IoT networks, the traditional wired and wireless networks utilise the classical routing techniques that are usually based on Internet Protocol, resulting in less scalability support in the latest IoT architecture [17]. To this end, several routing mechanisms were implemented that receive tremendous acceptance to cope with the latest IoT requirements [18]. In the literature, most of the existing routing techniques can be divided into three different categories: data-centric routing, location-based routing, and hierarchical routing [19]. The hierarchical routing technique received more popularity in terms of energy efficiency; therefore, in this literature review, we only focussed on hierarchical routing. In hierarchical routing, the IoT nodes are grouped into clusters where one IoT node is selected as a CLN and acts as a relay node for communication between member nodes and the BS [20]. In particular, the participants send their sensed information to their respective CLNs. The CLNs receive the data through the data collection scheme and compress it to a single packet and then aggregate it to the centralised BS. The responsibility of member nodes is to just sense the data and forward it to CLN, whereas the CLNs face extra computational responsibilities [21]. Those responsibilities cause energy consumption and delay issues, and

the energy delay index for the trade-off (EDIT) protocol is proposed [22]. This protocol uses a back-off timer mechanism to select the CLNs using the objective function that depends on the residual energy and distance to the BS with the number of neighbour nodes. However, this protocol does not guarantee that the selected CLNs are evenly distributed in the monitoring area, leading to unbalanced energy consumption and longer time-consumption due to the re-selection of all CLNs simultaneously. To solve this issue, another routing protocol delay-constrained energy multi-hop [23] is proposed to minimise the overhead of the CLNs selection and delay. In this protocol, the CHs are selected based on the back-off timer mechanism according to the residual energy and distance to BS for the nodes. However, this protocol has the same problems as the EDIT protocol.

The existing techniques utilise numerous methods to achieve energy efficiency in IoT networks. Data aggression is a critical aspect of delivering only meaningful information towards the control station to control the efficiency of network resources [24].

Those data reduction techniques exploit temporal data correlation among the consecutive data observed by the single IoT node. In DACA [25], the instantaneous information remains critical and vital to measure the data similarity. This approach shows some success in query-based clustering schemes but DACA lacks efficiency for continuous data gathering applications.

In EEDC [26], the data similarity is determined by measuring the magnitude similarity and trend similarity among a lengthy time series's elements. This technique utilises the efficient similarity estimation; however, EEDC suffers huge communication overhead due to direct communication with the BS. In ASAP [27], data similarity indexed is computed by the correlation coefficient among the time series of sampled IoT devices which is still lacking in the data size similarity. In EAST [28], major clustering criteria is latest communicated data in the previous communication round and computes the threshold-based reference value. Temporal correlation is critical criteria, but the EAST depends upon the data aggregation instead of temporal correlation. In DSCCF [29], the distributed similarity is determined through temporal and magnitude data correlation that conducts significant data reduction and achieves marginal energy efficiency. However, this technique suffers unbalanced CLNs generation that creates latency in lengthy data exchange. In DPF [30], the authors proposed a dual prediction framework that utilises the adaptive filters-based least mean square (LMS). This technique is advantageous because of its ability to predict the data without any prior model. However, this technique suffers from identifying the appropriate filters.

In a distributed similarity estimation, the long-range data exchange results in a huge communication overhead. To overcome this communication overhead, the proposed framework estimates data similarity among the IoT nodes and respective CLNs, using only the data series's filtered version.

In this research work, we utilise the data projection with help of Euclidean distance metric to formulate the data similarity metric. We analyse that data series's projection is significantly sufficient as compared to the whole data series. This proposed method provides the more effective implementation in which the similarity estimation requires much lesser data as compared to existing solutions. In this, we significantly reduce the computation overhead. The main advantage of using this method is that fewer data is enough to hold the trend information for the lengthy data series. Also, the industrial demands for effective success in IoT networks insist the researchers consider the SDN-based solutions for flexible management of CLNs generation. The SDN-based central management in IoT networks provides massive flexibilities to exercise and develop network management and applications. As a result, the IoT nodes become more efficient in terms of reprogramming the priority operations. Therefore, the proposed framework adopts centralised control through SDN controllers for efficient CLNs selection that manages the overall network topology and perform application-aware operations.

3 | OVERVIEW OF THE PROPOSED MULTI-HOP SIMILARITY-BASED-CLUSTERING FRAMEWORK FOR THE IOT-ORIENTED SOFTWARE-DEFINED WIRELESS SENSOR NETWORKS FRAMEWORK

In order to execute the proposed framework, we briefly explain the whole methodology as follows:

- **Overview**: The network area is divided into four regions. There are four types of IoT nodes distributed in each region, such as user devices, health-related equipment, temperature measurement, and home appliances. For constructing the data similar clusters, each IoT node senses particular data using the normalized least mean square (nLMS) filter. We deploy OpenFlow switches in all regions to select the suitable CLNs among these devices for each cluster. The OpenFlow switches estimate the residual energy and measure the delay of each IoT node. The node with the highest strength and minimum delay decided to become a CLN for the current round. The member nodes measure the Euclidean distance between their data projection and the CLN's data projection and associate themselves with the data similar CLNs. Afterwards, the CLNs find a link-aware route by identifying multiple pathways using nLMS-based multi-hop routing. Finally, the route with the most vital link and minimum hops will be selected, and the CLN's forwards respective data to the central BS, where the central BS has four SDN controllers. These SDN controllers are dedicated region-wise (one for each region) to manage the network traffic and balance the load equally. The network architecture of the proposed framework is given in Figure 1.

- **Data Prediction**: The digital filters are declared to measure the node's data in the time domain. These digital filters measure the short term linearity of the signal. Also, a linear combination of data history is used in order to predict future data. For future data prediction, we use an adaptive-nLMS filter that optimally reduces the cost of data communication while maintaining robustness in the network. For each communication round, the SDN controller maintains its region's historical data, either by data prediction or by data reception.
- **Data Aggregation**: The data aggregation is carried out through lossless inter-cluster and intra-cluster multi-hop routing. In particular, each CLN compares the new data with the previous round data (gain from the regional SDN controller) and estimate the difference. Among the difference, if the value of data is greater than the given threshold, the data is aggregated through multi-hop communication. The lengthy data is aggregated through inter-cluster communication, while the short data is aggregated through intra-cluster communication.

4 | SYSTEM MODEL

In this section, we briefly explain the system models of the proposed framework.

4.1 | The network model

In the proposed framework, we create a distributed network with multi-type IoT nodes such as user devices, health-related equipment, temperature measurement, and home appliances. The whole network is divided into four regions, and in each region, we deploy the OpenFlow switch to select a suitable CLNs for each cluster. The OpenFlow switches are connected with the central SDN controllers through mesh topology. Each SDN controller is dedicated to one region to maintain the traffic and manage the load equally. Each IoT node is equipped with GPS for its own and its neighbours' location identification. The data from these IoT nodes are gathered by CLNs, which are then aggregated to the central SDN controllers
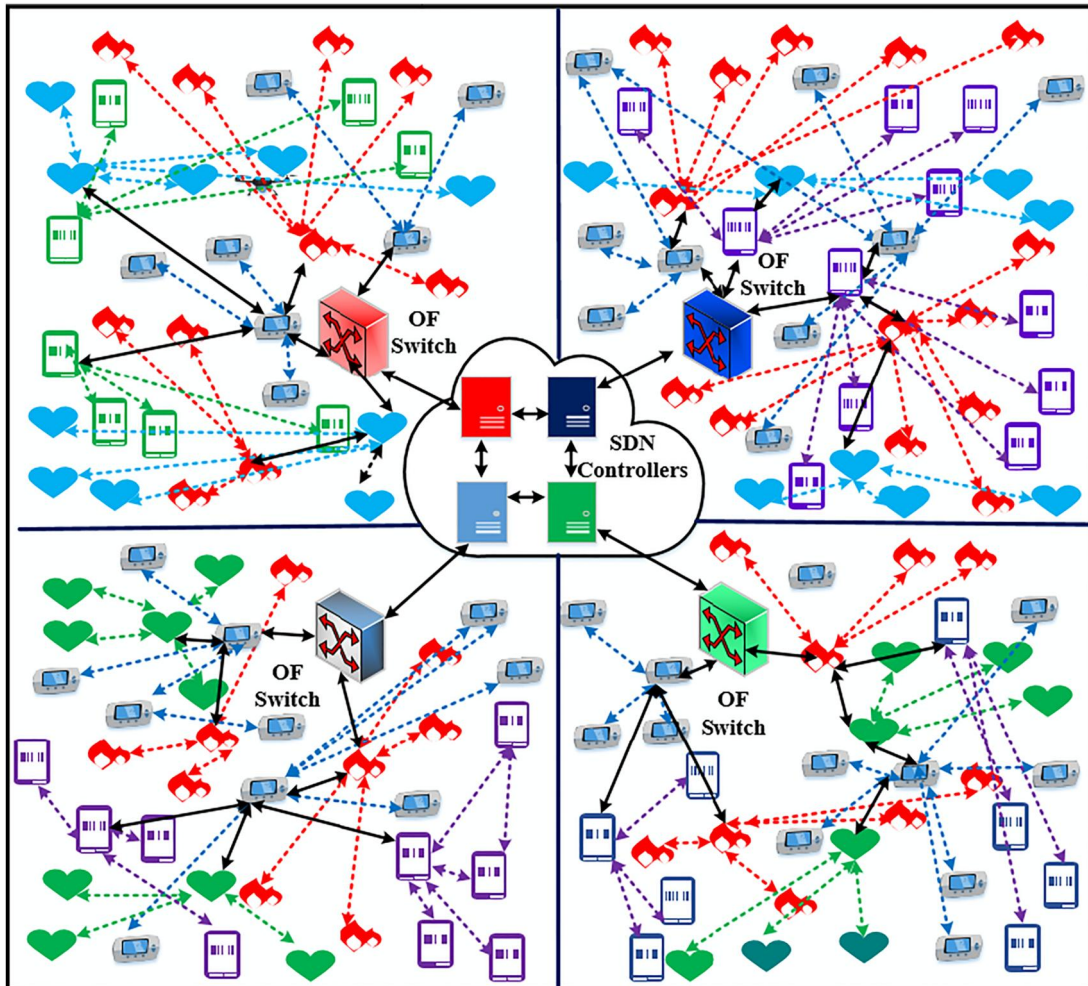

FIGURE 1 The network architecture of the proposed Multi-hop Similarity-based-Clustering framework for the IoT-oriented Software-Defined wireless sensor Networks (MSCSDNs) framework. Here, the network field is divided into four regions, and in each region, there are four types of IoT nodes distributed along with the OpenFlow switch. All the sensed data are forwarded to the base Station (BS), where the BS consists of four SDN controllers (one for each type of IoT nodes)

through lossless inter-cluster and intra-cluster multi-hop communication.

4.2 | The energy model

In the proposed MSCSDN framework, we use the log-distance path loss model (LDPLM), a well-known IoT application model [31]. The LDPLM is used with multiple log regions split at a distance R_d. The expression of this path loss at a distance d can be formalised as

$$P_{loss}(d) = \begin{cases} P_{loss}(d_x) + 10\, i \log(d) + \rho_{\omega_1} & \text{if } d \le R_d \\ P_{loss}(d_y) + 10\, j \log(d) + \rho_{\omega_2} & \text{if } d > R_d \end{cases} \tag{1}$$

where $P_{loss}(d)$ is the log region spilt at the distance R_d, and $P_{loss}(d_x)$ and $P_{loss}(d_y)$ represent the path loss from the reference distance d_x and d_y close to the transmitter. i and j are the path loss exponents and $\rho_{\omega_1}, \rho_{\omega_2}$ represents the random variable of zero-mean Gaussian with the standard deviation of ω_1 and ω_2, respectively.

The IoT nodes are adaptive to control the radius of transmission by controlling the power levels [43]. The power of the signal at the receiver-end $S_{(rx,d)}$, for a transmitted signal at power P from a distance d is expressed as

$$S_{(rx,d)} = S_{tx}(P) - P_{loss}(d). \tag{2}$$

The receiver successfully receives the data packets as

$$S_{(rx,d)} = R_s \tag{3}$$

where R_s represents the sensitivity of reception at IoT node. The dissipation energy for transmitting the data packets D_p from a distance d at power P is expressed as

$$E_{tx}(P, D_p) = REC_{tx}(P) T_{tx}(D_p) \tag{4}$$

where REC is the radio energy consumption for the total transmission (T_{tx}) of data packets (D_p) for mica2 IoT node at various energy levels as mentioned in Ref. [44]. Similarly, the dissipation energy for receiving those data packets can be computed as

$$E_{rx}(D_p) = REC_{rx} T_{tx}(D_p). \tag{5}$$

4.3 | Multi-type data and heterogeneity-aware clustering model

In IoT applications, clusters are constructed to achieve the optimal energy efficiency in the network. This efficiency is influenced by the spatial correlation between the IoT nodes [32]. In the case of similar sensed data, the transmission can be

reduced by eliminating the same information. In the proposed MSCSDN framework, we use four different types and IoT nodes: user devices, health-related equipment, temperature measurement, and home appliances to sense the heterogeneous data. Initially, the IoT nodes obtain the information from SDN controllers and measure the historical round data. Afterwards, the CLNs selection process begins, where the IoT nodes with higher energy and minimum delay will be selected as CLNs. Finally, the member nodes measure the Euclidean distance and obtain the data projection reports by comparing their data and CLNs data. If the similarity is greater than the threshold value, then the member associates themselves with the perspective CLNs. This communication between the member nodes and CLNs for multi-type data is captured in Figure 2.

5 | PROPOSAL OF THE PROPOSED MULTI-HOP SIMILARITY-BASED-CLUSTERING FRAMEWORK FOR IOT-ORIENTED SOFTWARE-DEFINED WIRELESS SENSOR NETWORKS FRAMEWORK

In this section, we explain the complete execution of the proposed framework.

5.1 | Measuring the data projection reports

In the proposed MSCSDN framework, the IoT nodes measure the data projection reports using the adaptive LMS filter [33]. The LMS filter create the samples of data stream G on the length l at an instant m. The data stream samples are denoted as $G[m]$, which computes the prediction $Q[m]$ using a linear combination from the existing samples through the corresponding weight vector $\eta[m]$.

$$Q[m] = \eta[m] \times G[m]. \tag{6}$$

The result $Q[m]$ is finally compared to the given threshold $T[m]$. The error ϖ generated during the prediction can be expressed as

$$\varpi[m] = Q[m] - T[m]. \tag{7}$$

The adaptation algorithm utilise the merger of the above computed error, which plays a key role to adjust the weights of filters of every instant m in the packet size φ to reduce the error of mean square. The adaptation of this error is expressed as

$$\eta[m + 1] = \eta[m] + \varphi G[m] \varpi[m]. \tag{8}$$

After the multiple rounds, the prediction is finalised to make aforementioned rigorous prediction and weight adaptation. The data is projected with the integration of relevant

element of the data series and the filter coefficients of LMS. The element represents the data series's current magnitude, and LMS coefficients reflect the data series trends. To estimate both data series and trend similarity, the IoT nodes calculate the distance between its data series and CLN data series.

5.2 | Selection of optimal cluster leader nodes

In general, CLNs are required to perform extra communication and computational responsibilities than normal IoT nodes [34]. Therefore, the IoT nodes with higher energy can efficiently balance the energy cost. In the proposed framework, all the IoT nodes in a particular region claim to become a CLN, and then the OpenFlow switches measure the residual energy

and latency of each IoT node in their respective region. The latency (T_d) of IoT node can be measured as

$$T_d = E\frac{T}{\eta_i} \qquad (9)$$

where T represents the total time for optimal CLN selection, E is a constant, and η_i is the weight of node i. To select optimal CLNs, each IoT node broadcasts a first message containing primary information to the OpenFlow switch as shown in Figure 3. In particular, the message contains the node ID, remaining energy, transmission power, and location. In detail, the node ID is used to identify the node, residual energy, and transmission power is used to select the most effective CLNs, and location is used to identify the neighbouring member nodes.
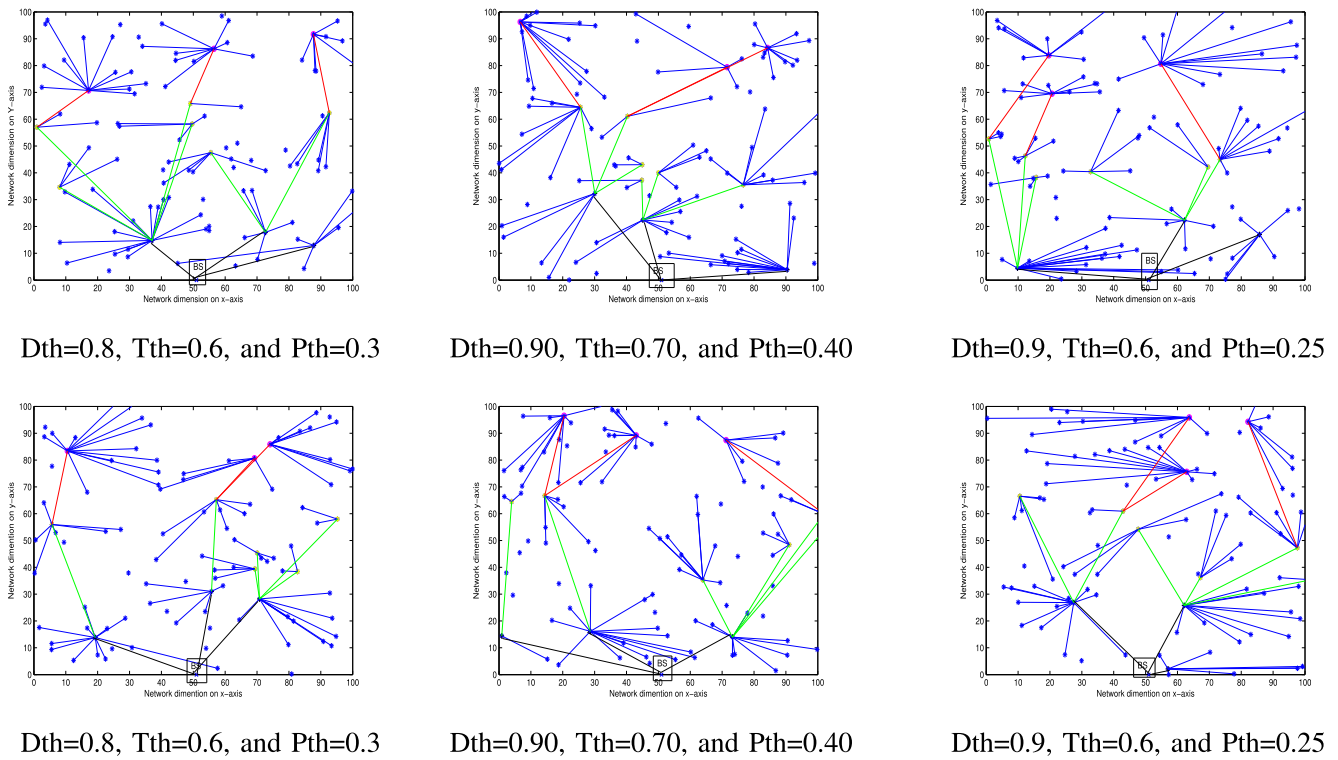

Dth=0.8, Tth=0.6, and Pth=0.3 Dth=0.90, Tth=0.70, and Pth=0.40 Dth=0.9, Tth=0.6, and Pth=0.25

Dth=0.8, Tth=0.6, and Pth=0.3 Dth=0.90, Tth=0.70, and Pth=0.40 Dth=0.9, Tth=0.6, and Pth=0.25

FIGURE 2 Multi-type data received region-wise at central SDN controllers through adaptive-nLMS inter-cluster and intra-cluster multi-hop communication in the proposed Multi-hop Similarity-based-Clustering framework for IoT-oriented Software-Defined wireless sensor Networks (MSCSDNs) framework
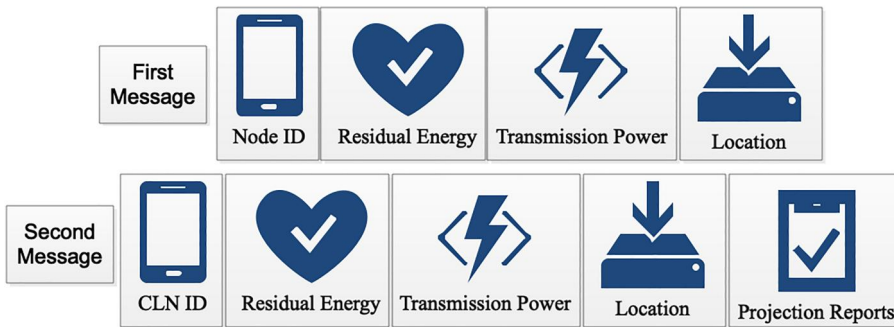

FIGURE 3 Each IoT node broadcast the first message to their respective OpenFlow switch in order to become a cluster leader node (CLN). Once the CLNs are selected, they will broadcast the second message with projection reports in order to identify the member nodes with similar data

After receiving the first message from each IoT node, the OpenFlow switch compares the remaining energy and latency of each IoT node. The node with the highest energy and minimum latency will be selected as CLN. Afterwards, selected CLN nodes broadcast the second message as shown in Figure 3. By hearing this message, the member nodes associate themselves with their respective CLNs based on data similar to projection reports. The projection report in Figure 3 is used to choose the member nodes with more similar data.

5.3 | Measuring the data similarity

The neighbouring nodes identify the CLN nodes with nearly similar data. In particular, the neighbouring nodes receive the broadcasted message from the CLNs. After the reception of these notifications, the indigenous IoT devices compare the own CLNs projection reports with arrived data projection reports. The indication of this similarity is measured through the Euclidean distance between the data projections of two-time series. If the similarity is greater than the similarity threshold, the node adds this CLN to its perspective CLN list. A brief data similarity measure is shown in Algorithm 1. The process continues until all the neighbouring nodes associate themselves with their perspective CLNs.

Algorithm 1: Association with the Data Similar CLNs

```
Input: Neighbouring IoT Nodes, CLNs ID,
       and Data Projection Reports
Output: Data Similar Association
1 Initialisation:
2 for ∀ IoT non-CLN Nodes
3 if Similarity(CLN_x) ≥ Similarity_Node_i
4   then
5 return (Associate Node_i → CLN_x)
6   else return (Find other CLN)
7 end
```

Furthermore, the proposed MSCSDN framework utilises the attraction score, where each non-CLN IoT node computes the attraction score for all the available CLNs. This attraction score is measured by the information broadcasted by CLNs, and the CLN with the highest score will become the CLN of that particular IoT node. Considering the aforementioned metrics and attraction score, the non-CLN nodes associate themselves with the most appropriate CLNs. In order to measure this attraction score, we compute

$$AS_{CLN_x} = \frac{E_{re}(CLN_x)}{E_{tx}(CLN_x)} CLN_x \in CLN_\forall \quad (10)$$

$$E_{tx}(CLN_x) = E_{elec} + E_{amp}d_{Node\ i\ to\ CLN_x}^{\alpha=2} \quad (11)$$

where AS_{CLN_x} is the attraction score for the CLN_x, E_{re} is the remaining energy of IoT node, E_{tx} is the transmission energy

for CLN, E_{elec} is the consumed energy by the radio, E_{amp} is the energy of amplifier and $d_{Node\ i\ to\ CLN_x}^{\alpha=2}$ represents the distance between IoT node i to CLN x. Once the IoT nodes associate themselves with their perspective CLNs, they send an association message containing LMS projections' information based on the nodes' recent data history. Afterwards, member nodes start sensing the environment and generate the new reports for the current running round.

5.4 | Link selection for routing

In general settings of IoT energy-efficient routing schemes consider either the direct transmission or the one-hop neighbour for transmitting the sensed data and ignores the whole path, which mostly harms the network lifetime [35–37]. The strength of the whole path lies on the weakest link, where one-hop might be more energetic while the other hops are weaker. In this case, the weaker nodes are pressurised to aggregate the data and ignores a comparatively better available alternative path [38]. In the proposed framework, we utilise link-aware routing in order to find the most suitable path for multi-hop transmission. Before going into details, below we define the terminologies used in the proposed MSCSDN's link-aware routing.

- **Transmitting node** is the one who is sending the reports to the BS.
- **Strong node** is the one who has enough energy resources for transmitting or receiving the reports for the next couple of rounds.
- **Moderate node** is the one who has enough energy resource for transmitting or receiving the reports for the current round.
- **Weak node** is the one who has not enough energy resource for transmitting or receiving the reports for the current round.

In this link-aware routing, the transmitting node is well aware of its neighbouring nodes, and the neighbouring nodes are aware of their neighbours. When a transmitting node announces for transmission, each neighbour node shares its transmission path with the transmitting node. The transmitting node compares the path strength and forwards the data to the neighbour with the strongest link.

In Figure 4, we present the link-aware routing scheme used in the proposed MSCSDN framework. The sensed information of node one can reach to the central SDN controller through 2, 5, 9, and 11 or through 3, 4, 7, and 8 or through 3, 6, 10, and 8. Considering only the neighbours' energy level, it seems obvious to choose 2, 5, 9, and 11. In this link, 9 and 11 are the weak nodes, and they may be pressured to forward the data, which results in poor latency.

In this regard, the proposed framework offers the best link path. In particular, node two offers two strong links and two weak links, whereas node three offers two routes, one with three moderate links and one strong link and the second route with
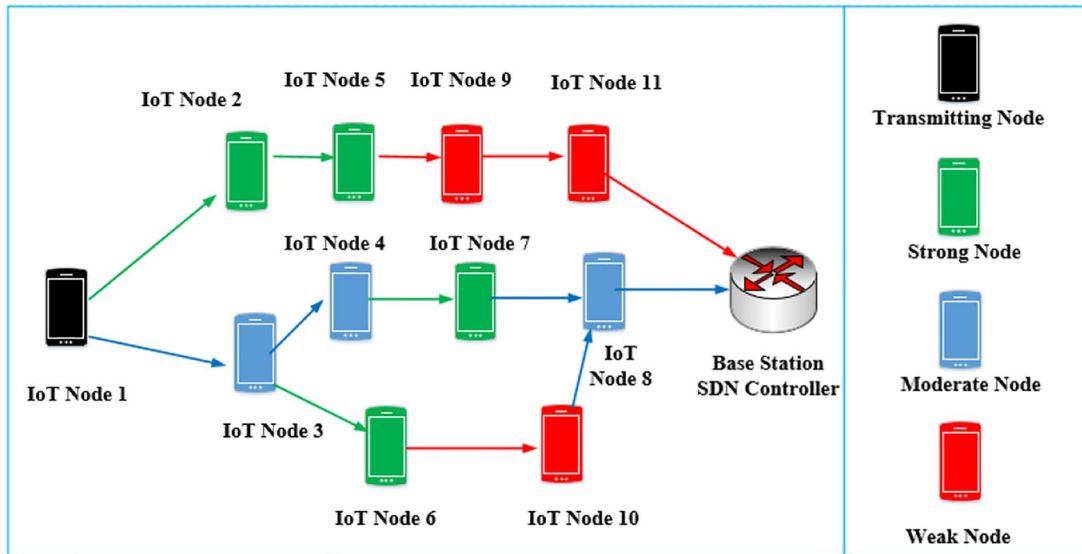
FIGURE 4   Link aware routing for multi-hop inter-cluster and intra-cluster communication

one strong link, two moderate links, and one weak link. This complete information of neighbouring nodes helps select the most suitable path, where the transmitting node chooses 3, 4, 7, and 8 route, which is comparatively better than the other route.

5.5 | Prediction-based reports generation

The proposed framework is leveraged with the prediction-based reporting for intra-cluster data sensing. The prediction-based reporting helps to prolong the lifespan of the IoT node. As the nodes are kept sensing the environment, the high sampling of data frequency leads to a temporal correlation, resulting in redundant data [39]. The proposed framework minimises this data redundancy using prediction-based reporting, where the node only transmits the subset of sensed data. Based on this transmitted data, the remaining data is predicted. In this regard, we use an adaptive nLMS prediction filter for the prediction reports. The prediction of data reports is made at the CLN and non-CLN nodes. The structure of the adaptive nLMS prediction system is given in Figure 5.

The CLN and non-CLN nodes both utilise identical filters for prediction-based reporting. At each instant m of a data sample, the CLN-members transmits the sensed reports to the CLN in the initial mode. At the same time, the source prediction engine updates its coefficients to converge based on variance. If this variance is less than the error threshold e_t, then the prediction is converged for T consecutive predictions. Afterwards, the CLN-member nodes switch their status to the standalone mode and communicate only the prediction model with their respective CLNs. In this mode, at each instant m of the data sample, the CLNs and non-CLNs predict the reports using the prediction model. To verify the prediction reports, CLNs compare the model's predicted report with the actual sensed report. For each instant m, the CLN consider the predicted report as e_t approximation of the CLN-member nodes.
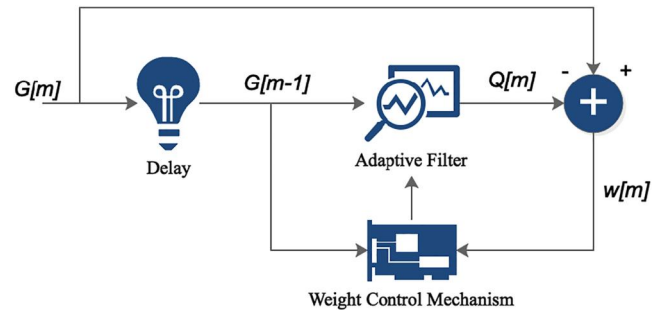

FIGURE 5   Adaptive normalized least mean square (nLMS) based prediction filter

The current trend cannot be repeated consistently by a model built using historical data as communication pattern show fluctuation over the network operation. Based on the particular higher threshold of prediction deviates e_t, the IoT devices update the prediction deviates more than the e_t periodically. Otherwise, the CLN-member node switches from standalone to the normal mode. During current communication mode, the sensed reports are delivered to CLNs. After that, the prediction model CLNs adjusts the weight to comprehend the prediction report with on-demand value. Hence, after the current prediction, the CLN-member diverts to the standalone mode. After receiving the prediction reports, the CLNs aggregate these reports to the local OpenFlow switch, which is responsible for forwarding them to the central SDN controllers.

6 | EXPERIMENTAL SETUP

In this section, we conduct extensive simulation experiments to evaluate the performance of the proposed Multi-hop Similarity-based Clustering for IoT Oriented Software-

Defined WSNs (MSCSDNs). The main reason to implement MSCSDN is to obtain energy efficiency in a highly dense network. In order to evaluate the performance, we consider the state-of-the-art approaches such as DACA [25], EEDC [26], ASAP [27], EAST [28], and DSCCF [29]. In all of our experiments, the proposed framework outperforms the existing techniques with marginal delay. All of our experiments are conducted in a MATLAB environment [40]. To investigate the performance, we consider a large set of system parameters where we deploy the various number of nodes in different sizes of network dimensions. The considered metrics for evaluation are network lifespan, report delivery to CLNs and SDN controllers, end-to-end delay, packet delivery ratio, and CLN generations. The benchmark system parameters are defined in Table 1.

7 | PERFORMANCE ANALYSIS

In this subsections, we provide a detailed performance analysis on aforementioned metrics.

7.1 | Network lifespan

The most important metric to measure the energy-efficient data collection system is the network lifespan. Two terms can measure the network lifespan: the overall network lifetime and network stability. The network lifetime counts the total number of communication rounds until the last IoT node dies. In contrast, the network stability counts the total number of communication rounds until the first IoT node's death. The proposed framework extends network efficiency by reducing the number of data per communication. We deploy 100 IoT nodes in the network dimensions of $100 \times 100m$ and run the experiments for 5000 communication rounds in experiments. The rest of the system parameters are the same as defined in Table 1. Figure 6 shows the network lifetime and network stability concerning the number of alive and dead nodes for the

TABLE 1 Parameters used in simulation

Parameter	Value
Network size	100^2, 250^2, and 500^2
Number of nodes	100, 200, 300, and 500
Adaptive nLMS length	5
Initial energy	0.5, 1.0, and 2.5
P	0.1
Data aggregation energy cost	50pj/bit j
Packet size	200 bit
Transmitter electronics E_{TX}	50nj/bit
Receiver electronice E_{RX}	50nj/bit
Transmit amplifier E_{fs}	10pj/bit/m^2

communication rounds, respectively. In particular, DACA, ASAP, DSCCF lose their IoT nodes in the early communication rounds, while EEDC and EAST try to compete with the proposed MSCSDN for a longer period. The first graph of Figure 6 shows the network lifetime. The proposed MSCSDN has lower performance than EAST in early communication rounds but outperforms EAST in later communication rounds and keeps alive until 4457 rounds, where EAST dies after around 3200 rounds. The remaining schemes DACA, EEDC, ASAP and DSCCF die after 1300, 2980, 2890, 3021 rounds, respectively.

Similarly, in the second graph of Figure 6, the proposed scheme outperforms the existing schemes in terms of network stability. In particular, the proposed scheme has better network stability than DACA, EEDC, ASAP, and DSCCF, where it secure at least 2% more stability than other schemes. The EAST performs better in terms of network stability due to the centralised data collection and static nature of deployed IoT nodes.

7.2 | Reports delivery to cluster leader nodes and SDN controllers

As the proposed framework exploits the temporal correlation in the sensed reports, therefore the inter-cluster and intra-cluster communication cost is reduced. We extend our experiments and measure the communicated number of data reports from member-CLNs to CLNs and CLNs to SDN controllers to prove this performance. In Figure 7, we present the number of transmitted reports to CLNs and SDN controllers, respectively. In particular, in the first graph of Figure 7, the proposed framework obtains 40%, 25%, 35%, 30% and 20% better reports delivery to CLNs than DACA, EEDC, ASAP, EAST, and DSCCF, respectively. In DSCCF, the authors also use the temporal correlation and used compressed forwarding to achieve better performance. Despite the compressed forwarding, the proposed framework MSCSDN beats the DSCCF. In the case of delivery to SDN controllers, the proposed framework is leveraged with the OpenFlow switches that maintain the highest delivery compared with existing approaches. The second graph of Figure 7 shows the number of reports delivered to SDN controllers. It can be seen that the proposed framework achieves better performance due to regional OpenFlow switches in the network.

7.3 | Delay and report delivery ratio

The reliability of an energy-efficient system depends on the latency. In the proposed work, we use nLMS system to minimise this latency. For this evaluation, we compare the end-to-end delay and report delivery ratio. Here, we consider DACA, ASAP, and DSCCF for the comparison with the proposed MSCSDN. In Figure 8, we present the end-to-end delay and report delivery ratio, respectively. The results in Figure 8 prove
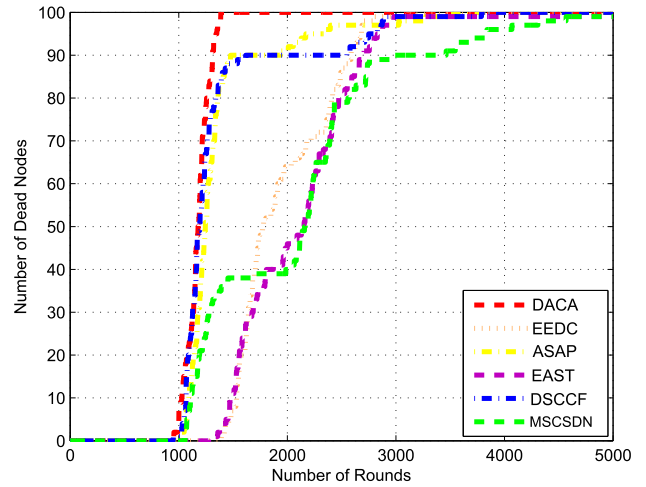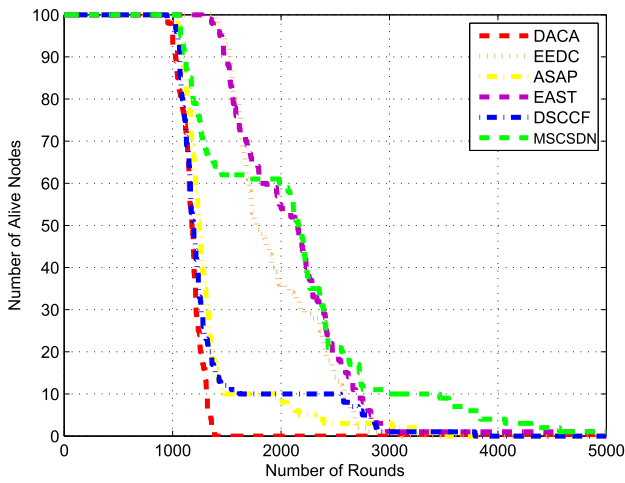
FIGURE 6 Network lifetime and network stability, where we deploy 100 IoT nodes in 100 × 100 network dimensions with the initial energy $E_o = 0.5$ in graph 1 and 2, respectively
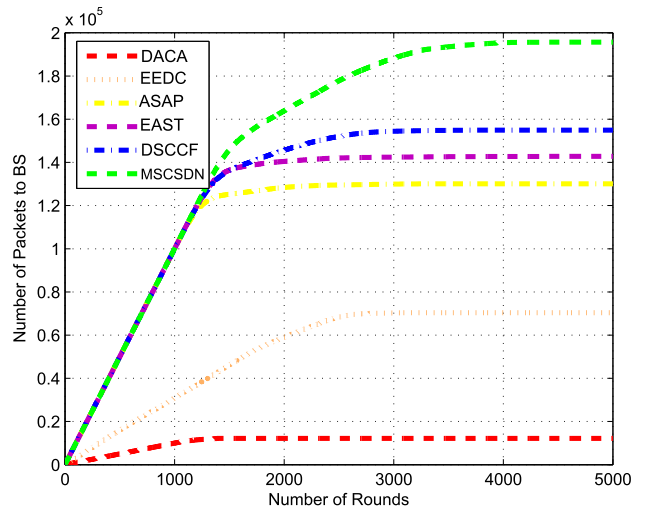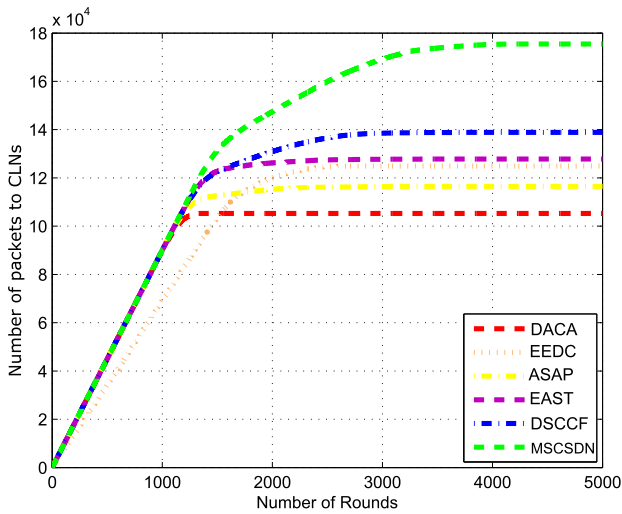

FIGURE 7 Number of reports delivered to cluster leader nodes (CLNs) and centralised SDN controllers, where we deploy 100 IoT nodes in 100 × 100 network dimensions with the initial energy $E_o = 0.5$ in graph 1 and 2, respectively
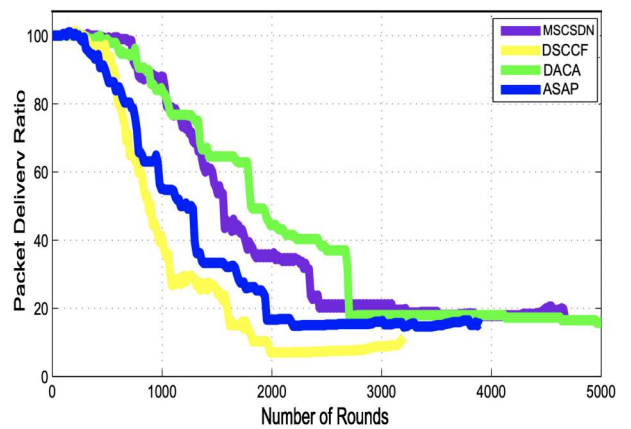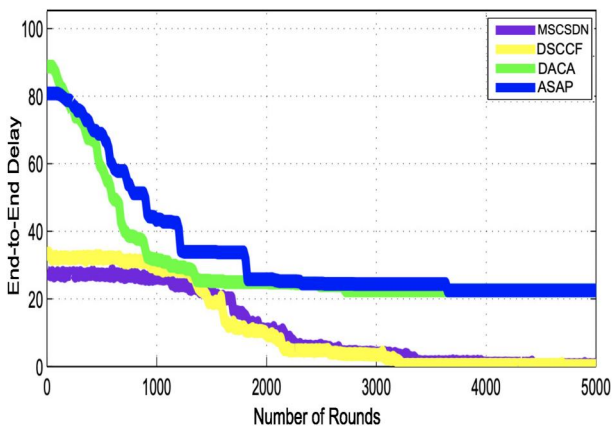

FIGURE 8 End-to-End Delay and Packet Delivery Ratio where, $N = 100$ deployed in $ND = 100 \times 100$ with the initial energy $E_o = 0.5$ in graph 1 and 2, respectively

that the proposed framework has a minimum delay and maximum delivery ratio.

7.4 | Large scale network

To prove the performance in large scale networks, here we consider huge network dimensions with a more dense deployment of IoT nodes. In particular, we consider two scenarios for dense deployment. In scenario one, we consider the $250 \times 250m$ size of network area and deploy 200 IoT nodes and run the network simulations for 7000 communication rounds with the initial energy of 1.0. Figure 9 shows the energy efficiency in terms of network lifetime and network stability in scenario one. The graph in Figure 9 demonstrates that the proposed framework works smoothly and achieves higher energy efficiency in a more extensive network than the existing approaches. In particular, the IoT nodes in the

proposed framework are still alive and communicating until the 6800 communication rounds, while DACA, EEDC, ASAP, EAST, and DSCCF die after 1921, 5129, 4260, 6341, 5212 communication rounds, respectively.

In the second scenario, we further extend the network dimensions to $500 \times 500m$ network area and deploy 300 IoT nodes for a more dense network and run the simulations for 10,000 communication rounds with the initial energy of 2.5. Figure 10 shows the simulation results in terms of network lifetime and network stability for the second scenario. The graph in Figure 9 proves that the proposed framework achieves higher efficiency than the other approaches. In particular, the IoT nodes are still active until the end of 9103 communication rounds, while the existing approaches DACA EEDC, ASAP, EAST, and DSCCF are active only until 3811, 7460, 6102, 8200, 4921 communication rounds. We attribute this higher efficiency to the regional deployment of OpenFlow switches, similarity-based clustering and the prediction-based reporting
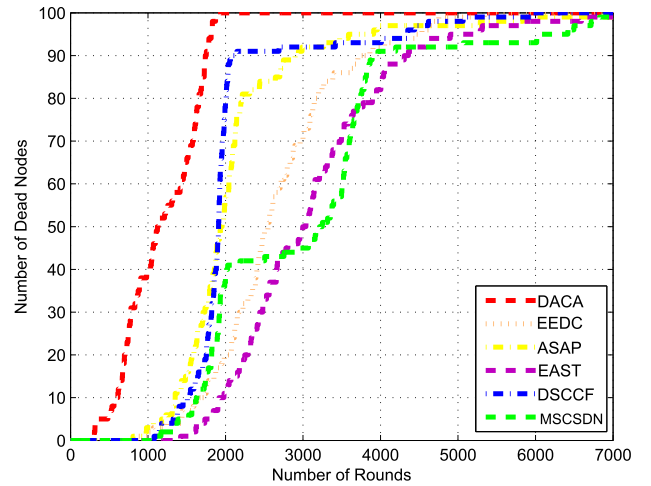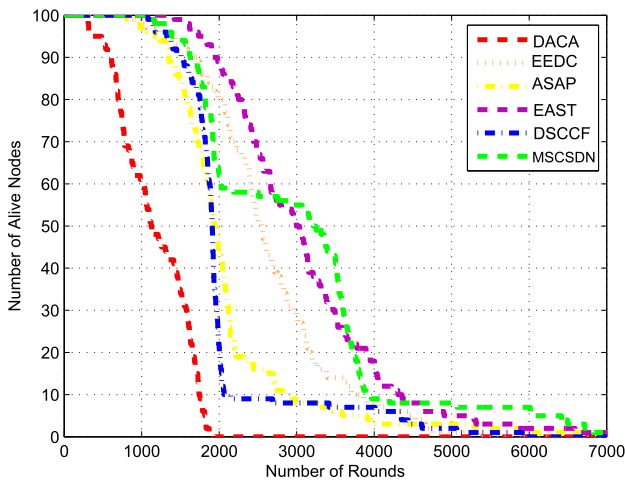

FIGURE 9 Network lifetime and network stability, where we deploy 200 IoT nodes in 250×250 network dimensions with the initial energy $E_o = 1.0$ in graph 1 and 2, respectively
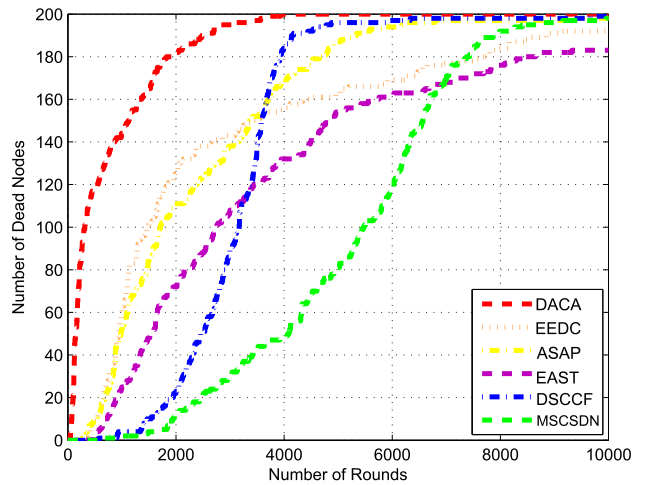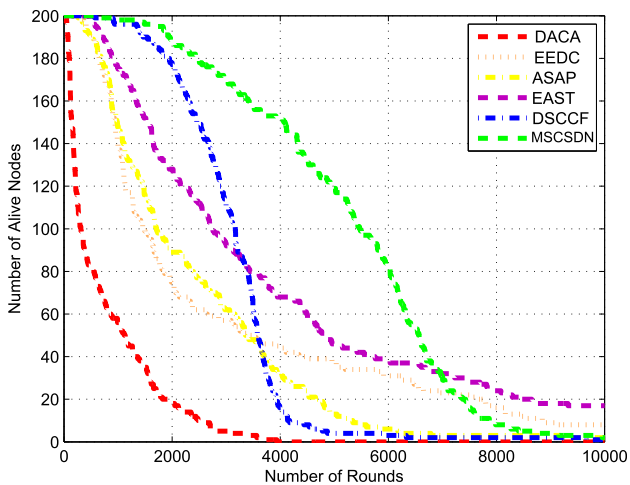

FIGURE 10 Network lifetime and network stability, where we deploy 300 IoT nodes in 500×500 network dimensions with the initial energy $E_o = 2.5$ in graph 1 and 2, respectively

which reduces the communication load from the IoT nodes. Further, the adaptive nLMS system also helps minimise the communication load through link aware routing.

7.5 | Generation of cluster leader nodes

In Section 3.5, we show the optimal selection of CLNs in the proposed framework. Here, we prove the efficiency of this selection criteria in comparison with existing approaches. We consider the approaches mentioned above and compare the CLNs generation in three different scenarios. In particular, we consider three sizes of network area {100 × 100, 250 × 250 and 500 × 500}, where we deploy 100, 200 and 300 IoT nodes and run the simulation for 5000, 7000, and 10,000 communication rounds with the initial energy of 0.5, 1.0, and 2.5, respectively. The graphs in Figure 11 show the CLNs generations for each scenario, respectively. The x-axis indicates the fluctuation of CLNs generation, while the y-axis represents the latency period per round. It is demonstrated clearly that the proposed framework MSCSDN has more fluctuation with minimum latency than the existing approaches. We attribute this higher performance to the involvement of OpenFlow switches to select the optimal IoT nodes for CLNs operations.

To remove the noise from the above graphs, we show the number of CLNs generated per round of all three scenarios in Figure 12 for MSCSDN. In particular, the box plot in Figure 12 shows the number of CLNs, where the blue line in each box shows the average number of CLNs generated in a particular round.

7.6 | Lifespan in heterogeneous settings

We further prove the network performance in heterogeneous settings. Here, we consider network lifetime, network stability, and network instability to prove the performance of the proposed framework in heterogeneous settings. In particular, we set the initial energy to several levels, such as 3, 4, 5, 6, 7, 8, 9, and 10, and measure the network performance for a huge number of communication rounds. Here, we choose the DSCCF approach for comparison as the said approach has the same criteria for cluster creation. The only difference is that DSCCF is deployed with normal sensor nodes while the proposed framework is leveraged with smart IoT nodes and regional OpenFlow switches with central SDN controllers. Figure 13 shows the performance comparison in terms of network lifetime, network stability, and network instability period concerning multiple heterogeneity levels. It is demonstrated that the proposed framework has a great advantage of Open switches and central SDN controllers that helps to maintain higher network performance.
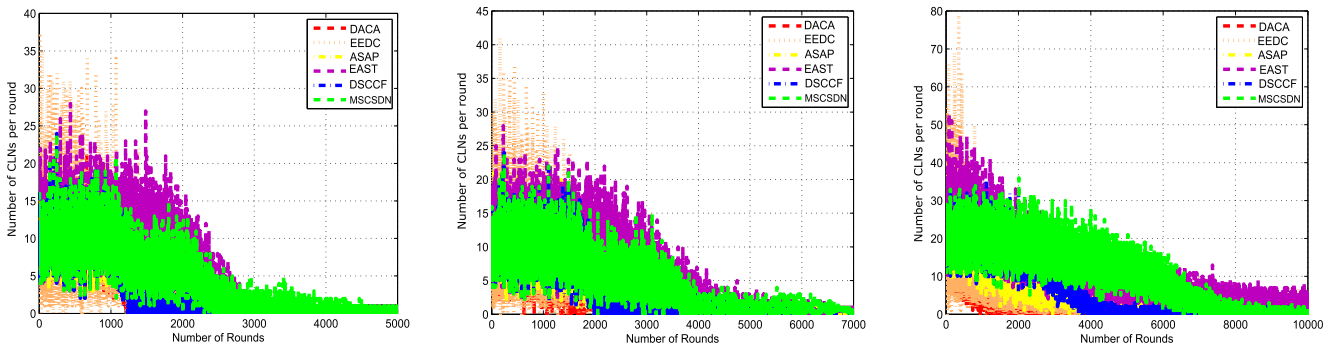

FIGURE 11 CLNs generation in three different scenarios where, we deploy {100, 200, 300} IoT nodes in the network dimensions of {100 × 100, 250 × 250, 500 × 500} with the initial energy of {0.5, 1.0, 2.5} in graph 1, 2 and 3, respectively
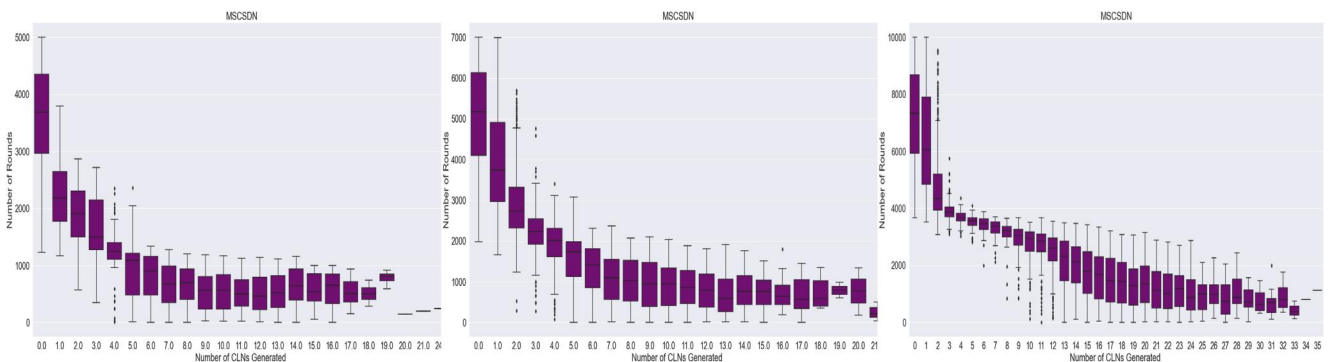

FIGURE 12 CLNs generations of Multi-hop Similarity-based-Clustering framework for IoT-oriented Software-Defined wireless sensor Networks (MSCSDN) in three different scenarios where, we deploy {100, 200, 300} IoT nodes in the network dimensions of {100 × 100, 250 × 250, 500 × 500} with the initial energy of {0.5, 1.0, 2.5} in graph 1, 2 and 3, respectively
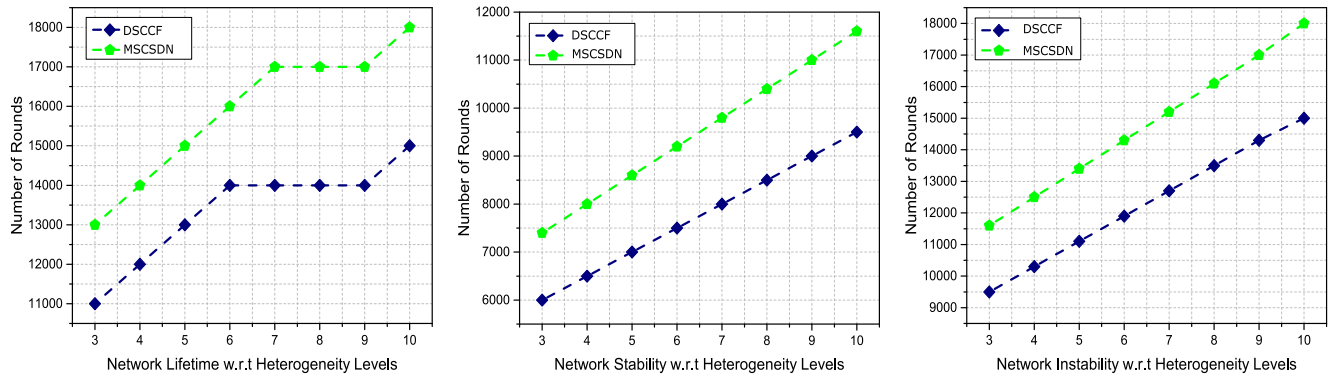
FIGURE 13 Network Lifetime, Stability and Instability period, where we deploy 500 IoT nodes in the network dimensions of 500 × 500 with different value of heterogeneity levels in graph 1, 2 and 3, respectively

8 | CONCLUSION

This paper proposes a novel multi-hop similarity-based-clustering for IoT-oriented software defined WSNs (MSCSDN) framework. The network is leveraged with regional OpenFlow switches which are directly connected to the four centralised SDN controllers. The CLN selection and member association is based on similar data that reduce the clustering overhead. The prediction based reporting minimises the communication overhead. Further, the adaptive nLMS- based route decision minimises the inter-cluster and intra-cluster communication cost. The experimental evaluation on various network settings with multiple system parameters prove that the proposed framework performs better than the existing solutions in terms of energy efficiency. In the future work, we plan to extend the proposed framework to reduce data reports size through compressed forwarding.

CONFLICTS OF INTEREST
The authors declares no conflict of interest regarding publication of this research article.

DATA AVAILABILITY STATEMENT
The authors were unable to find a valid data repository for the data used in this study. These data are available from [aye-shashafique] at [njust.edu.cn]".

ORCID
Muhammad Asad https://orcid.org/0000-0003-0036-1714

REFERENCES

1. Pirbhulal, S., et al.: A novel secure IoT-based smart home automation system using a wireless sensor network. Sensors. 17(1), 69 (2017)
2. Maheswari, A., Yadav, R.K., Nath, P.: Data congestion prediction in sensors based IoT network. J. Sci. Ind. Res. 80(12), 1091–1095 (2022)
3. Sankari, B.S., Nemani, R.: Squirrel search algorithm based support vector machine for congestion control in WSN-IoT. Wireless. Pers. Commun., 1–16 (2022)
4. Piran, M.J., et al.: Energy-efficient transmission range optimization model for WSN-based Internet of things. Comput. Mater. Continua (CMC). 67(3), 2989–3007 (2021)
5. Scarselli, F., et al.: The graph neural network model. IEEE Trans. Neural. Network. 20(1), 61–80 (2008)
6. Raj, J.S., Basar, A.: QoS optimization of energy efficient routing in IoT wireless sensor networks. J. ISMAC. 1(01), 12–23 (2019)
7. Asad, M., et al.: IoT enabled adaptive clustering based energy efficient routing protocol for wireless sensor networks. Int. J. Ad Hoc Ubiquitous. Comput. 32(2), 133–145 (2019)
8. Nguyen, N., Nguyen, Q.C., Le, M.T.: A novel autonomous wireless sensor node for IoT applications. Telkomnika. 17(5) (2019)
9. Saad, G., et al.: Correlation-based sensor activity scheduling mechanisms for wireless sensor networks. In: 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), pp. 1–8. IEEE (2019)
10. Altinel, D., Kurt, G.K.: Modeling of multiple energy sources for hybrid energy harvesting IoT systems. IEEE Internet Things J. 6(6), 10846–10854 (2019)
11. Basavaraj, G.N., Jaidhar, C.D.: Low Latency and Energy Efficient Cluster Based Routing Design for Wireless Sensor Network (2019)
12. Diwakaran, S., Perumal, B., Devi, K.V.: A cluster prediction model-based data collection for energy efficient wireless sensor network. J. Super-comput. 75(6), 3302–3316 (2019)
13. Ding, Z., et al.: Energy-efficient relay-selection-based dynamic routing algorithm for IoT-oriented software-defined WSNs. IEEE Internet Things J. 7(9), 9050–9065 (2020)
14. Theodorou, T., Mamatas, L.: CORAL-SDN: a software-defined networking solution for the Internet of Things. In: 2017 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), pp. 1–2. IEEE (2017)
15. Kobo, H.I., Abu-Mahfouz, A.M., Hancke, G.P.: A survey on software-defined wireless sensor networks: challenges and design requirements. IEEE Access. 5, 1872–1899 (2017)
16. Banerjee, A., Hussain, D.M.: SD-EAR: energy aware routing in software defined wireless sensor networks. Appl. Sci. 8(7), 1013 (2018)
17. Younis, O., Fahmy, S.: HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Trans. Mobile. Comput. 3(4), 366–379 (2004)
18. Qiu, M., et al.: Informer homed routing fault tolerance mechanism for wireless sensor networks. J. Syst. Architect. 59(4-5), 260–270 (2013)
19. Lotf, J.J., Hosseinzadeh, M., Alguliev, R.M.: Hierarchical routing in wireless sensor networks: a survey. In: 2010 2nd International Conference on Computer Engineering and Technology, Vol. 3, pp. V3-650. IEEE (2010)

20. Abbasi, A.A., Younis, M.: A survey on clustering algorithms for wireless sensor networks. Comput. Commun. 30(14-15), 2826–2841 (2007)

21. Al-Fuqaha, A., et al.: Internet of things: a survey on enabling technologies, protocols, and applications. IEEE Commun. Surv. Tutor. 17(4), 2347–2376 (2015)

22. Thakkar, A., Kotecha, K.: Cluster head election for energy and delay constraint applications of wireless sensor network. IEEE Sensor. J. 14(8), 2658–2664 (2014)

23. Huynh, T.T., Dinh-Duc, A.V., Tran, C.H.: Delay-constrained energy-efficient cluster-based multi-hop routing in wireless sensor networks. J. Commun. Network. 18(4), 580–588 (2016)

24. Dias, G.M., Bellalta, B., Oechsner, S.: A survey about prediction-based data reduction in wireless sensor networks. ACM Comput. Surv. 49(3), 1–35 (2016)

25. Bahrami, S., Yousefi, H., Movaghar, A.: DACA: data-aware clustering and aggregation in query-driven wireless sensor networks. In: 2012 21st International Conference on Computer Communications and Networks (ICCCN), pp. 1–7. IEEE (2012)

26. Liu, C., Wu, K., Pei, J.: An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation. IEEE Trans. Parallel Distr. Syst. 18(7), 1010–1023 (2007)

27. Gedik, B., Liu, L., Philip, S.Y.: ASAP: an adaptive sampling approach to data collection in sensor networks. IEEE Trans. Parallel Distr. Syst. 18(12), 1766–1783 (2007)

28. Villas, L.A., et al.: An energy-aware spatio-temporal correlation mechanism to perform efficient data collection in wireless sensor networks. Comput. Commun. 36(9), 1054–1066 (2013)

29. Arunraja, M., Malathi, V., Sakthivel, E.: Distributed similarity based clustering and compressed forwarding for wireless sensor networks. ISA Trans. 59, 180–192 (2015)

30. Santini, S., Romer, K.: An adaptive strategy for quality-based data reduction in wireless sensor networks. In: Proceedings of the 3rd International Conference on Networked Sensing Systems (INSS 2006), pp. 29–36. TRF, Chicago (2006)

31. Jung, S., Lee, C.O., Han, D.: Wi-Fi fingerprint-based approaches following log-distance path loss model for indoor positioning. In: 2011 IEEE MTT-S International Microwave Workshop Series on Intelligent Radio for Future Personal Terminals, pp. 1-2. IEEE (2011)

32. Singh, M., Soni, S.K.: Fuzzy based novel clustering technique by exploiting spatial correlation in wireless sensor network. J. Ambient Intell. Hum. Comput. 10(4), 1361–1378 (2019)

33. Xiong, K., Wang, S.: Robust least mean logarithmic square adaptive filtering algorithms. J. Franklin Inst. 356(1), 654–674 (2019)

34. Mehta, D., Saxena, S.: MCH-EOR: multi-objective cluster head based energy-aware optimized routing algorithm in wireless sensor networks. Sustain. Comput: Inform. Syst. 28, 100406 (2020)

35. Cao, X., et al.: Towards energy-efficient wireless networking in the big data era: a survey. IEEE Commun. Surv. Tutor. 20(1), 303–332 (2017)

36. Ejaz, W., et al.: Efficient wireless power transfer in software-defined wireless sensor networks. IEEE Sensor. J. 16(20), 7409–7420 (2016)

37. Jan, B., et al.: Energy Efficient Hierarchical Clustering Approaches in Wireless Sensor Networks: A Survey. Wireless Communications and Mobile Computing. (2017)

38. Lai, X., et al.: Energy efficient link-delay aware routing in wireless sensor networks. IEEE Sensor. J. 18(2), 837–848 (2017)

39. Arunraja, M., Malathi, V.: Collective prediction exploiting spatio temporal correlation (CoPeST) for energy efficient wireless sensor networks. TIIS. 9(7), 2488–2511 (2015)

40. Higham, D.J., Higham, N.J.: MATLAB Guide. Society for Industrial and Applied Mathematics (2016)