Hindawi
Mobile Information Systems
Volume 2022, Article ID 4681583, 15 pages
https://doi.org/10.1155/2022/4681583

Research Article

Internet of Things-Enabled Optimal Data Aggregation Approach for the Intelligent Surveillance Systems

Mohammad Khalid Imam Rahmani,[1] **Fazlullah Khan**,[2] **Abdul Wahab Muzaffar**,[1] **and Mian Ahmad Jan**[2]

[1]*College of Computing and Informatics, Saudi Electronic University, Riyadh, Saudi Arabia*
[2]*Department of Computer Science, Abdul Wali Khan University Mardan, Mardan, Pakistan*

Correspondence should be addressed to Mohammad Khalid Imam Rahmani; m.rahmani@seu.edu.sa

Received 15 March 2022; Revised 8 June 2022; Accepted 25 June 2022; Published 10 August 2022

Academic Editor: Hasan Ali Khattak

The Internet of Things (IoT)-based intelligent surveillance systems in smart cities are a challenging issue as various devices capture the data. These devices, deployed close to the underlined phenomenon, such as cameras, are duplicated or redundant as accuracy is the main requirement of these systems. For this purpose, sensor nodes are deployed to provide 24/7 monitoring of a smart city, which minimizes the security risks and enables quick response in case of any disaster. However, due to a large number of devices, huge data are generated; thus, controlling traffic congestion in case of undesirable circumstances, for example, in case of accident or intention blockage of the road, is desperately needed. Numerous data aggregation mechanisms were reported in the literature to address this issue with smart city surveillance systems. However, these approaches were designed for either specific application environments or complex environments, making the implementation process hard. In this article, we have developed an Internet of Things-enabled optimal data aggregation approach, specifically designed for the intelligent surveillance systems in smart cities, to convert raw data values into the refined ones with minimum possible data loss ratio. Moreover, the proposed scheme bounds every server to perform or carry out the data refinement process to maintain the expected ratio of accuracy and precision. In this approach, ordinary devices are forced to capture and forward data in raw form, preferably without any or minimum possible processing. This reduces the load on ordinary devices in intelligent surveillance systems. Additionally, we have developed a novel approach to eliminate or reduce (if elimination is not possible) noisy data or outliers. It is implemented along with existing state-of-the-art techniques to verify the exceptional performance of the proposed data aggregation approach. These algorithms were compared using various performance evaluation metrics such as refinement ratio, data loss ratio, energy efficiency, and lifetime. The simulation results have verified that the proposed scheme's performance is better than the existing approaches.

1. Introduction

The world was struck by heart wrenching and unfortunate incident of 9/11 that led to an unending terrorism wave around the globe. Although this problem needs a political solution, technology has to go hand in hand to implement policies sought by world leaders. Therefore, the research community is trying to ensure the safety and security of humans using intelligent surveillance systems [1]. As a result, the smart city concept is initiated to control and monitor various challenges such as traffic congestion, safety, and security. However, studies focusing on the identification of suspicious activities have been limited [2]. Therefore, this project aims at building a smart city model that monitors human behavior and any doubtful movements. Different features are extracted from the data collected through simulations. Every suspicious activity will be reported to the concerned security agencies using an alarm system, and the related data will be stored in a centralized database for subsequent processing. The concept of a smart city is mostly deployed for overcoming the crisis raised due to the massive migration of people to the urban areas. In urban areas, citizens are efficiently utilizing the resources from different perspectives. These include a clean and sustainable environment, robust and secured public mobility and transportation, IT connectivity and digitalization, e-governance, health, and education in line with the Vision 2030 of the Kingdom [3]. Smart cities have gained worldwide acceptance and have been deployed for various applications such as waste management, traffic management, water management, resource utilization, smart parking, healthcare, and supply chain and control. These applications are deployed mostly in technologically advanced countries with sufficient resources to make the concept of smart cities a reality. The Obama administration, in Nov 2015, took the initiative of making the cities smarter for the sake of controlling and monitoring various challenges such as traffic congestion, climate change, and improving the economy and services to the citizens [4]. The concept of a smart city is incomplete without extracting useful and distinctive patterns from the collected data for countering terrorism and controlling crime by the concerned agencies. Unfortunately, few studies have been dedicated to investigating and implementing the security constraints in smart cities, for example, in London and Singapore, where initiatives were taken to predict and identify any illegal and suspicious activity. However, monitoring a person's suspicious movement and activities is very scarce in the literature, especially in the middle east and Asia. The European Innovation Partnership on Smart Cities and Communities (EIP-SCC) has envisioned a Strategic Implementation Plan in 2013 that is intended to accelerate the industrial deployment of smart cities on a very large scale for energy management, transportation, waste management, and other applications. After 9/11 and July 2005 London underground blasts, the UK government brought dramatic changes to the existing infrastructure by deploying sensors. These were initially intended for surveillance but were later used to monitor the lifts, escalators, and heating, ventilation, and air conditioning (HVAC) controls to closed-circuit video and communication systems. In the Spanish city of Port of Santander, various sensing devices have been deployed in streetlights and facades to observe the temperature, noise, crowd detection, and air pollution emissions [5]. In Padova, Italy, numerous multimedia sensor devices were deployed on street light poles to observe the light intensity and precise operation of these poles. These sensors gather data about air temperature, humidity, vibration, CO_2 level, noise, etc. To evaluate security shortcomings and noise at outdoor events, the European Commission focused on wearables for smart ecosystems. The "MONICA" project uses various wearable devices and connected sensors and actuators for communication with a cloud-based infrastructure. These wearables, sensors, and actuators provide data about a targeted region for developing security applications to prevent undesirable situations and events. Moreover, MONICA provides visual and haptic clues on smartphones and wristbands that guide the masses to safer locations in an emergency. In Bremerhaven, Germany, the researchers have been working on deploying dynamic access points to facilitate the tourists and locals in the form of Internet connectivity. The ubiquitous connectivity to the people helps the realm of the smart city concept as all the services are readily available. In Melbourne, Australia, smart parking has been the core focus of the local government in the context of smart cities. For example, the buses are connected to the streetlights that provide real-time information about the location of the buses. Moreover, the proposed smart city application detects the occupancy status of each parking lot and directs the drivers to any available space in the lot. Furthermore, some similar projects have been carried out in the existing literature. Therefore, in this project, we are aiming to design a smart city model for the urban areas of KSA. The proposed model will be validated using simulation to monitor human behavior, body language, and doubtful movements. The activities of citizens will be monitored using simulation in a particular area. In the simulation, various types of data such as human behavior, body language, and any doubtful movements will be generated in a number of scenarios to consider different factors in a specified time. We will apply machine learning algorithms to the generated data to extract valuable information. In this context, only alarming data will be transferred to the database for further processing and to the government agencies for immediate investigation and action to control the crimes. In the Internet of Things (IoT), especially in smart cities, a large number of devices C_i are deployed to capture data (preferably in raw form) by directly interacting with the environment through sensors [6, 7]. These devices $C - i$ are forced to forward it to a centralized server or cluster head S_j for onward processing, and analysis [8]. In these networks, member devices C_i are deployed in the concerned area. It is highly likely that these devices, specifically those deployed in the vicinity, generate spatially and temporally correlated data streams or values that need to be processed before transmission activity is initiated as the transmission is assumed to be more power-hungry than processing activity [9, 10]. Therefore, researchers have focused on the idea of minimizing the transmission ratio of those packets that are redundant. It is because every entity of the underlined IoT network suffers heavily if it has to process and transmit these streams (preferably duplicate ones). Additionally, forwarding or transmitting these data streams (specifically redundant) exposes the IoT network to numerous possible problems such as depletion of energy, bandwidth utilization, and communication overheads [11]. Additionally, in the IoT networks, resource-constrained and miniature devices C_i are severely affected by various activities carried out by the concerned device C_i, that is, sharing of data, processing, and sensing (capturing data values by interacting with the environment directly). Among these operations, data transmission consumes approximately more energy than other operations. Generally, the transmission of duplicate data (preferably higher volumes of raw redundant streams) results in rapid depletion of energy of these devices C_i. Therefore, limited or constrained energy of member devices C_i and approximately higher consumption of energy associated with communication lead to an imbalance in the IoT network, which is designed for the smart cities [12, 13].

Data aggregation or fusion is one of the common approaches that are used to resolve the aforementioned challenging issue that is tightly coupled with the resource-limited networks in general, and IoT in particular [14–16]. Data aggregation or fusion approach refers to the local in-network data processing specifically from different source devices C_i. In contrast, these data are needed to be transmitted to a centralized server module S_j for onward processing. These approaches are designed to manage data redundancy or duplication through refinement of redundant data using various mechanisms and transmit a refined version of the original data upstream towards the server module or base station S_j. Data aggregation or refinement not only reduces the expected communication cost but efficiently utilizes the available onboard battery power of the resource-constrained member devices C_i in the operational IoT networks, especially for the smart cities. In literature, existing data aggregation techniques are designed to minimize either computation or communication cost overhead but not both simultaneously [17]. Additionally, the existing state-of-the-art data aggregation techniques are primarily based on duplicate-insensitive functions, which suffer heavily when eliminating or minimizing duplicate ratio [18]. Likewise, these techniques are inherently insensitive or susceptible to duplicate, redundant, and erroneous readings from faulty devices C_i in the IoT networks for smart cities. To resolve these issues, various extensions or even novel data aggregation approaches were presented in literature and were successful in achieving their goal [19, 20]. However, these mechanisms were designed particularly for resource-efficient devices C_i, that is, CH, server, or ordinary devices. Although these techniques have resolved some of the challenging issues, which are closely linked with the IoT networks for smart cities, they incur higher computational and space complexities. Hence, these approaches are not feasible for the tiny, resource-constrained member devices C_i in the IoT networks [21, 22]. Apart from it, accuracy and precision ratios of the refined data values are major concerns associated with existing approaches [23]. Therefore, the scientific community desperately needs the design and development of an optimal and global data aggregation scheme, which will be a valuable contribution if presented.

In this article, we will propose a global and optimal data aggregation scheme specifically designed for the resource-limited networks deployed in smart cities, especially for traffic control. The proposed scheme has the built-in capacity to minimize duplicate data values up to the expected level without compromising on other performance metrics such as accuracy and precision ratios. The main contributions of this article are given as follows:

(1) An algorithm is developed for controlling traffic congestion in case of undesirable circumstances, for example, in case of accident or intention blockage of the road. The law enforcement agencies will dispatch the concerned authorities to resolve such circumstances well in advance.

(2) A novel data aggregation scheme that can ensure 24/7 monitoring of a smart city minimizes the security risks and enables quick response in case of any disaster.

(3) A sophisticated mechanism to ensure accuracy and precision ratio of the surveillance system deployed in the smart cities.

(4) A dedicated mechanism is presented to detect or separate noisy data values from accurate ones and how these values are rectified preferably without consulting the source device in the IoT network designed for the smart cities.

The rest of the article is organized as follows. In Section 2, a brief literature review is presented with a strong emphasis on data aggregation and fusion techniques for heterogeneous WSNs. Next, Section 3 describes our proposed scheme for data fusion and surveillance, aggregation, and vulnerability-aware routing. In Section 4, performance evaluations are described in detail, and a comparative study of the proposed algorithms with field-proven algorithms on real-time datasets is presented. Finally, concluding remarks and future research directives are discussed in Section 5.

2. Literature Review

Redundant or duplicate data values are common issues associated with a surveillance system, especially when the data are captured by resource-limited devices C_i, that is, sensors, in the Internet of Things and wireless sensor networks. The situation becomes even more complicated if multiple devices are deployed nearby; thus, it is highly likely that the data captured by these devices are similar. Additionally, it primarily occurs due to correlation among the data captured by the randomly deployed member devices, specifically spatial and temporal, in the IoT networks for smart cities [24]. This high correlation in data values leads to the huge consumption of substantial network resources. It causes congestion through the IoT network, directly affecting the quality of network server metrics. In literature, various mechanisms have been proposed to address these issues. However, data fusion and surveillance is an ideal solution that has the capacity to minimize redundant data values, save energy, and prolong network lifetime without compromising on other evaluation metrics [14–16, 25, 26]. Harb et al. [14] have proposed a simplified data aggregation technique to refine capture data before its transmission, and it has two different phases. (i) Phase-I: In this phase, member devices are forced to conduct the data refinement process locally, that is, in-node processing. (ii) Phase-II: Server modules or CH is authorized to further refine the captured data values of various member devices in the IoT networks in this phase. It is to be noted that during phase-I, similarity functions are utilized periodically to convert raw captured data values into a refined version. Then, it is forwarded to the intended destination module, CH. Likewise, in Phase-II, also known as global aggregation, every CH module first finds and then computes how two datasets captured by different devices deployed in the field are different from each other. For this purpose, a similarity function, known as ANOVA, preferably one way, is utilized along with various tests and distance functions. It is evident from the literature that the Euclidean distance-based operation outperforms other distance-based functions in terms of energy consumption. The proposed approach is developed particularly for periodic activity-based WSNs. However, this approach's reliability and feasibility must be thoroughly investigated in environments where communication activity is either query-based or continuous. Moreover, a similar function, that is, an assumption, is used for data discrimination that can negatively affect the accuracy of the underlying network. In Ref. [15], a novel data aggregation technique based on the *k*-means clustering approach, an enhanced version of the traditional *K*-mean (EK-mean), is presented to minimize the ratio of duplicate data values. This approach operates at two different levels: (i) sensors and (ii) aggregators. In the former approach, a well-known distance measure, that is, Euclidean distance, is utilized to find those data values that are different from each other. In the latter case, an extended version of the traditional *K*-mean approach is employed as an aggregator or refiner. It is primarily used to form various groups of those fairs (preferably similar data values), captured by various member devices deployed in the real experimental setup. This mechanism can minimize the expected ratio of duplicate data values in the captured dataset, at both local and global levels, and minimize the transmission activity of the redundant data values in the network. The main benefit of EK-mean is that it removes many drawbacks associated with *K*-mean, such as complexity and computation costs, particularly in dense networks, which accelerates computation and throughput. Despite its advantages, such as efficient clustering of similar data, the proposed technique does not devise any data comparison mechanism at the cluster head and thus incurs a higher computational cost. Furthermore, like Ref. [14], the proposed technique uses a similar function and, as such, suffers from low accuracy. Moreover, this technique operates at the expense of higher energy consumption that adversely affects the resources of sensor nodes.

In Ref. [27], a classical, hierarchical, function-based data aggregation technique, that is, Tiny Aggregation (TAG), is presented. TAG relies on the tree topology for the computation of *Min*, *Max*, *Average,* and *Sum* for distributed and centralized WSNs. The proposed approach consists of two phases: distribution and collection. During the distribution phase, a query is distributed among the nodes. Next, data are generated in response to the query from the previous phase and collected from leaves towards the branches and finally towards the tree's root. TAG significantly conserves energy and prolongs network lifetime by reducing communication overhead and congestion in the network. However, it is inefficient for dynamic topologies and link/device failure. Moreover, TAG incurs additional energy consumption with topological changes. A distributed data fusion and surveillance algorithm, that is, a summarized form of data values in the node by parameters (DSNP), is presented to resolve the issues above, preferably with the available approaches or infrastructures [16]. In this scheme, data refinement at the node level is refined through different distance functions, that is, min, max, and average. These functions are used to refine the collected data transmitted remotely by the deployed IoT application-specific CH or server. In this mechanism, the Bollinger analyzer is utilized to determine whether these data values are needed to report back to the intended server module or not. Bollinger analysis uses arithmetic moving average (AMA) to calculate a tolerable range below and above the mean. DSNP mechanism is beneficial in providing refined data values with high quality, accuracy, and precision ratio. This mechanism utilizes previous mean values and how these values have deviated from the current data values in the network. However, missing data values have serious consequences, preferably on the performance of the DSNP scheme. Apart from it, this scheme has effectively utilized the moving average function and assumed that the data values captured by various sensor nodes are noise-free. Dynamic Message List-Enabled Data Aggregation (DMLDA) scheme was reported in the literature to refine data values in the heterogeneous wireless networks [25]. The primary objective of the DMLDA scheme is to develop a lightweight (preferably computation-based), efficient resource filtering and real-time scheme for the distributed environment of WSNs. This scheme has defined and developed a sophisticated data structure preferably at every node or device, which is bounded to keep track of previously transmitted data values. It is to be noted that the data structure, which is formed in the previous step, is utilized to perform data refinement (preferably real time) through various filters placed to remove the duplicate, irrelevant, and redundant data values from captured data by various nodes. However, this mechanism has achieved data aggregation specifically at the CH level only, and thus, data refinement, preferably at the member node level, is desperately needed to be developed. Therefore, CH modules consume their energy more rapidly than other devices in the network. Additionally, this mechanism has not considered an interesting metric, that is, the transmission of the ratio of the average packet in the network. However, this approach relies on a particular list obtained after the refinement process. If this list is incomplete, the whole refinement process is compromised, which is an alarming situation in the networks. In addition, DMLA has not developed any strategy for situations where the feasibility of the underlined list is questionable. Thus, this mechanism is not very useful in scenarios where query-enable, continuous, or periodic data communications are needed to realize. Fajar et al. [26] have presented two different data refinement or aggregation approaches specifically designed for the WSN and IoT networks. In this scheme, various devices were deployed in the agriculture field to capture data values, that is, air humidity, temperature, and soil humidity. Initially, a simple moving average (SMA) is used to refine data values, particularly at the node level. For this purpose, data are divided and placed into equal slots where devices are bounded to calculate the SMA of each slot such that duplicate data values are removed, which are shared or sent to the Sink. Likewise, in the second phase, a modified version of the traditional threshold-sensitive energy-efficient sensor network (TEEN) protocol [28] was utilized. This mechanism is carried out on devices with a single board for sensing; that is, these devices are forced to perform one task at a particular time interval. The traditional TEEN approach is a reactive mechanism [29] that is feasible for real-time applications. This approach has considerably extended the lifetime of the operational network by forwarding those packets that are not duplicated. Although these techniques have proved to be energy-efficient, precise, and accurate [26, 30], these techniques lack outlier detection. As a result, these approaches are highly susceptible to various outliers or noisy values, which are quite common in resource-constrained networks. Furthermore, the applicability of these approaches is needed to be examined on a large scale of WSNs, because the performance of the reactive protocols is inversely proportional to the size of the underlined network.

3. Proposed Data Aggregation and Surveillance Approach for the IoT Networks in Smart Cities

This section describes the proposed hybrid data aggregation and surveillance approach, named Aggregation Mod-L. The proposed approach can minimize duplicate data values, preferably those captured by devices C_i deployed nearby or in the IoT networks deployed in smart cities. Fusion and aggregation mechanisms are dedicated to improving the accuracy and precision of data captured by the deployed devices C_i, with a strong emphasis on maintaining the overall reliability and integrity of the refined data. However, preliminary definitions of various terminologies are defined before explaining the proposed hybrid data aggregation schemes.

3.1. Structure of the IoT Networks for the Surveillance System in the Smart Cities. The proposed approach is designed for heterogeneous IoT networks where server or cluster head (CH) module S_j is more powerful than ordinary member devices C_i in terms of onboard battery, processing, and communication power, as shown in Figure 1. It is since the proposed scheme forces every member device C_i to transmit its captured data to the concerned or nearest server device S_j, which is responsible for refining and then these data to the base station module. The proposed mechanism is based on the following assumptions:

(1) Every member device C_i is assumed to be static or motionless and is randomly deployed.

(2) Base station module is static and deployed closed to all clusters in the IoT networks for smart cities.

(3) Every device C_i is equipped with a similar onboard battery, processing, and communication power.

(4) Both member devices C_i and server or cluster head S_j have communication capacity; however, data refinement is supposed to be performed by the concerned server, not the member device in the IoT networks for smart cities.

(5) Transmission power (T_x) and residual energy E_r may vary due sending data.

3.2. Server or Cluster Head Formation Mechanism in the IoT Networks for the Surveillance System in the Smart Cities. The selection process of cluster head or server module S_j is quite similar to the well-known techniques called low-energy adaptive clustering hierarchy (LEACH) but with a slight modification to enhance the coverage area and lifetime of the operational IoT networks for the smart cities. If a device C_i is elected as a server or cluster head, then it is a prospect for the next rounds, preferably as a server or cluster head, in a $1/p$ ratio of rounds. In the proposed cluster selection and formation process, if a device has served as cluster head or server module S_j in the previous round, it is relieved from the service only if its residual energy E_r level is below the defined threshold value. However, if the E_r value is greater than the defined threshold value, then the same device C_i is permitted to act or serve as a server or cluster head for the next round. Alternatively, if a server or cluster head S_j has less residual energy than the defined or expected threshold value, then it is relieved from the service of CH and continues to serve as an ordinary member device C_i in the IoT networks for smart cities. Permitting a server module S_j to continue its service for the next round enhances the lifetime of the underlined IoT networks as cluster selection is an energy starving process.

Additionally, the proposed approach bounds every member device C_i to transmit data, preferably in raw form, to the intended or nearest CH module S_j. The concerned server module is forced to refine captured data of its member devices in the IoT networks. A device C_i is bounded to the utilized equation (1) for selecting the ideal cluster head or server module S_j in the IoT networks. Moreover, if a device C_i is deployed with an equal distance from two or more server modules S_j, then it is permitted to select one of these server modules randomly. Likewise, every member device C_i is assumed to have initial energy of 0.5j, and the threshold value is defined as residual energy E_r that is enough for a member device C_i to serve as a server module for a particular round in the IoT networks:

$$\begin{cases} \forall_{i=0\ldots n} C_i \in \sigma\left(S_j\right) iff \dfrac{\sqrt{\left(C_{x_i} - S_{x_j}\right)^2 + \left(C_{y_i} - S_{y_j}\right)^2}}{\left(x_i + y_j\right)} < \delta, \\ \exists_{i=0\ldots n} C_i \in \sigma\left(S_j\right) iff \dfrac{\sqrt{\left(C_{x_i} - S_{x_j}\right)^2 + \left(C_{y_i} - S_{y_j}\right)^2}}{\left(x_i + y_j\right)} == \delta, \\ \text{where C\&S represents device and server modules respectively,} \end{cases} \tag{1}$$

where δ is the threshold value or allowable distance among member devices in the IoT networks.

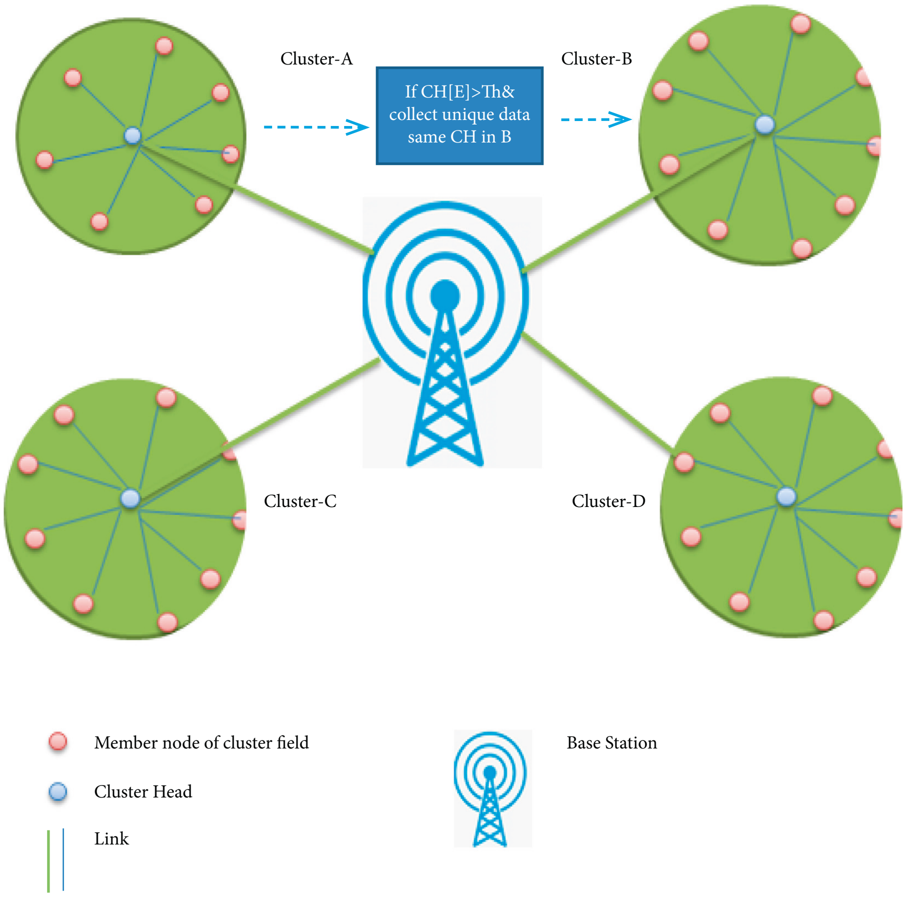
Cluster-A
If CH[E]>Th& collect unique data same CH in B
Cluster-B
Cluster-C
Cluster-D
Member node of cluster field
Cluster Head
Link
Base Station

Figure 1: Generalized infrastructure of the proposed heterogeneous IoT network.

3.3. Communication Mechanisms for the Surveillance System in the Smart Cities. In IoT networks and smart cities, the design and development of energy-efficient communication protocols are among the challenging issues. Heterogeneous networking infrastructure is one of the common mechanisms used to resolve this issue in the resource-constrained networks in general and IoT networks in particular. In this approach, communication activity is divided into two groups: (i) intranetwork and (ii) internetworks communication. In intranetwork or cluster communication, every member device C_i is forced to communicate directly with the nearest server or cluster head module S_j such that every device C_i is allocated a dedicated time slot. Time slots are used to reduce the collision probability of packets from various devices interested in starting packet transmission simultaneously. Moreover, a device C_i is forced to wait for a particular time interval as defined by the following equations:

$$L_\rho = \frac{\rho^2\left(1 + C_s^2\right)\left(C_a^2 + \rho^2 C_s^2\right)}{2\left(1 - \rho\right)\left(1 + \rho^2 C_s^2\right)}, \tag{2}$$

$$L_\rho = \frac{\lambda^2 \sigma^2 + \rho^2}{2\left(1 - \rho\right)}, \tag{3}$$

where λ and σ are used to represent the mean and variance of the packet arrival time, respectively, in the IoT networks. By substitution, we get the following:

$$C_a^2 = \frac{\sigma_s^2}{\left(1/\lambda^2\right)}. \tag{4}$$

Likewise, if μ and σ are used to represent service rate and service time, then waiting time becomes as follows:

$$C_s^2 = \frac{\sigma_s^2}{\left(1/\mu^2\right)}. \tag{5}$$

As soon as the concerned server or cluster head module S_j receives data from approximately all member devices C_i, its expected or defined waiting period is expired. It has to refine these data values before activating the transmission activity. For this purpose, we have proposed a novel idea to reduce possible comparisons among data from different devices C_i in the operational IoT network. In this scheme, every server module S_j performs data aggregation to match the data values of those devices C_i, deployed in the same geographic area within the smart building environment. For example, four, C_1, C_2, C_3, C_4, different devices C_i are deployed in room-1 and three, that is, C_5, C_6, C_7, in room-2, respectively. Additionally, we assume that these devices are members of the same server module, say "S_1." The server module compares captured data values of C_1, C_2, C_3, C_4 with each other as it is highly likely that their values are similar. Likewise, data values captured by devices deployed in room-2 are compared for possible duplication. It is to be noted that the proposed scheme forces the server module to avoid matching data values of devices C_i deployed in a different location in the particular smart building environment. Although it is possible that some data values, which are collected by various devices C_i deployed at different locations, may be the same, it is very rare and can be neglected. Additionally, the overhead associated with refining these values is more than the ratio of the duplicate data values in the IoT networks.

Apart from it, the proposed data refinement scheme can detect outliers (if any) in the captured data by various devices C_i in the IoT networks, which are designed specifically for smart cities. Once outliers are detected, a sophisticated method is used to rectify those readings before the data refinement process. Detection and correction measures were incorporated in the proposed scheme because it is highly likely that data captured by resource-constrained devices C_i contain irrelevant or false data that must be rectified before processing. Additionally, the proposed scheme can enhance the accuracy and precision ratio of the underlined real-time decision support system if it is solely based on the captured readings of these devices in the operational IoT networks. It is to be noted that the proposed outlier's detection mechanism is simple and resource-efficient that it is easily executable on the ordinary device as well. However, the proposed algorithm is designed to be executed on server or cluster head level, which is usually more powerful than ordinary devices C_i deployed in a real experimental environment of smart buildings. The proposed server- or cluster head (S_j)-level data aggregation algorithm is presented in Algorithm 1.

Once data values, preferably those captured by ordinary member devices C_i, are refined, the next step is transmitting or sharing these values with a common destination device or server that is the base station in this case. For this purpose, the concerned server module transmits the refined version of data values captured by member devices in packets using a wireless communication mechanism. In IoT networks, a server or cluster head module will likely have the capacity to communicate directly with the intended destination device, which is the base station in this case. However, suppose the concerned server module's coverage area is limited and cannot communicate directly with the concerned destination device. In that case, a multihop communication strategy is adopted where packets are transmitted to the nearest server module in the IoT networks. In these scenarios, cluster heads act as relay devices for the particular source device, or it is assumed as a bridge between the base station and the concerned server or cluster head module. The coverage area of a particular device C_i or server module S_j, if we assume that transmitters or antennas of these devices are omnidirectional, is computed using the following equation:

$$C_omni = \frac{3 * \sqrt{3}}{2} * R^2, \tag{6}$$

where R is used to represent cellular or coverage area radius. Moreover, it will be more beneficial if omnidirectional antennas are set apart for 120 C^0 to the concerned base station in the IoT networks. Likewise, the coverage area of the respective base station module is calculated by using the following equations:

$$BS_{\text{sector}} = \sum \frac{A_i}{360}, \tag{7}$$

$$BS_{\text{sector}} = \frac{\pi * R_i^2}{360}, \tag{8}$$

where R_i is used to represent the radius of every antenna pattern angle such as i degree, which is preferably determined through the link budget in order to remove a maximum allowable path loss in the omnidirectional antennas. Furthermore, the path loss ratio of particular device C_i and server S_j is needed to be reduced, which is carried out in the proposed scheme by reducing the transmission ratio of various communication channels. The path loss ratio of an ordinary device C_i or server module S_j, particularly in terms of wavelength, is computed using the following equation:

$$C_{FSPL} = \left(\frac{2\pi d}{\lambda}\right)^2, \tag{9}$$

where FSPL and λ are used to represent the path loss ratio of free space and signal wavelength (preferably in meters), respectively, in the IoT networks. Likewise, the path loss ratio of an ordinary device C_i or server module S_j, particularly in terms of frequency, is computed using the following equation:

$$C_{FSPL} = \left(\frac{2\pi \mathrm{d} f}{c}\right)^2, \tag{10}$$

where d, f, and c are used to represent the distance between the source and destination transmitters, frequency (preferably in hertz), and speed of light, respectively, in the IoT networks.

3.4. Proposed 3-Dimensional Logical CH-Enabled DHT and Flooding Free Routing Mechanisms for the Surveillance System in the Smart Cities. In the proposed data aggregation approach, we have assumed that every member device C_i has a specific cluster or group's logical identifier (LID). Additionally, every device C_i is assumed to have a unique public identifier known as a universal identifier (UID), which is formed by either integrating IP, and media access control (MAC) addresses or separately used depending on the application requirement. However, we have used a hybrid version in the proposed mechanism to enhance IoT networks' accuracy and precision ratio. The proposed scheme, that is, Aggregated$_M OD_L$each, is implemented and tested on the IoT network, depicted in Figure 2 where three different clusters (X, Y, and Z) are formed, each with a unique cluster

```
Input: Captured Data
Output: Return Refined Data
Matched←zero;
Non – matched←zero;
Counter←0;
$\text{device}_1 \leftarrow 0$;
$\text{device}_2 \leftarrow 0$;
for every captured data by device 1 & 2 do
    if distance $(\text{data} - \text{value}(\text{device}_1), \text{data} - \text{value}(\text{device}_2)) \leq \delta$then
        $\text{Matched}_i \leftarrow \text{data} - \text{value}(\text{device}_1)$ OR $\text{data} - \text{value}(\text{device}_2)$;
        $i \leftarrow i + 1$;
    else
         OR $\text{data} - \text{value}(\text{device}_2)$;
        $j \leftarrow j + 1$;
    end if
    if $(\text{data} - \text{value}(\text{device}_1), \text{data} - \text{value}(\text{device}_2)) \in$ outliers then
        $\text{Avg} = (\sum_{i=0}^{n} (\text{device}_{\text{val}})/n - \sum_{i=0}^{m} (1))$;
        $\text{data} - \text{value}(\text{device}_1), \text{data} - \text{value}(\text{device}_2)_i \leftarrow \text{Avg}$
    end if
end for
return Refined Data Values
```

Algorithm 1: Proposed server-side data aggregation and surveillance algorithm for the smart cities.

head or server module responsible for transmitting the captured data to the destination module. Apart from it, the existing state-of-the-art mechanism, that is, 3-DcRP logical clustering techniques, is integrated with the proposed approach. For example, various member devices C_i, that is, $Z_{1\ldots4}$, of this cluster, say "Z," capture data values at a particular time interval, that is, 10 : 00 : 00. These devices share their captured data values with the cluster head or server module, which is forced to refine them before sending them to the centralized server module in the IoT network as these devices are neighbors, as described in Figure 2, and data values captured by these devices likely contain duplicate values. Additionally, the expected probability of duplicate values is highly increased if these readings are captured by neighboring devices C_i and preferably at the same time interval, that is, 10 : 00 : 00 in this case. The server module or cluster head of the particular area thoroughly examines these readings by utilizing the proposed algorithm and eliminating duplicate data values. Thus, a refined version, preferably duplication free, of the captured data values is transmitted towards the intended destination module, which is the sink in this case. Apart from it, it is highly likely that a server device S_j may not be able to communicate directly with the intended destination module; thus, a multihop communication mechanism is used where data packets are transmitted through other servers or clusters head modules, as shown in Figure 2. Moreover, a relaying cluster head either transmits the received packets directly or may apply data aggregation to refine it further. However, we have observed that performing data aggregation at different levels creates a hurdle and is an additional overhead. Thus, we have preferred and bounded cluster head or server module S_j, particularly those which serve as relay devices, to forward packets without performing aggregation activity. Additionally, we have observed that the proposed algorithm performance, specifically in terms of accuracy and precision ratio of the refined data values, is not affected by an increasing number of devices C_i in the IoT network.

4. Performance Evaluation: Results

In this section, we have thoroughly evaluated the expected performance of the proposed data aggregation algorithm in refining the captured data values with the utilization of minimum possible resources in the IoT networking infrastructure. Furthermore, the proposed scheme is compared with the existing state-of-the-art techniques in terms of various performance evaluation metrics used for networking in general and IoT networks in particular. These algorithms, that is, proposed and existing, were developed in MATLAB, which is sophisticated software for simulating newly developed algorithms in various application areas. Furthermore, we ran these experiments multiple times, and these algorithms' average values were used as approximate results. We have used similar topologies (preferably logical, which defined how the IoT networks operate in a realistic environment), the transmitter's coverage area, onboard batteries, nodes, and sink device. Apart from it, a random deployment strategy was utilized for both devices and server or cluster head modules to streamline it with the real-time implementation of these networks, where 90% of the devices were assumed as ordinary devices and 10% as cluster head or server modules. The coverage area of ordinary devices was assumed to be 500 m, which is in line with the commercialized version of the Waspmote PRO Board available in the market. Likewise, every cluster was assumed to either balanced or nonbalanced infrastructures as balance clustering mechanisms are difficult to realize in a real-world

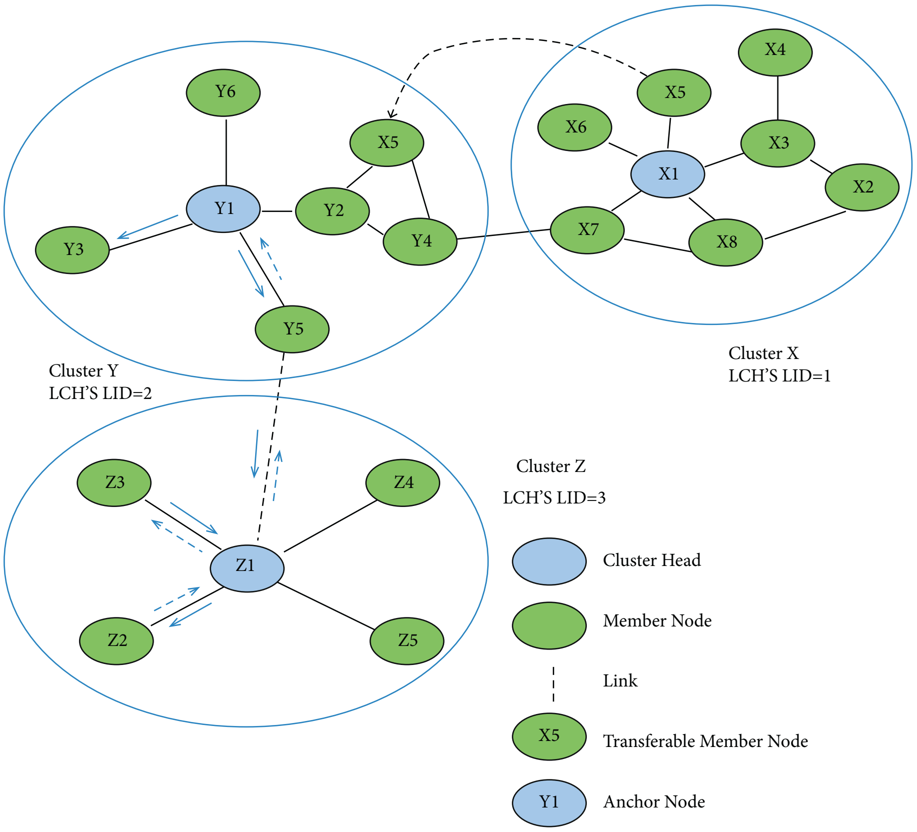

Figure 2: An exemplary scenario of the proposed and 3-DcRP approaches.

environment. Numerous parameters used in the simulation setup are depicted in Table 1.

We have utilized various measures to verify the exceptional performance of the proposed hybrid data aggregation mechanism in the realistic environment of IoT networks. These evaluation metrics are the time of first and last device energy consumption, end-to-end delay (both among devices, server modules, and sink), lifetime, and average packet delivery ratio. Furthermore, the proposed scheme is compared against the existing state-of-the-art approaches, preferably data aggregation, and was thoroughly examined for how long a particular algorithm maintains maximum live devices in the IoT networks.

4.1. Detection of Dead Devices in the IoT Network for the Surveillance System in the Smart Cities. A device C_i is assumed dead if it has consumed the available onboard battery power completely in the resource-limited IoT networks. Moreover, it is assumed as one of the crucial evaluation metrics used to verify the performance of newly developed data aggregation schemes in the IoT networks. For this purpose, the proposed data aggregation scheme is compared with the existing state-of-the-art approaches in terms of the time when the first device is encountered, which has consumed its onboard battery power completely in the IoT networks. Likewise, the proposed and existing state-of-the-art approaches are compared in terms of when half (50%) of member devices have to consume their onboard battery power completely. The performance of these algorithms in terms of these two measures is represented in Figures 3–5 respectively. We have observed that the proposed scheme's performance in terms of both metrics is better than the existing data aggregation approaches.

Table 1: Simulation setup of heterogeneous WSN.

Parameters	Values
IoT network area of deployment	700 m × 700 m
Member devices	1500
Server or cluster head modules	10%
Base station	1
Device's initial energy	0.5 J
Device's residual energy	Available power
Energy consumption of the transceiver (T_i)	5,026 mA
Transmission range (T_r)	100 m
Energy required for amplification (devices)	10 pJ/bit/m2
Energy required for amplification (server)	Efs/10
Initial HC of an ordinary node (T_r)	∞
Maximum distance between nodes	250 m
Coverage area of device	500 m
Coverage area of cluster head	500 m
Coverage area of base station	Maximum

Apart from that, the proposed data aggregation scheme performance is compared against the existing state-of-the-art approaches in scenario(s) where the last device in the IoT

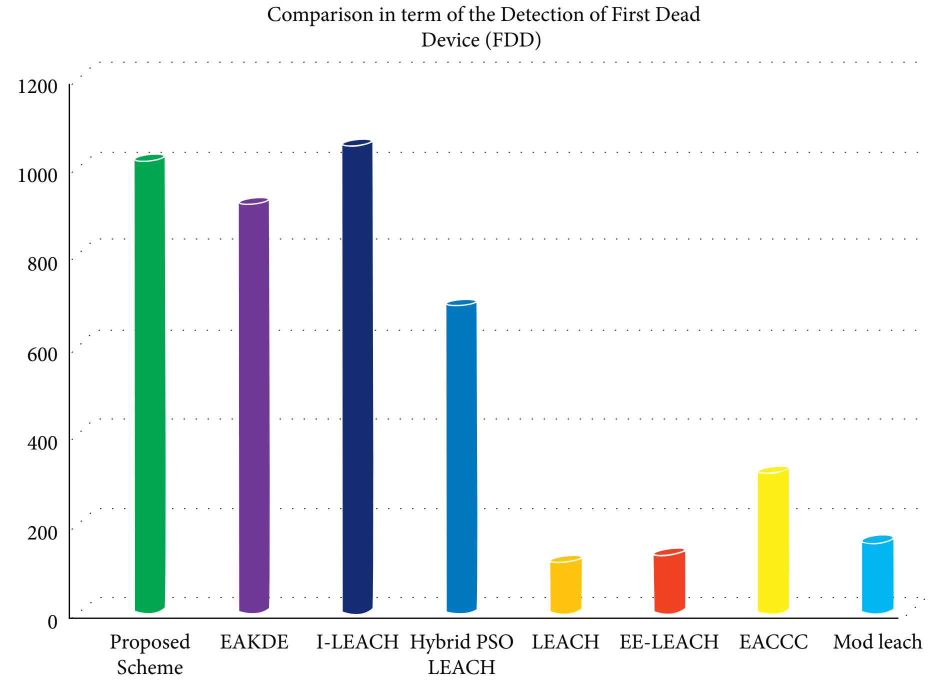

Figure 3: A realistic scenario where the first device has completely consumed the available power.

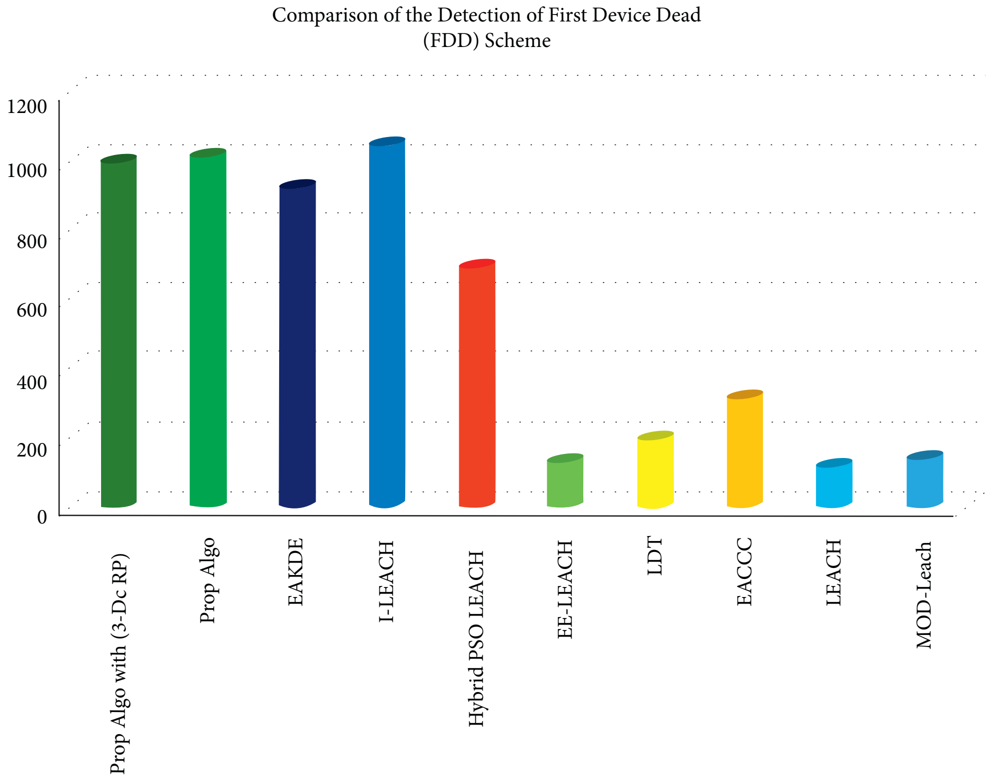

Figure 4: A realistic scenario where the first device has completely consumed the available power.

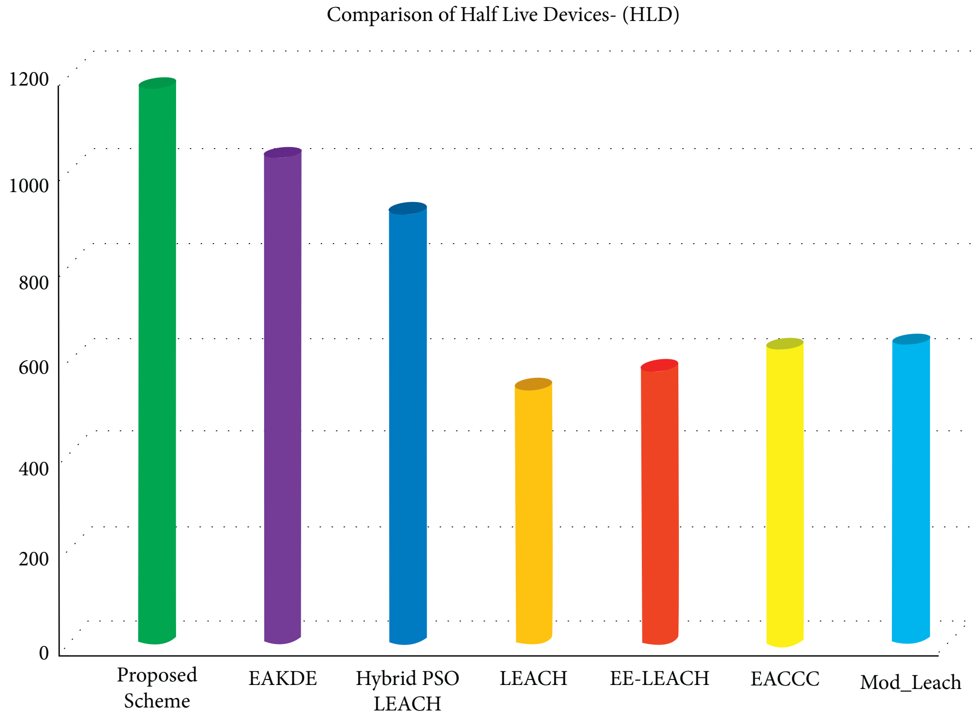

Figure 5: A realistic scenario where half of the member devices have completely consumed the available power.

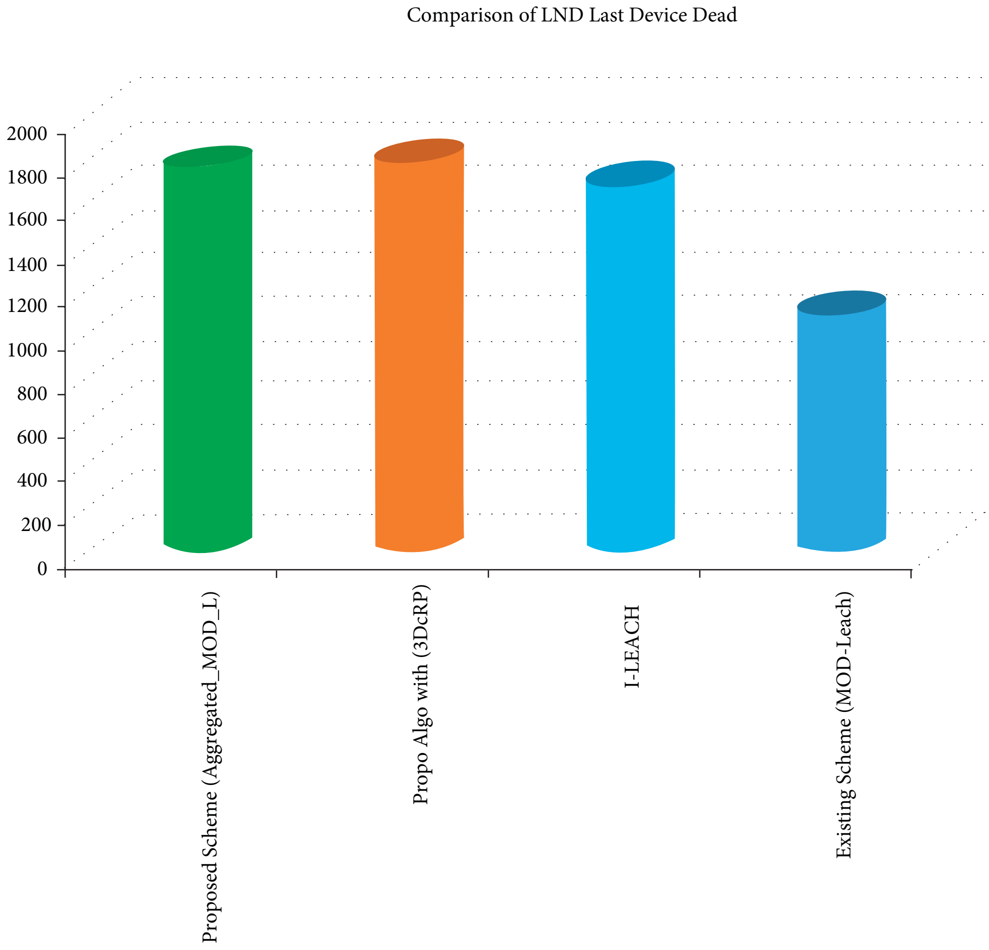

Figure 6: A realistic scenario where the last device has completely consumed the available power.

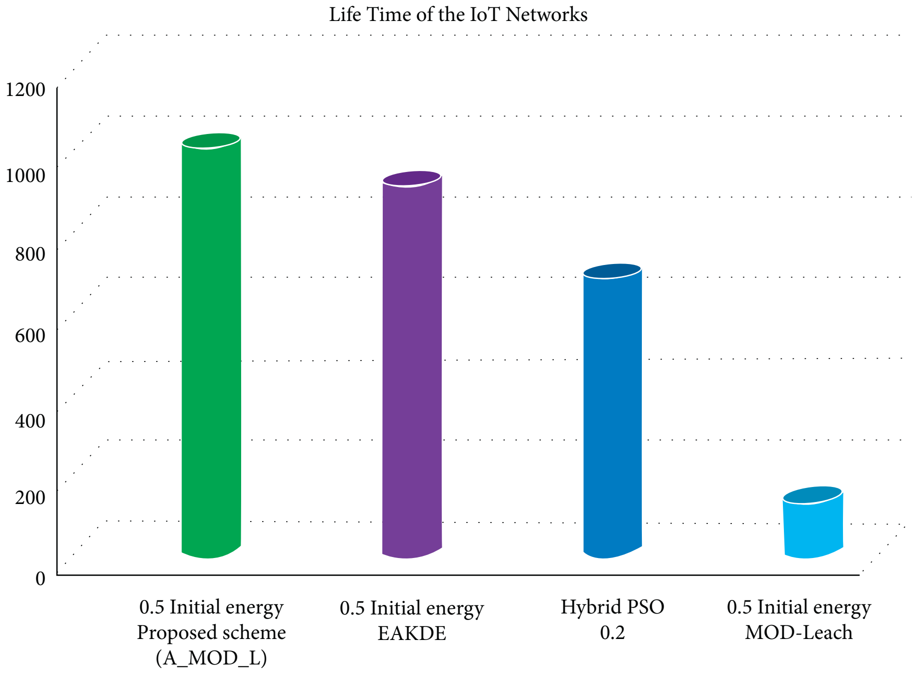

Figure 7: Lifetime of the proposed data aggregation and existing state-of-the-art approaches in the IoT networks.

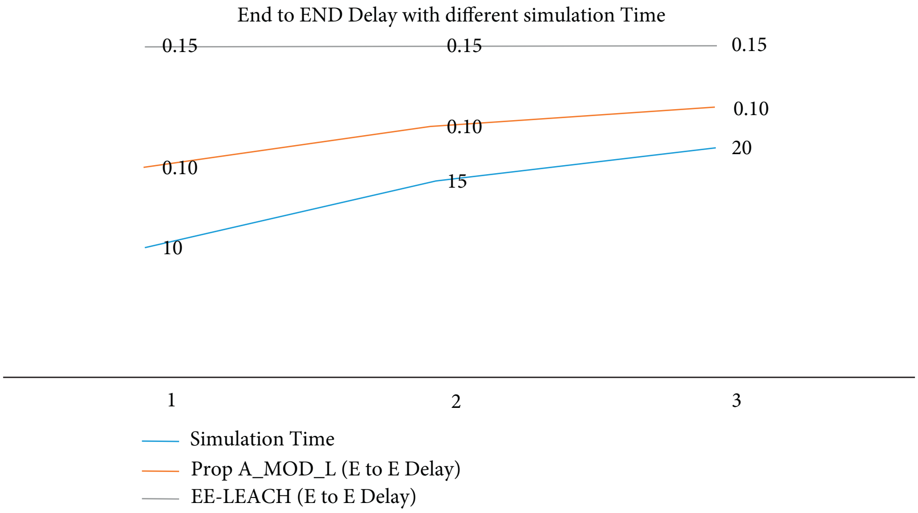

Figure 8: End-to-end delay of the proposed data aggregation and existing state-of-the-art approaches in the IoT networks.

networks consumes its onboard battery. The simulation results have verified that the proposed data aggregation scheme is better than existing approaches, as shown in Figure 6.

4.2. Lifetime of the IoT Networks for the Surveillance System in the Smart Cities. The underlined IoT networks' lifespan is assumed to be one of the vital evaluation metrics specifically used to judge the expected performance of the newly developed data aggregation approaches in the IoT networks. For this purpose, the proposed data aggregation scheme's performance is compared against state-of-the-art existing approaches, particularly regarding the network's lifetime. We have observed that the proposed scheme has outperformed existing approaches in terms of lifetime performance evaluation metrics, as shown in Figure 7. It is to be noted that the lifetime of the underlined IoT network, which is based on the proposed data aggregation scheme, is longer than existing approaches, which is because the proposed scheme has minimized (if not completely avoided) the transmission of duplicate data packets, specifically of neighboring devices in the IoT networks.

4.3. End-to-End Delay of the IoT Networks for the Surveillance System in the Smart Cities. End-to-end delay of the underlined IoT networks is assumed as one of the vital

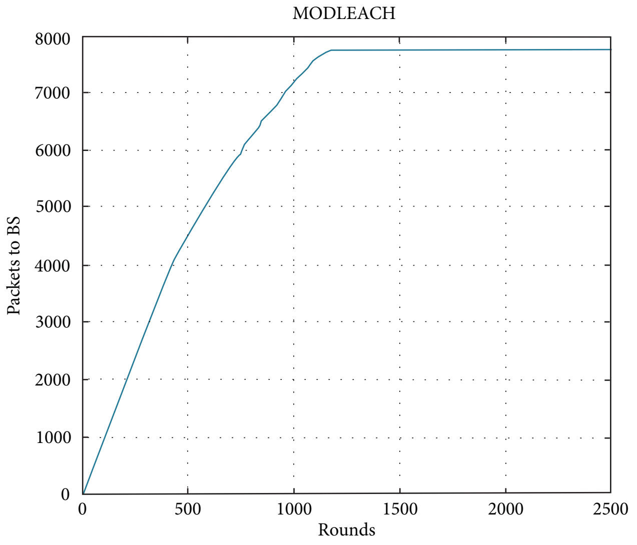

Figure 9: Average packet delivery ratio measured at server or cluster head module in the IoT networks.

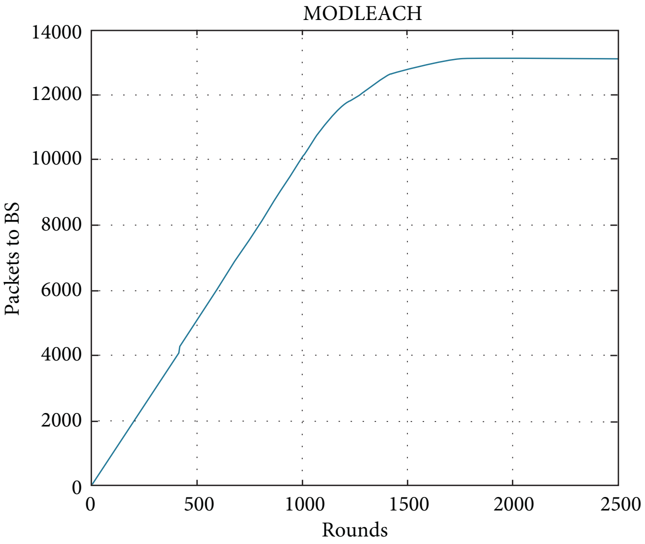

Figure 10: Average packet delivery ratio measured at the base station module in the IoT networks.

evaluation metrics, which are specifically used to judge the expected performance of the newly developed data aggregation approaches in the IoT networks. For this purpose, the proposed data aggregation scheme's performance is compared against state-of-the-art existing approaches, particularly in terms of latency. We have observed that the proposed scheme has outperformed existing approaches in terms of end-to-end delay performance evaluation metrics, as shown in Figure 8. It is to be noted that the end-to-end delay of the underlined IoT network, which is based on the proposed data aggregation scheme, is shorter than that of existing approaches, which is because the proposed scheme has adopted the direct communication infrastructure, specifically in terms of member devices, in the IoT networks.

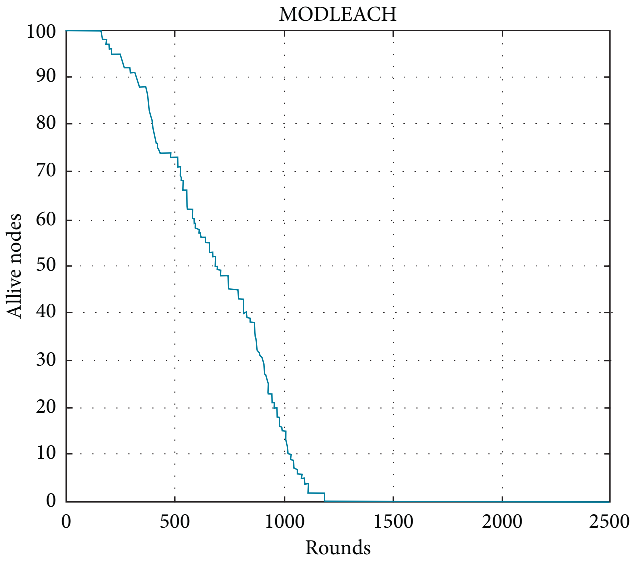

Figure 11: Alive devices in the IoT networks (proposed scheme).

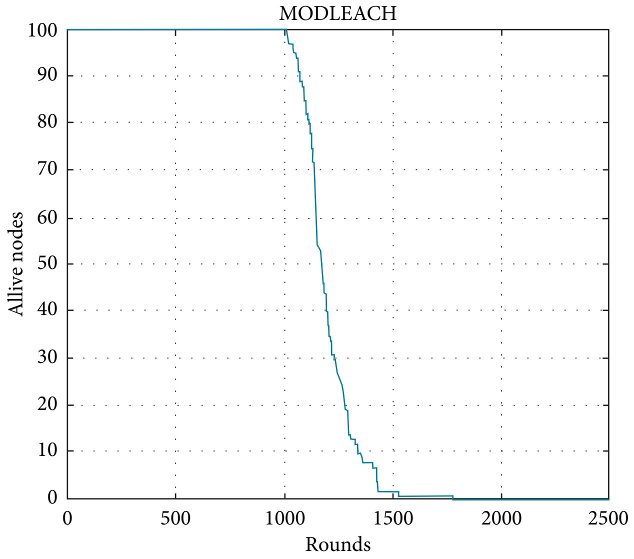

Figure 12: Alive devices in the IoT networks (existing schemes).

4.4. Packet Delivery Ratio in the IoT Networks for the Surveillance System in the Smart Cities. The average packet delivery ratio (APDR) is one of the common measures used to evaluate the performance of the newly developed data aggregation and routing approaches in the IoT networks. The proposed data aggregation and existing schemes were evaluated based on the APDR ratio. We have observed that the proposed scheme has achieved the maximum possible APDR value than the existing state-of-the-art approaches in the IoT networks. Additionally, we have repeated these experiments at least five times to improve these results' accuracy and precision ratio. From Figures 9 and 10, it is clearly evident that the proposed scheme is better than the existing state-of-the-art approaches in both scenarios, that is, cluster head and base station modules.

4.5. Alive Devices in the IoT Networks for the Surveillance System in the Smart Cities. An interesting evaluation metric in the IoT network is reporting or identifying how many devices are active at a particular time interval. Moreover, this approach becomes more important if we are interested in finding the ratio of active devices in a particular area within the smart building infrastructures. For this purpose, the proposed scheme is thoroughly examined to verify how many devices are active at specified time intervals and sophisticated locations in the smart building where the proposed scheme-based IoT network is implemented. The proposed scheme has enabled member devices to operate longer compared to the existing state-of-the-art approaches, as shown in Figures 11 and 12.

5. Conclusion and Future Work

The world was struck by heart wrenching and unfortunate incident of 9/11 that led to an unending terrorism wave around the globe. Although this problem needs a political solution, technology has to go hand in hand to implement policies sought by world leaders. Data fusion and surveillance are the common approaches used to minimize the ratio of duplicate data values in resource-limited networks. These approaches try to reduce this ratio by comparing data values from various devices deployed in a real environmental setup. In this article, we have presented an extended version of the existing state-of-the-art LEACH protocol and proposed an ideal data aggregation approach for the IoT networks. The proposed scheme has bounded every member device to send captured data directly to the nearest server or cluster head modules in the IoT networks. The concerned server module is responsible for refining these values before sending them to the base station module. Additionally, the server module is bound to match data values of the neighboring devices or devices deployed in the same region, that is, room or floor in the IoT networks. Moreover, sophisticated mechanisms are presented to address the outlier issue, which is tightly coupled with the IoT networks. The simulation results have verified the exceptional performance of the proposed data aggregation approach, where it is evident from various results that the proposed scheme is an ideal solution for heterogeneous IoT networks. In the future, we are interested in extending the operational capacities of the proposed data aggregation scheme to homogeneous IoT networks where member devices are allowed to move from one place to another.

Data Availability

The data used in this research is available within the article.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflicts of Interest

All authors declare that they have no conflicts of interest.

Acknowledgments

The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through project no. 7965.

References

[1] M. Tahir and S. Anwar, “Transformers in pedestrian image retrieval and person re-identification in a multi-camera surveillance system,” *Applied Sciences*, vol. 11, no. 19, p. 9197, 2021.

[2] Q. Wang, D. Lin, P. Yang, and Z. Zhang, “An energy-efficient compressive sensing-based clustering routing protocol for wsns,” *IEEE Sensors Journal*, vol. 19, no. 10, pp. 3950–3960, 2019.

[3] N. Awan, S. Khan, M. K. I. Rahmani et al., “Machine Learning-Enabled Power Scheduling in Iot-Based Smart Cities,” 2021.

[4] M. Shorfuzzaman, M. S. Hossain, and M. F. Alhamid, “Towards the sustainable development of smart cities through mass video surveillance: a response to the covid-19 pandemic,” *Sustainable Cities and Society*, vol. 64, Article ID 102582, 2021.

[5] C. Englund, E. E. Aksoy, B Astrand, F. Alonso-Fernandez, M. D. Cooney, and S. Pashami, “Ai perspectives in smart cities and communities to enable road vehicle automation and smart traffic control,” *Smart Cities*, vol. 4, no. 2, pp. 783–802, 2021.

[6] X. Ding, J. Guo, D. Li, and W. Wu, “An incentive mechanism for building a secure blockchain-based internet of things,” *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 1, pp. 477–487, 2021.

[7] M. Tahir, N. N. Hamadneh, and M. K. I. Rahmani, “Machine learning and deep learning are crucial to the existence of iot and big data,” *A Step Towards Society*, vol. 5, pp. 69–77, 2021.

[8] R. Khan, I. Ali, M. A. Jan et al., “A Hybrid Approach for Seamless and Interoperable Communication in the Internet of Things,” *IEEE Network*, vol. 35, pp. 202–208, 2021.

[9] M. C. Vuran, Ö B. Akan, and I. F. Akyildiz, “Spatio-temporal correlation: theory and applications for wireless sensor networks,” *Computer Networks*, vol. 45, no. 3, pp. 245–259, 2004.

[10] M. K. I. Rahmani, “Blockchain technology: principles and algorithms,” in *Blockchain Technology and Computational Excellence for Society 5.0*, pp. 16–27, IGI Global, Pennsylvania, USA, 2022.

[11] S. Abbasian Dehkordi, K. Farajzadeh, J. Rezazadeh, R. Farahbakhsh, K. Sandrasegaran, and M. Abbasian Dehkordi, “A survey on data aggregation techniques in iot sensor networks,” *Wireless Networks*, vol. 26, no. 2, pp. 1243–1263, 2020.

[12] M. Alam, M. Albano, A. Radwan, and J. Rodriguez, “Context based node discovery mechanism for energy efficiency in wireless networks,” in *Proceedings of the IEEE International*

Conference on Communications (ICC), pp. 6008–6012, IEEE, Ottawa, ON, Canada, June 2012.

[13] X. Zheng, A. Idrees, F. Khan et al., "A reliable communication and load balancing scheme for resource-limited networks," *IEEE Access*, vol. 8, pp. 179921–179930, 2020.

[14] H. Harb, A. Makhoul, D. Laiymani, and A. Jaber, "A distance-based data aggregation technique for periodic sensor networks," *ACM Transactions on Sensor Networks*, vol. 13, no. 4, pp. 1–40, 2017.

[15] M. Rida, A. Makhoul, H. Harb, D. Laiymani, and M. Barhamgi, "Ek-means: a new clustering approach for datasets classification in sensor networks," *Ad Hoc Networks*, vol. 84, pp. 158–169, 2019.

[16] L. F. C. Maschi, A. S. R. Pinto, R. I. Meneguette, and A. Baldassin, "Data summarization in the node by parameters (dsnp): local data fusion in an iot environment," *Sensors*, vol. 18, no. 3, p. 799, 2018.

[17] G. Wei, Y. Ling, B. Guo, B. Xiao, and A. V. Vasilakos, "Prediction-based data aggregation in wireless sensor networks: combining grey model and kalman filter," *Computer Communications*, vol. 34, no. 6, pp. 793–802, 2011.

[18] B. Kang, P. K. H. Nguyen, V. Zalyubovskiy, and H. Choo, "A distributed delay-efficient data aggregation scheduling for duty-cycled wsns," *IEEE Sensors Journal*, vol. 17, no. 11, pp. 3422–3437, 2017.

[19] G. Dhand and S. S. Tyagi, "Data aggregation techniques in wsn: Survey," *Procedia Computer Science*, vol. 92, pp. 378–384, 2016.

[20] T. Kiruthiga and N. Shanmugasundaram, "In-network data aggregation techniques for wireless sensor networks: a survey," *Computer Networks, Big Data and IoT*, vol. 66, pp. 887–905, 2021.

[21] K. Sarangi and I. Bhattacharya, "A study on data aggregation techniques in wireless sensor network in static and dynamic scenarios," *Innovations in Systems and Software Engineering*, vol. 15, no. 1, pp. 3–16, 2019.

[22] M. Arif and K. I. Rahmani, "Adaptive ara (aara) for manets," in *Proceedings of the Nirma University International Conference on Engineering (NUiCONE)*, pp. 1–6, IEEE, Ahmedabad, India, November 2012.

[23] W. Alghamdi, M. Rezvani, H. Wu, and S. S. Kanhere, "Routing-aware and malicious node detection in a concealed data aggregation for wsns," *ACM Transactions on Sensor Networks*, vol. 15, no. 2, pp. 1–20, 2019.

[24] I. Horiya Brahmi, S. Djahel, D. Magoni, and J. Murphy, "A spatial correlation aware scheme for efficient data aggregation in wireless sensor networks," in *Proceedings of the IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops)*, pp. 847–854, IEEE, Clearwater Beach, FL, USA, December 2015.

[25] T. Du, Z. Qu, Q. Guo, and S. Qu, "A high efficient and real time data aggregation scheme for wsns," *International Journal of Distributed Sensor Networks*, vol. 11, no. 6, Article ID 261381, 2015.

[26] M. Fajar, J. Litan, M. Abdul, and A. Halid, "Energy efficiency using data filtering approach on agricultural wireless sensor network," *International Journal of Computer Engineering and Information Technology*, vol. 9, no. 9, p. 192, 2017.

[27] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "Tag: a tiny aggregation service for ad-hoc sensor networks," *ACM SIGOPS - Operating Systems Review*, vol. 36, pp. 131–146, 2002.

[28] A. Manjeshwar and D. P. Agrawal, "Teen: arouting protocol for enhanced efficiency in wireless sensor networks," *Ipdps*, vol. 1, p. 189, 2001.

[29] I. Y. Mohammed, "Comparative analysis of proactive & reactive protocols for cluster based routing algorithms in wsns," *World Scientific News*, vol. 124, no. 2, pp. 131–142, 2019.

[30] R. Singla, N. Kaur, D. Koundal, S. A. Lashari, S. Bhatia, and M. K. Imam Rahmani, "Optimized energy efficient secure routing protocol for wireless body area network," *IEEE Access*, vol. 9, pp. 116745–116759, 2021.