

ORIGINAL RESEARCH

Open access

Optimized IoT Routing Protocol for Energy Efficiency in Smart Cities: A Marine Predators Algorithm Approach

Yuki Yamamoto^{1*}, Kenji Ito¹

Abstract

This research proposes a novel, low-power routing protocol tailored for IoT networks that merges intelligent data-aggregation clustering with optimized path selection to overcome critical data-handling challenges in smart urban environments. The protocol relies on a layered architecture that includes Data Fusion Heads (DFH), intermediate relay DFHs, and the marine predators algorithm—an effective nature-inspired optimizer equipped with a purpose-built fitness function. This function simultaneously refines multiple criteria: the geographic clustering density of Sensor Nodes (SNs) inside each data fusion group (DFG), the total hop distance toward the sink, intra-group node proximity, residual battery level (RE), Average Scale of Building Occlusions (ASBO), and the rotation interval of the Primary DFH (PDFH). One of the protocol's core advancements is the systematic application of in-network data fusion, which drastically reduces duplicate packet forwarding and markedly improves the accuracy and usefulness of collected information across all DFGs. By aggregating readings from multiple sensors, the volume of outbound traffic is greatly reduced, yielding substantial improvements in energy efficiency. Two complementary forwarding strategies are supported: single-hop forwarding, where already fused payloads travel directly to the sink, and multi-hop forwarding, in which an appropriate PDF relay is dynamically elected using a weighted relay selection metric that incorporates remaining energy (RE), sink distance, and ASBO. By integrating robust fault-tolerance procedures, intelligent redundancy elimination, and efficient fusion mechanisms, the protocol achieves markedly greater dependability and sustains strong operation under adverse or unpredictable network conditions. Detailed simulation experiments, along with side-by-side benchmarking against contemporary methods, confirm the protocol's clear advantage across essential performance indicators—namely, network longevity, total energy expenditure, data throughput, and mean end-to-end latency. Compared with leading recent approaches such as the Particle Swarm Optimization-based energy-efficient clustering protocol (PSO-EEC), linearly decreasing inertia weight PSO (LDIWPSO), Optimized Fuzzy Clustering Algorithm (OFCA), and Novel PSO-based Protocol (NPSOP), the presented solution delivers outstanding gains. Concretely, network lifetime is prolonged by 299% versus PSO-EEC, 264% versus LDIWPSO, 306% versus OFCA, and 249% versus NPSOP. Energy usage drops by 254% relative to PSO-EEC, 264% relative to LDIWPSO, 247% relative to OFCA, and 253% relative to NPSOP. Throughput rises by 67% over PSO-EEC, 59% over LDIWPSO, 53% over OFCA, and 50% over NPSOP. By skillfully combining data aggregation with carefully tuned routing decisions, this protocol establishes a new benchmark for power-efficient organization and communication within IoT-based data fusion groups, delivering a dependable and adaptable framework suitable for diverse smart-city and IoT scenarios.

Keywords IoT, Data fusion, Optimization algorithm, MPA, Routing, Scheduling

*Correspondence:

Yuki Yamamoto

yuki.yamamoto@kyoto-u.ac.jp

¹ Department of Electric Power Networks, Faculty of Engineering, Kyoto University, Kyoto, Japan

Introduction

The Internet of Things (IoT) has transformed numerous fields—such as intelligent transportation, smart urban systems, security and surveillance, medical services, precision agriculture, public safety, water resource management, and ecological observation—through pervasive device interconnectivity and real-time data sharing [1, 2]. At the core of these IoT systems are Sensor Nodes (SNs), which persistently sense and relay environmental information. Deployed in large numbers, these SNs constitute the foundational layer of IoT infrastructures, making effective coordination and resource management indispensable for achieving high performance and long-term viability [3].

To boost both operational efficiency and capability in IoT deployments, data fusion emerges as a fundamental mechanism [4]. Data fusion involves combining inputs from multiple SNs to generate more reliable, precise, and actionable insights than any single node could provide on its own [5]. By aggregating and refining data at different tiers—whether at the level of individual sensors, within clusters, or at gateway points—data fusion approaches eliminate duplicates, suppress interference, and enhance analytical outcomes [6]. These strategies also substantially reduce network traffic and power consumption by avoiding the broadcast of unprocessed readings, thereby prolonging the operational lifespan of the entire system.

Moreover, data fusion groups (DFGs) organize SNs into cohesive clusters, each overseen by a designated leader, the Data Fusion Head (DFH), and assisted by ordinary members, the DFG Members (DFGMs). This tiered organization streamlines data fusion activities, promotes judicious resource allocation, conserves battery life, and lengthens network endurance through localized processing, reduced superfluous transmissions, and equitable distribution of processing and communication burdens across nodes [7]. Embedding data fusion capabilities into clustered IoT frameworks delivers scalable, power-thrifty, and dependable performance, which remains vital for the ongoing advancement of IoT solutions in a wide array of application areas [8, 9].

Problem statement and motivation

Managing large-scale IoT sensor deployments poses substantial challenges, particularly in terms of power efficiency, reliable data delivery, and overall network resilience. A primary concern involves the election of DFHs, which play a central role in cluster formation [10]. DFHs collect readings from associated DFGMs, execute data consolidation, and forward the processed results either directly to the sink or via intermediate relay DFHs. Yet the DFH position demands considerable energy expenditure, and poorly chosen heads can cause accelerated battery drain, unequal workload sharing, and drastically reduced network lifetime [11]. Hence, intelligent DFH selection stands as a cornerstone of successful clustering [12].

An additional difficulty lies in determining the ideal count of DFGs and ensuring balanced node membership within them. IoT environments are inherently volatile, with frequent shifts in topology and operating conditions that render conventional DFH election strategies—often based exclusively on remaining energy (RE)—insufficiently responsive [13]. Such static techniques frequently fail to accommodate rapid changes, leading to energy disparities and early node exhaustion. Recent advances in machine learning introduce valuable adaptability and situational awareness to DFH election processes, effectively overcoming these shortcomings [14, 15].

Energy optimization within DFGs is equally pressing, given that most SNs depend on constrained, non-rechargeable power supplies such as batteries [16]. Achieving low energy expenditure alongside dependable, low-latency data delivery becomes crucial, particularly in time-sensitive domains like remote patient monitoring, where prompt, trustworthy information is non-negotiable. The fluid nature of IoT networks necessitates algorithms that can respond to topological changes, fairly distribute workloads, and prudently manage scarce computing and memory resources.

Data fusion itself introduces further complexities in resource-limited IoT settings. Synthesizing and interpreting measurements from diverse SNs to yield coherent, high-fidelity results is resource-heavy and constrained by hardware limitations [17]. Suboptimal fusion procedures can trigger unnecessary packet retransmissions, increased power consumption, and longer delays. Maintaining superior fused-data quality is indispensable for domains such as pollution

tracking and factory automation, where critical decisions hinge on precise, up-to-the-minute inputs. Although advanced fusion methods exist, incorporating them into constrained IoT platforms raises issues of expandability, flexibility, and power frugality [18].

The incorporation of cutting-edge technologies—including machine learning, Artificial Intelligence (AI), and especially Computational Intelligence (CI)—into DFH election, DFG formation, and fusion operations creates both promising avenues and notable hurdles [19]. Evolutionary and swarm-based techniques such as Genetic Algorithms (GA) [20], Particle Swarm Optimization (PSO) [21], and Ant Colony Optimization (ACO) [22] have shown strong capability to support multi-objective, context-sensitive choices by weighing factors like RE, link quality, traffic burden, and information overlap [23]. When appropriately applied, these approaches markedly improve clustering quality and fusion effectiveness, yielding more uniform energy profiles, lower redundancy, and extended network service life [24]. Selecting the most suitable optimization strategy, however, demands careful alignment with specific network traits—node mobility, population density, device diversity, and resource scarcity—highlighting the necessity for lean, responsive, real-time algorithms customized to particular IoT use cases [25].

Routing mechanisms are likewise indispensable, governing the efficient transit of information packets from SNs to DFHs and then to sink or relay points [26, 27]. Well-designed routing ensures consistent delivery, curbs power usage, and prevents overburdening individual nodes. In contrast, flawed routing protocols can produce load imbalances, prolonged delays, and wasteful energy expenditure. Routing schemes that dynamically account for node-to-sink distance, current RE, and aggregation opportunities are vital for preserving performance amid fluctuating conditions [28].

Ultimately, proficient DFH election, optimized DFG organization, and flexible routing form the essential pillars for resolving IoT network difficulties. Delivering power-efficient, trustworthy, and highly scalable architectures is critical to realizing IoT's transformative impact across smart urban infrastructure, industrial control systems, healthcare, and beyond. Because these issues are deeply interconnected, only integrated approaches that jointly optimize data fusion grouping and routing can provide comprehensive solutions. Although AI- and machine-learning-driven innovations hold great promise for tackling these problems, their effective deployment in resource-scarce, constantly changing IoT contexts still requires substantial investigation.

The present work seeks to address these issues by examining leading-edge methods for DFH selection, data-fusion group optimization, in-network fusion, and dynamic routing. By combining advanced optimization strategies with rigorous assessment in realistic settings, this research aims to develop efficient, scalable, and highly adaptive IoT solutions that enable maximum utility across numerous application domains.

Research gaps, our methodology, and contributions

Significant shortcomings remain in the IoT networking domain, particularly in clustering mechanisms, optimization strategies, and data fusion processes used in the Data Fusion Head (DFH) election and routing design. Present-day clustering and fusion approaches frequently exhibit limited scalability and poor responsiveness to shifting network dynamics, which leads to suboptimal resource allocation and shorter overall network durability. Furthermore, current optimization techniques for DFH selection require further improvement to achieve better energy load balancing, equitable task distribution across nodes, and stable connectivity. Many existing solutions fail to adequately address heterogeneous deployments, where Sensor Nodes (SNs) differ widely in residual energy and functional capabilities. Routing protocols also struggle to handle the variable, unpredictable traffic patterns common in IoT use cases. A large proportion of these protocols cannot perform real-time route adaptation, which often results in elevated end-to-end delays and unnecessary power drain. Overcoming these limitations requires the development of more sophisticated, flexible algorithms that deliver consistent performance across diverse IoT settings while prioritizing reliability, energy conservation, and extended service lifetime.

There is growing attention to leveraging Computational Intelligence (CI) and data fusion techniques to design routing solutions for IoT systems. Such applications demand straightforward, dependable, and cost-effective communication frameworks. At the same time, enhancing network efficiency constitutes a highly intricate optimization challenge. To tackle these issues, nature-inspired evolutionary algorithms—including Particle Swarm Optimization (PSO) [29] and Ant Colony Optimization (ACO) [22]—have become increasingly prominent for identifying efficient routing paths. These methods iteratively evaluate a population of candidate solutions, compute the Data Fusion Fitness Function (DFFF), and compare each candidate against the current global best to converge on superior routing decisions. The precise formulation of the DFFF plays a decisive role in determining high-quality DFH choices. In much prior work, DFFFs have relied on fixed or randomly assigned weighting coefficients, without accounting for the spatial distribution of SNs or their current remaining energy (RE). Moreover, selecting the most suitable optimization method itself presents a non-trivial decision.

Network performance degradation becomes more pronounced as both the physical scale of the deployment and the total number of deployed SNs increase. In response, the IoT research community has developed various energy-aware routing protocols to enable prolonged autonomous operation—particularly in challenging or remote environments—without frequent battery intervention. Against this backdrop, the central goal of the present study is to introduce a practical routing protocol that improves overall network performance and addresses the deficiencies identified in earlier proposals. Only a handful of metaheuristic algorithms have achieved widespread adoption, largely because many suffer from high computational demands and poor energy profiles, which prevent them from guaranteeing sufficient remaining energy to prolong network lifetime meaningfully. For this reason, careful selection of the most fitting metaheuristic becomes essential.

Notably, the current work directly addresses the weaknesses of prevailing clustering and data fusion schemes while adopting a more effective metaheuristic—the Marine Predators Algorithm (MPA)—to perform Primary DFH (PDFH) selection with a strong focus on energy efficiency [30]. Within the proposed framework, network lifetime is distributed uniformly across operational rounds. The initial round encompasses both the setup and steady-state phases, whereas subsequent rounds focus exclusively on steady-state activities related to data aggregation and forwarding. The setup phase may be re-executed whenever PDFH reselection (regrouping) becomes necessary. A distinctive characteristic of this protocol is that SNs remain capable of sensing and transmitting data continuously throughout each Round Time (RT), enabling professional-grade data fusion handling.

The network sensing field is modeled as a rectangular area in which SNs are randomly deployed, and the sink node initiates the process by broadcasting an information request to all SNs. To reduce control message overhead during DFG formation, the sink performs the setup phase only once at the outset. In reply, each SN transmits a message containing its geographic coordinates and current energy level. Using this information, the sink partitions the sensing region into equal-sized hexagonal DFGs. It then designates a PDFH for each cluster.

In summary, the innovative protocol introduced in this paper implements a data-fusion-centric approach in which the MPA—whose popularity has surged in recent years, evidenced by its 2029 citations on Google Scholar as of February 8, 2025—selects the most suitable PDFH for each group. The protocol's originality lies in the design of its DFFF, which incorporates critical factors, including the average separation among DFG Members (DFGMs), their individual RE values, the distance from each DFGM to the sink node, the Average Scale of Building Occlusions (ASBO), and the PDFH rotation frequency. Crucially, each parameter receives a tailored weighting coefficient derived through extensive trial-and-error experimentation involving a large number of simulation runs, with the final weights chosen to maximize overall network performance. This carefully crafted DFFF serves as the objective function that the MPA seeks to minimize. In other words, the configuration yielding the smallest combined value across these weighted parameters represents the optimal PDFH choice. Beyond the PDFH election, the proposed protocol operates in a batch-oriented manner. This means that regrouping (PDFH reselection) is not performed in every round but only after completing a batch of several rounds. To enable this mechanism, the remaining energy (RE) of the currently selected PDFH is compared against the average RE of the DFG Members (an established energy threshold). Should the PDFH's RE drop below this threshold, the regrouping

process is triggered: the PDFH broadcasts a disjoin message to the sink node and all Sensor Nodes (SNs) across the network, signaling the initiation of PDFH reselection and cluster reformation.

For data aggregation and forwarding, the proposed protocol offers two variants. In the version without multi-hop routing, SNs transmit their readings directly to their assigned PDFH. The PDFH then performs data fusion on the collected information and forwards the aggregated result straight to the sink node using single-hop communication. To coordinate transmissions efficiently, each SN is assigned a dedicated Time Slot (TS) via an innovative Time Division Multiple Access (TDMA) scheduling scheme.

In contrast, the multi-hop routing variant operates as follows: SNs first send data to their respective PDFHs, which aggregate the incoming information. If the distance between a PDFH and the sink node is shorter than the crossover distance (87 m), the PDFH transmits the fused data directly to the sink (single-hop mode). When this distance exceeds the crossover threshold, the protocol activates a state-of-the-art Relay Data Fusion Cost Function (RDFCF) to select appropriate Primary Relays (PRs) and forward the data hop-by-hop until it reaches the sink node (multi-hop mode). The RDFCF is constructed from three key influencing parameters: the distance between the candidate PDFH and potential upper-layer PDFHs, the remaining energy (RE) of those candidates, and the Average Scale of Building Occlusions (ASBO). Because the objective is to choose the relay PDFH exhibiting the lowest RDFCF value, the function is formulated to be minimized.

The concepts of relay PDFHs and the RDFCF bring several advantages: they substantially reduce control message overhead, shorten the effective communication distance to the sink, and intelligently alleviate congestion around PDFHs located near the sink by more evenly distributing forwarding responsibilities. This load-balancing strategy ensures that SNs positioned far from the sink conserve energy more effectively. Collectively, the described techniques significantly reduce energy consumption and markedly extend the network's operational lifetime.

To the best of our knowledge, no prior publication in the literature has addressed or encompassed the distinctive innovations presented in this work—most notably the seamless integration of advanced data fusion techniques into clustered IoT architectures—which sets this proposal apart from existing approaches. In view of this uniqueness, the principal contributions of the paper are outlined below:

- Introduction of a novel, adaptive, and dynamic method for dividing the network into Data Fusion Groups (DFGs) using a hexagonal topology. This division process automatically adapts to the current placement and density of SNs, delivering optimal area coverage, equitable node distribution, and excellent scalability that naturally aligns with the evolving deployment patterns of IoT sensor networks.
- Presentation of a highly efficient DFG formation mechanism capable of continuous adaptation to fluctuating network densities. The clusters dynamically reorganize in response to real-time changes in conditions, thereby optimizing resource usage, balancing computational and communication loads, and preserving low latency alongside robust performance.
- Development and implementation of an advanced selection strategy—powered by the Marine Predators Algorithm (MPA)—for identifying the optimal Primary DFH (PDFH), First Secondary DFH (FSDFH), and Second Secondary DFH (SSDFH) within each DFG. The method employs a sophisticated Data Fusion Fitness Function (DFFF) that incorporates multiple critical factors: distance of SNs to the sink node, individual RE levels, average inter-DFGM separation, ASBO, and PDFH rotation frequency. This comprehensive evaluation effectively safeguards SN's energy reserves and substantially extends the overall network's longevity.
- Introduction of a completely new parameter, the Average Scale of Building Occlusions (ASBO), which—according to the authors' review of the literature—has never before been incorporated or thoroughly explored within DFFF or RDFCF formulations.

- For multi-hop relaying, deployment of an innovative Relay Data Fusion Cost Function (RDFCF) to elect the Primary Data Fusion Relay (PDFR), First Secondary DFR (FSDFR), and Second SDFR (SSDFR). Selection relies on the candidate's RE, its distance to both the sink node and the originating PDFH, and ASBO values, enabling efficient identification of optimal relays—especially those farther from the sink—and delivering measurable energy savings for PDFHs.
- Optimization of resource allocation and effective mitigation of data redundancy through the application of data fusion techniques that handle device/link failures and recovery, redundancy elimination, and building occlusion effects. Particular attention is given to redundancy issues arising among spatially proximate SNs; an efficient fusion management strategy is proposed to ensure protocol coherence and performance. Specialized recovery procedures are integrated to manage failures robustly, while the impact of building occlusions at different heights is analyzed together with targeted countermeasures.
- Provision of a clear, step-by-step illustrative example demonstrating the MPA-based selection processes for PDFH, FSDFH, and SSDFH within the proposed protocol, thereby improving reader comprehension and practical insight into these mechanisms.
- Thorough performance assessment of the proposed protocol via extensive simulation experiments, evaluating essential metrics such as network lifespan, energy consumption, throughput, and average end-to-end delay. Detailed comparative analyses against state-of-the-art protocols highlight the superior effectiveness of both the single-hop and multi-hop variants of the proposed scheme.

Related Works

This section surveys the relevant existing literature. Over the past two decades, the proliferation of the Internet of Things (IoT) has attracted intense interest from researchers in numerous fields, profoundly impacting many aspects of everyday life. The fundamental principle of IoT is to connect virtually every object, enabling people to interact with the physical world through diverse Sensor Nodes (SNs). In IoT deployments, where SNs are typically battery-powered and deployed in remote or hard-to-reach locations, implementing energy-efficient techniques is essential given their limited power supply. Minimizing energy usage is therefore critical to overcoming this constraint. A wide range of strategies has been developed to extend the operational lifetime of IoT systems and improve overall network durability. Among these, Cluster Head (CH) selection is a key method for maintaining network stability: SNs are grouped into clusters across the sensing area using machine learning, and CHs are elected from within those clusters based on carefully defined metrics. The growing application of Computational Intelligence (CI) in Wireless Sensor Networks (WSNs), IoT ecosystems, and crowd-sensing systems has proven particularly valuable for accomplishing this goal [31].

Extensive research has investigated clustering approaches that rely on metaheuristic optimization algorithms to manage uncertainties inherent in WSNs. Although many CH election schemes have been proposed, achieving optimal energy efficiency remains a major unresolved issue in IoT environments.

Darabkh *et al.* [32] introduced two advanced CH selection protocols: Energy and Distance-Based CH Selection (EDB-CHS) and EDB-CHS with Balanced Objective Function (EDB-CHS-BOF). Both protocols aim to prevent early node failures, correct uneven energy depletion among SNs, and significantly prolong the entire network's lifetime. For the EDB-CHS protocol, the authors adopt a hexagonal clustering layout and derive a closed-form mathematical expression to determine the optimal number of CHs. They also present an efficient CH election mechanism driven by a threshold probability that incorporates several important factors: each SN's likelihood of becoming a CH, its proximity to the sink node, and its remaining energy (RE). In contrast, the EDB-CHS-BOF protocol specifically targets long-range inter-CH communications. It defines a new per-round threshold probability for CH candidacy and introduces a balanced objective function to promote a more even spatial distribution of CHs throughout the network. Experimental results demonstrate

that both proposed protocols outperform competing methods by a substantial margin, particularly in terms of throughput and network longevity.

Heidari [33] sequentially applies two metaheuristic algorithms for clustering and routing. They first use the Whale Optimization Algorithm (WOA) to form clusters and identify optimal CHs through a procedure named WOA-clustering. Routing from these CHs to the sink node is then handled by the Harris Hawks Optimization (HHO) algorithm, implemented as HHO-Routing. This combined strategy markedly lowers the energy required to deliver data to the sink. To evaluate the effectiveness of their approach, the authors compare it against several established algorithms. Simulation results indicate clear improvements in data delivery rates, overall energy consumption, and the number of rounds until the first node depletes its energy.

Driven by the imperative to improve energy efficiency and prolong operational duration in underwater Wireless Sensor Networks (WSNs), He *et al.* [34] proposed an energy-aware clustering and multi-hop routing protocol that leverages metaheuristic optimization algorithms. Nevertheless, most existing metaheuristic approaches use separate algorithms for clustering and multi-hop path selection, leading to greater computational complexity, inconsistent initialization routines, and difficulties in adjusting hyperparameters. To overcome these drawbacks, the authors introduced a new hierarchical framework, the Hierarchical Chimp Optimization Algorithm (HChOA), that unifies clustering and multi-hop routing. The effectiveness of HChOA was rigorously tested across multiple performance indicators through detailed simulations and benchmarked against several established protocols. The findings indicate that HChOA delivers superior results in terms of network lifetime and overall energy expenditure.

In a related effort, Darabkh *et al.* [35] developed innovative techniques to reduce total network energy usage while accounting for multiple critical operational parameters. Their clustering mechanism evenly partitions the network into balanced groups, with each cluster electing a Cluster Head (CH) via Particle Swarm Optimization (PSO) that considers essential criteria such as a Sensor Node's (SN's) remaining energy (RE) and its distance to the Mobile Sink (MS). In addition, the protocol coordinates the MS's mobility pattern to achieve uniform energy depletion across all SNs. This is accomplished by employing a circular path with an adaptive radius and fixed angular speed: the MS begins at the network center. She moves outward and inward along radial lines. Such a trajectory ensures that nodes closer to the sink consume less energy than those farther away, thereby boosting energy efficiency and network performance. The MS also periodically broadcasts its starting location, allowing each SN to estimate the sink's current position at any moment. Moreover, clustering control overhead is greatly reduced because the setup phase is triggered only rarely. Extensive simulation experiments confirm that the proposed protocol markedly outperforms comparable schemes in network longevity and power consumption across random numbers of rounds.

Another contribution comes from Somula *et al.* [36], who presented SWARAM (Osprey Optimization Algorithm-based Energy-Efficient CH Selection), a new optimization method tailored for CH selection in WSNs within IoT contexts. The SWARAM procedure consists of two main stages: initial cluster formation based on Euclidean distances among nodes, followed by CH selection using the Osprey Optimization Algorithm. Performance evaluations comparing SWARAM against other CH selection strategies show that it improves packet delivery ratio and network lifetime by 10% each, leading to noticeably better overall network efficiency.

Ensemble clustering has proven highly effective in unsupervised learning tasks for many years, yet conventional methods relying on Co-Association (CA) matrices continue to exhibit important weaknesses. These approaches typically emphasize only intra-cluster pairwise associations while overlooking useful inter-cluster relationships that could provide richer data insights. They also depend on separate downstream clustering algorithms (such as spectral clustering) to extract final partitions from the CA matrix and lack mechanisms to assign adaptive weights to different base clusterings, treating each base result as equally valuable, regardless of its quality. To resolve these issues, Bian *et al.* [37] introduced WEC-FCA, an advanced Weighted adaptive Ensemble Clustering technique built around a Fuzzy Co-Association (FCA) matrix. The FCA matrix fundamentally improves ensemble clustering by simultaneously capturing both intra- and inter-cluster associations, yielding a more comprehensive data representation. By embedding a rank constraint into the FCA

matrix, the authors devised an optimization model that directly produces the best ensemble FCA matrix, naturally consistent with the actual number of clusters, and removes the reliance on external clustering tools. In addition, WEC-FCA applies Shannon entropy to dynamically weight each base clustering outcome, prioritizing the most informative contributions throughout the optimization. Experiments on synthetic and real-world datasets consistently show that WEC-FCA matches or exceeds the performance of leading contemporary methods while effectively overcoming the core limitations of traditional ensemble techniques.

With the rapid evolution of next-generation communication systems, precise location data has become indispensable for delivering context-sensitive, geographically aware smart services. Modern Intelligent Transportation Systems (ITS) impose especially rigorous requirements for real-time, high-accuracy, and privacy-conscious positioning information. To meet these demands, Zhou *et al.* [38] proposed the Spatial–Temporal Federated Transfer Learning (ST-FTL) framework, which fuses multi-sensor measurements with state-of-the-art privacy-preserving methods to advance collaborative positioning in dense urban ITS environments. The framework adopts a three-tier architecture that merges Federated Learning (FL) with Transfer Learning (TL), resulting in quicker global model convergence, lower communication overhead, and higher prediction precision—even when local datasets are sparse, as is common in urban street canyons. A density-driven spatial–temporal clustering method identifies the most suitable source domains for knowledge transfer, guaranteeing domain similarity and boosting cross-area model effectiveness. The framework further incorporates a convolutional-gated unit to discard irrelevant features and construct new, valuable representations, thereby improving global model initialization and weight aggregation. On the device side, a multi-sensor fusion model integrates GPS and inertial measurement unit readings using an enhanced time-aware asymmetric attention mechanism that adaptively weights each sensor modality based on its contextual importance. Additionally, a lightweight Siamese network architecture enables efficient data augmentation by leveraging each vehicle's historical trajectories, mitigating the problem of limited training samples. Evaluations conducted on two publicly available datasets confirm that ST-FTL outperforms leading approaches, delivering higher positioning accuracy, faster convergence, and reliable performance under typical urban ITS conditions.

Rawat and Chauhan [39], Choudhary *et al.* [40], Giri *et al.* [41], and Huangshui *et al.* [42] discuss studies highly relevant to the protocol we propose.

In Rawat and Chauhan [39], the authors present a particle swarm optimization (PSO)-based energy-efficient clustering protocol (PSO-EEC) designed to extend network lifetime. This protocol applies the PSO algorithm to choose both Cluster Heads (CHs) and relay Sensor Nodes (SNs) throughout the network. CH selection relies on a Fitness Function (FF) generated via PSO, which evaluates the SN's energy ratio (comparing initial energy to remaining energy (RE)), the distance between SNs and potential CHs, and the node degree to identify the most appropriate candidate for the CH position. For forwarding aggregated data to the sink node, PSO-EEC uses another fitness metric based on the CH's RE and its distance to the sink to select suitable relay nodes for multi-hop communication. Simulation experiments show that PSO-EEC surpasses previous methods across key performance indicators, including energy usage, network longevity, and throughput.

Choudhary *et al.* [40] introduce a linearly decreasing inertia weight PSO-based Clustering technique (LDIWPSOC) to form high-quality clusters. In LDIWPSOC, PSO begins with a particle encoding scheme where each particle represents the indices of SNs chosen as CHs, using their spatial coordinates. During initialization, particle positions and velocities are assigned randomly. An FF is then formulated that incorporates intra-cluster distance, the distance from CHs to the sink node, and the RE of the selected CHs to search for both personal and global best solutions. Particle positions and velocities are updated using a linearly decreasing inertia weight, which promotes broad exploration early in the process and gradually shifts toward precise exploitation. This mechanism helps achieve greater search precision and enables rapid convergence to the global optimum while avoiding local optimum traps. However, nodes with highly uneven neighbor distributions may still be selected as CHs, potentially leading to excessive energy consumption within clusters.

In Giri *et al.* [41], an Optimized Fuzzy Clustering Algorithm (OFCA) is proposed for CH election, combined with a routing mechanism to deliver data to the sink node. CHs in OFCA are determined using a type-1 fuzzy logic system that considers distance to the sink node, remaining energy (RE), and node concentration (density). For energy-efficient inter-cluster routing, PSO is employed to select relay CHs. The PSO Fitness Function is designed to identify the local and global best particle positions, thereby prolonging network lifetime. Following standard PSO procedure, particles are first initialized and then iteratively updated until one particle achieves the optimal fitness value, marking the global best solution. As a result, relay CHs are selected from those with higher node concentration, lower distance to the sink, and superior RE levels. The PSO-based routing in OFCA identifies optimal relay points to create efficient data-forwarding paths, thereby reducing the energy burden on CHs during inter-cluster communication and extending the overall network duration. Nevertheless, relying on a single fixed path from source to destination may compromise Quality of Service (QoS) guarantees. Simulation results indicate that OFCA delivers a longer network lifetime and supports higher message delivery rates to the sink than competing approaches.

Huangshui *et al.* [42] present a novel PSO-based clustering and routing protocol called NPSOP, with the primary goal of maximizing network lifespan while achieving energy efficiency and balanced load distribution. NPSOP uses PSO to select CHs simultaneously and define routing paths for each CH, constraining particle components using relevant metrics (RMs), such as distance to the sink node, remaining energy (RE), and node centrality, to accelerate convergence. A composite Fitness Function evaluates particle quality by accounting for both energy consumption and network load balancing. An adaptive inertia weight mechanism is also incorporated to update particle states, helping avoid local optima and guiding the search toward the global optimum through iterative refinement. Comprehensive experimental comparisons assess NPSOP against prior methods using metrics including energy consumption, throughput, network lifetime, the standard deviation of residual energy, and load distribution. The results reveal that NPSOP extends network lifetime by 29.94% compared to PSO-EEC, 24.16% compared to LDIWPSO, and 13.67% compared to OFCA. Similarly, it reduces energy consumption by 24.08% compared to PSO-EEC, 19.16% compared to LDIWPSO, and 10.95% compared to OFCA.

In addition to the discussion above, Table 1 summarizes the protocols covered in Rawat and Chauhan [39], Choudhary *et al.* [40], Giri *et al.* [41], and Huangshui *et al.* [42] that are most closely aligned with our work and highlight the distinctive features of our proposed methodology relative to these earlier contributions. The comparison focuses on critical dimensions addressed in those studies, including network area partitioning method, CH selection approach, parameters included in the Data Fusion Fitness Function (DFFF), clustering/data fusion grouping strategy, number of clusters/DFGs, inter-cluster communication type, scheduling mechanism, RDFCF parameters, and the presence of illustrative examples demonstrating the CH selection process.

Table 1. A comparative analysis table of closely related studies focusing on similar contribution aspects as in our proposed work.

Protocols	Primary objectives	PDFH selection algorithm	DFFF parameters (key factors in fitness function)	Data fusion grouping scheme	Number of DFGs	RDFCF parameters (relay selection criteria)	Handles impairments	Novel aspects
PSO-EEC [39]	Extend overall network lifetime	PSO	Ratio of each SN's initial energy to remaining energy (RE), distance from DFGMs to	Distributed	Predefined	Remaining energy of PDFH, the distance separating PDFH from the sink node	X	X

			PDFH, and degree (connectivity) of SN					
LDIWPSOC [40]	Maximize energy efficiency	PSO	SN's remaining energy (RE), mean distance between PDFH and its associated DFGMs, and distance of PDFH to the sink node	Hybrid	Predefined	Distance to the nearest PDFH	✗	✗
OFCA [41]	Prolong network operational lifetime	Fuzzy logic	Aggregate remaining energy of chosen PDFHs, average distance from selected PDFHs to the sink node, and total concentration of chosen PDFHs	Centralized	Predefined	SN's remaining energy, distance to the sink node, and concentration level of the PDFH	✗	✗
NPSOP [42]	Maximize network survival duration	PSO	SN's remaining energy (RE), the mean distance between PDFH and its DFGMs, and the distance of PDFH from the sink node	Distributed	Predefined	Energy expenditure levels and load distribution balance	✗	✗
Proposed protocol (without routing)	Prolong network lifetime, minimize total energy	MPA	SN's remaining energy (RE), average distance among	Hybrid	Dynamic	—	✓	✓

	usage, ensure balanced load distribution		DFGMs, distance to the sink node, PDFH rotation frequency, ASBO factor					
Proposed protocol (with multi-hop routing)	Prolong network lifetime, minimize total energy usage, ensure balanced load distribution	MPA	SN's remaining energy (RE), average distance among DFGMs, distance to the sink node, PDFH rotation frequency, ASBO factor	Hybrid	Dynamic	Distance between PDFH and candidate PDFR, distance between candidate PDFR and the sink node, remaining energy of candidate PDFR, ASBO factor	✓	✓

Hybrid means that the clustering schemes have both centralized and distributed algorithms. (✓) indicates that an aspect is fully addressed in the corresponding article. (x) indicates that an aspect is not covered.

Fundamentals of Marine Predators Algorithm and Its Practical Applications

The Marine Predators Algorithm (MPA), first presented in 2020 [43], is a biologically inspired, population-based metaheuristic for optimization. It draws inspiration from the foraging behavior of marine predators, particularly their use of Lévy flights and Brownian motion, as well as the established predator–prey encounter rates and optimal foraging strategies observed in ocean ecosystems [44]. Since its introduction, MPA has rapidly gained favor over several well-known optimization approaches, including Genetic Algorithms (GA) [20], Particle Swarm Optimization (PSO) [21], and Ant Colony Optimization (ACO) [22]. Table 2 provides a side-by-side comparison of MPA with these methods, emphasizing its advantages in search effectiveness, convergence speed, robustness to parameter changes, and flexibility across practical scenarios.

MPA has been successfully applied to a broad range of continuous optimization challenges in diverse fields such as image segmentation, fog/edge computing resource allocation, photovoltaic parameter estimation, hybrid wind–solar energy system modeling, various engineering design problems, medical data classification, task scheduling, sentiment analysis, and feature selection tasks [45]. Consistent with other metaheuristic algorithms, MPA iteratively improves a collection of candidate solutions, starting from random initialization and progressively refining them through structured update rules to converge toward the global optimum [46].

Table 2. A comparison table of the MPA with different optimization methods regarding various features.

Feature	MPA (marine predators algorithm)	GA (genetic algorithm)	PSO (particle swarm optimization)	ACO (ant colony optimization)
Biological/Natural inspiration	Hunting and foraging strategies of marine predators	Principles of natural selection, genetics, and evolution	Collective social behavior observed in bird flocks or fish schools	Foraging patterns and trail-following behavior of ants
Parameter dependency	Low – minimal parameters, making tuning straightforward and user-friendly	High – highly sensitive to mutation rate, crossover rate, and related genetic operators	Moderate – depends on careful adjustment of inertia weight, cognitive coefficient, and social coefficient	High – strongly influenced by pheromone evaporation rate, heuristic parameters, and trail reinforcement factors
Core search mechanism	Combines Lévy flight for broad exploration, Brownian motion for local search, and adaptive predator-prey hunting phases	Relies on the selection of fitter individuals, crossover (recombination), and mutation operations	Updates particle velocity and position using personal best and global best information	Builds solutions through pheromone trails that guide probabilistic path selection, combined with heuristic information
Implementation complexity	Relatively low – straightforward structure with few operations and limited parameters	Moderate to high – involves more elaborate genetic operators (selection, crossover, mutation)	Moderate – conceptually simple velocity/position update rules, but requires proper parameter tuning	High – involves repeated pheromone updates, evaporation, and path construction over multiple iterations
Exploration capability	Strong – enhanced by adaptive mechanisms and Lévy flights that promote effective global search	Moderate – exploration level depends heavily on mutation probability	Strong – supported by random components in velocity updates and diversity from swarm particles	Moderate – driven mainly by pheromone evaporation, which gradually reduces focus on previously explored paths
Exploitation capability	Well-balanced – adaptive strategies shift focus to refine promising regions effectively	Moderate – tends to intensify search around high-fitness solutions after crossover phases	Good – particles converge toward known high-quality regions guided by the best solutions	Good – strong local refinement via pheromone reinforcement, but risks premature convergence to suboptimal areas
Convergence speed	Generally, fast—dynamic adaptation accelerates convergence without excessive stagnation	Moderate to slow – genetic operations often result in slower progress toward optima	Fast – especially when parameters are appropriately tuned, leading to rapid convergence	Slow to moderate – speed heavily depends on how quickly pheromone trails stabilize and evaporate

Scalability	High – maintains strong performance even on large-scale, high-dimensional optimization problems	Moderate – can face difficulties and increased computational cost on very large problem instances	High – scales effectively across a wide range of problem sizes and dimensions	Moderate – computational overhead grows significantly with larger problem sizes due to pheromone management
Main limitations	Relatively recent algorithm – still undergoing extensive validation across diverse problem domains	Susceptible to premature convergence and trapping in local optima	Can stagnate in local optima if parameter settings are suboptimal	Often exhibits slow convergence and a high risk of becoming trapped in local optima

The current section describes the development of MPA as a straightforward yet powerful metaheuristic for optimization. Its biological motivation is originally presented in Faramarzi *et al.* [43], while the mathematical formulation is explained in the subsections that follow.

The MPA's population initialization

MPA operates as a population-based algorithm, in line with most metaheuristic approaches. It starts by defining the initial algorithm parameters and then generates the first set of candidate solutions by uniformly distributing them throughout the search space according to the equation below [43]:

$$\vec{X} = \vec{X}_{min} + \vec{rand} \times (\vec{X}_{max} - \vec{X}_{min}), \quad (1)$$

Where $\vec{X}_{min}$ and $\vec{X}_{max}$ represent the lower and upper bounds of the decision variables in the optimization problem, and $\vec{rand}$ denotes a vector of uniformly distributed random numbers in the interval [0, 1].

$\vec{Prey}$ and $\vec{Elite}$ matrices construction and fitness evaluation

Inspired by the principle of survival of the fittest observed in natural systems, the most capable predators typically demonstrate the strongest hunting skills. In optimization algorithms, this concept translates to selecting the best-performing solution—treated as the top predator—to populate a structure called the $\vec{Elite}$ matrix, as expressed in Eq. 2. This matrix, composed of repeated vectors, enables predators to exploit positional information to locate and pursue prey more effectively [43].

$$\vec{Prey} = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,d} \\ X_{2,1} & X_{2,2} & \dots & X_{2,d} \\ \dots & \dots & \dots & \dots \\ X_{n,1} & X_{n,2} & \dots & X_{n,d} \end{bmatrix}, \quad (2)$$

Where $\vec{X_I}$ is the vector corresponding to the current best predator (top solution), which is replicated n times to construct the entire $\vec{Elite}$ matrix; here, n denotes the total number of search agents (population size), and d represents the problem dimensionality. Importantly, both predators and prey are modeled as search agents actively exploring the solution

space to find better resources. Thus, while predators hunt prey, the prey themselves are simultaneously searching for their own food sources. After each iteration, the $\overrightarrow{Elite}$ matrix is refreshed whenever a newly discovered solution outperforms the existing top predator.

A second matrix of identical size, termed $\overrightarrow{Prey}$, is maintained to update the predators' positions. At initialization, the $\overrightarrow{Prey}$ matrix is randomly generated, and the fittest individual among the prey (which becomes the initial top predator) is used to form the $\overrightarrow{Elite}$ matrix. The $\overrightarrow{Prey}$ matrix is structured as follows [43]:

$$\overrightarrow{Prey} = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,d} \\ X_{2,1} & X_{2,2} & \dots & X_{2,d} \\ \dots & \dots & \dots & \dots \\ X_{n,1} & X_{n,2} & \dots & X_{n,d} \end{bmatrix}, \quad (3)$$

Where X_i represents the j^{th} dimension (variable) of the i^{th} prey (search agent). It should be emphasized that the core mechanism of the entire optimization procedure is tightly coupled to these two matrices. Once the $\overrightarrow{Prey}$ matrix has been initialized, the Marine Predators Algorithm (MPA) computes the fitness of each prey individual using the problem's objective function.

The MPA's optimization process

The optimization procedure in MPA is divided into three distinct phases, each corresponding to different predator–prey velocity ratios and collectively mimicking the full behavioral lifecycle of predators and their prey. The iterations are distributed across these phases in accordance with natural movement patterns observed in marine ecosystems. The three phases are described below:

Phase 1: Exploration phase

When the velocity ratio is high ($V \geq 10$), meaning the predator moves significantly faster than the prey, this phase dominates the early portion of the search (roughly the first third of the total iterations), where broad exploration of the search space is most valuable. In such conditions, the most effective strategy for the predator is to remain still while the prey continues to move. This behavior is mathematically modeled as follows [43]:

$$\begin{aligned} &\text{While}(\text{Iter} < \frac{1}{3} \text{Max_Iter}) \\ &\overrightarrow{StepSize}_i = \vec{R}_B \otimes (\overrightarrow{Elite}_i - \vec{R}_B \otimes \overrightarrow{Prey}_i), \quad i = 1, \dots, n \end{aligned} \quad (4)$$

$$\overrightarrow{Prey}_i = \overrightarrow{Prey}_i + P \cdot \vec{R} \otimes \overrightarrow{StepSize}_i, \quad (5)$$

Where Iter and Max_Iter represent the current iteration counter and the total number of iterations, respectively. $\overrightarrow{StepSize}_i$ indicates the step size vector for the i^{th} predator at the current iteration, while $\vec{R}^B$ is a vector of random numbers drawn from a standard normal (Gaussian) distribution, simulating Brownian motion. The operator $\otimes$ denotes

element-wise (Hadamard) multiplication. In this equation, the term $\vec{R} \otimes \vec{B} \otimes \vec{Prey}$ effectively captures the prey's random movement. The parameter P is a fixed constant equal to 0.5, and $\vec{R}$ is a vector of uniformly distributed random numbers in the range $[0, 1]$.

Phase 2: Transition phase between the exploration and exploitation phases

In the phase where the velocity ratio is close to unity (predator and prey move at roughly the same speed), both parties are actively pursuing their objectives. This middle stage of the optimization process requires a careful balance between exploring new areas and refining promising solutions. Accordingly, the population is split evenly: one half prioritizes exploration while the other concentrates on exploitation. Within this framework, the prey performs exploitation-oriented movements, whereas the predator carries out exploration.

The underlying natural rule states that when the velocity ratio is near 1 ($V \approx 1$), the predator achieves its best outcome by following a Brownian motion pattern if the prey follows a Lévy distribution. Consequently, the Marine Predators Algorithm (MPA) implements Lévy-driven movement for the prey and Brownian motion for the predator.

$$\text{While}(\frac{1}{3} \text{Max_Iter} < \text{Iter} < \frac{2}{3} \text{Max_Iter})$$

For the first half of the population, MPA applies the following update rule [43]:

$$\vec{StepSize}_i = \vec{R}_L \otimes (\vec{Elite}_i - \vec{R}_L \otimes \vec{Prey}_i), \quad i = 1, \dots, n/2 \quad (6)$$

$$\vec{Prey}_i = \vec{Prey}_i + P \cdot \vec{R} \otimes \vec{StepSize}_i, \quad (7)$$

Where $\vec{R}^L$ is a vector of random numbers sampled from a Lévy distribution, representing Lévy-style movement. The term $\vec{R}^L \otimes \vec{Prey}$ simulates the Lévy motion of the prey, and adding the computed step size to the prey's current position produces its actual displacement. Because most steps in a Lévy distribution are short-range, this update predominantly supports local exploitation.

For the second half of the population, the update rule is formulated as follows [43]:

$$\vec{StepSize}_i = \vec{R}_B \otimes (\vec{R}_B \otimes \vec{Elite}_i - \vec{Prey}_i), \quad i = n/2, \dots, n \quad (8)$$

$$\vec{Prey}_i = \vec{Elite}_i + P \cdot CF \otimes \vec{StepSize}_i, \quad (9)$$

Where CF denotes the convergence factor—a dynamically adjusted parameter that controls the magnitude of predator movement across iterations. This factor is computed according to the formula [43]:

Phase 3: Exploitation phase

When the velocity ratio is low (predator significantly outpaces the prey), the process enters a strong exploitation-dominated stage, typically in the later portion of the optimization run. At a velocity ratio of $V = 0.1$, the predator's most

effective strategy is to adopt a Lévy movement pattern. This final phase is mathematically described as [43]:

$$\text{While}(\text{Iter} > \frac{2}{3} \text{Max_Iter}) \quad (10)$$

$$\overrightarrow{\text{StepSize}}_i = \overrightarrow{R}_L \otimes \left(\overrightarrow{R}_L \otimes \overrightarrow{\text{Elite}}_i - \overrightarrow{\text{Prey}}_i \right), \quad i = 1, \dots, n$$

$$\overrightarrow{\text{Prey}}_i = \overrightarrow{\text{Elite}}_i + P.CF \otimes \overrightarrow{\text{StepSize}}_i, \quad (11)$$

Here, the product $\overrightarrow{R}^L \otimes \overrightarrow{\text{Elite}}$ emulates the predator's Lévy-based movement, while adding the resulting step size to the $\overrightarrow{\text{Elite}}$ position simulates the predator's advance, thereby updating the prey's location toward more promising regions.

Eddy formation and fish aggregating devices' effect

Drawing on real-world predator behavior and environmental influences—such as eddy formation or the presence of Fish Aggregating Devices (FADs)—predators allocate 80% of their effort to searching locally around known productive areas and 20% to exploring distant regions that may contain different prey. This FAD's effect is modeled mathematically as follows [43]:

$$\overrightarrow{\text{Prey}}_i = \begin{cases} \overrightarrow{\text{Prey}}_i + CF \left[\overrightarrow{X}_{\min} + \overrightarrow{R} \otimes (\overrightarrow{X}_{\max} - \overrightarrow{X}_{\min}) \right] \otimes \overrightarrow{U}, & \text{if } r \leq \text{FADs}, \\ \overrightarrow{\text{Prey}}_i + [\text{FADs} (1 - r) + r] (\overrightarrow{\text{Prey}}_{r_1} - \overrightarrow{\text{Prey}}_{r_2}), & \text{if } r > \text{FADs}, \end{cases} \quad (12)$$

Where r is a uniformly distributed random number in $[0, 1]$, the parameter $\text{FADs} = 0.2$ reflects the degree of influence exerted by FADs on the position update. $\overrightarrow{U}$ is a binary vector containing 0s and 1s. $\overrightarrow{\text{Prey}}_{r_1}$ and $\overrightarrow{\text{Prey}}_{r_2}$ represents two randomly selected predators (solutions) from the current population.

Marine's memory saving

Marine predators are known for their remarkable ability to recall locations where they previously had successful hunts. MPA replicates this trait through a memory-saving mechanism. After updating the $\overrightarrow{\text{Prey}}$ matrix and applying the FADs effect, the fitness of each solution is evaluated once more, and the $\overrightarrow{\text{Elite}}$ matrix is refreshed accordingly. The fitness value for each solution in the current iteration is compared with its value from the previous iteration; the better-performing solution is retained. This comparison-and-replacement process steadily improves solution quality and emulates the natural tendency of predators to revisit proven foraging sites after successful encounters. The entire iterative procedure continues until a predefined stopping condition is satisfied (i.e., the maximum number of iterations, Max_Iter , is reached).

System Model

This section describes the network assumptions, the network architecture, and the energy consumption model employed in the proposed protocol.

Protocol's assumptions

Before presenting the detailed operation of the proposed protocol, it is essential to outline the foundational assumptions on which it is built. These assumptions are formulated to accommodate the diverse and application-specific nature of real-world IoT deployments.

1. The dimensions of the network sensing area are tailored to suit the requirements of the targeted IoT application.
2. A large number of IoT Sensor Nodes (SNs) are randomly scattered across a rectangular geographic region, with a single sink node deployed within the IoT network.
3. The sink node is located outside the perimeter of the sensing field, has an unlimited power supply, and is capable of broadcasting messages that can reach every SN throughout the entire network area.
4. All IoT SNs and the sink node are assigned unique identifiers and remain stationary after deployment.
5. Every SN is homogeneous, meaning all nodes begin with the same initial energy level and possess identical processing and communication capabilities, although each has a limited total energy budget.
6. Each SN is equipped with a GPS module, enabling it to determine its own location as well as the coordinates of the sink node and all other SNs in the network.
7. A Sensor Node is considered dead once its energy reserve is fully depleted.
8. SNs can dynamically adjust their transmission power based on the estimated distance to the intended receiver.
9. Specific Time Slots (TSSs) are assigned to SNs for transmitting data to their Data Fusion Head (DFH) during the data fusion process. Within these allocated slots, SNs are assumed always to have data ready for transmission, as the protocol operates in a data-driven (continuous sensing) mode rather than an event-triggered mode.
10. The distance separating a transmitting node from a receiving node can be reliably estimated using the strength of the received radio signal.

Energy consumption model

In IoT networks, the dominant sources of energy expenditure are wireless communication activities—namely, the transmission of data from SNs to their respective Data Fusion Heads (DFHs) and from DFHs to the sink node or to other DFHs (in the case of multi-hop routing)—along with the computational overhead associated with data fusion. During data fusion, each DFH collects readings from multiple SNs, performs aggregation operations (such as filtering, compression, and merging), and forwards a reduced dataset, thereby incurring additional energy costs for processing.

To model these energy requirements accurately, the well-established radio energy dissipation and propagation models originally presented in Darabkh *et al.* [47] and Shahzad *et al.* [48] are adopted for all simulations. The energy consumed by a transmitter to send an R -bit packet over a distance d is given by Darabkh *et al.* [47] and Shahzad *et al.* [48]:

$$E_{TX}(R, d) = \begin{cases} R \times E_{elec} + R \times \epsilon_{fs} \times d^2, & \text{if } d < d_0, \\ R \times E_{elec} + R \times \epsilon_{mp} \times d^4, & \text{if } d \geq d_0. \end{cases} \quad (13)$$

When an R -bit packet is received, the receiver expends the following amount of energy [47, 48]:

$$E_{RX}(R, d) = E_{elec}(R), \quad (14)$$

In Eqs. 13 and 14, E_{elec} (expressed in nJ/bit) represents the energy required by the transmitter or receiver circuitry to process one bit of data. The terms ϵ_{fs} and ϵ_{mp} (in pJ/bit/m² and pJ/bit/m⁴, respectively) denote the energy consumed per bit by the power amplifier in the free-space and multipath fading propagation models, respectively. The variable d indicates the separation distance between transmitter and receiver. The threshold distance (also known as the crossover distance d_0), which determines the switch between the two propagation models, is calculated as follows [47, 48]:

$$d_0 = \sqrt{(\epsilon_{fs} / \epsilon_{mp})}. \quad (15)$$

In addition, the energy expended for data aggregation (fusion) at the DFH level is modeled by the expression:

$$E_{DA} = N \times E_{fusion}, \quad (16)$$

Where N is the number of SNs whose data are aggregated by the DFH, E_{fusion} is the energy cost (in nJ/bit/signal) required to fuse or process one bit of incoming data, and E_{DA} represents the total energy consumed for performing data aggregation. Therefore, when each of the N SNs sends an R -bit data packet to the DFH, the overall energy required for aggregation is:

$$E_{DA} = N \times R \times E_{fusion}. \quad (17)$$

Network model

In the proposed protocol, Sensor Nodes (SNs) are randomly scattered across a square sensing field measuring $W \times W$. Each SN is assigned a unique identifier ranging from 1 to N , so the set of nodes is denoted as $SN = \{SN_1, SN_2, \dots, SN_n\}$. In addition to its position relative to the field center, each node maintains location information in the form (x_i, y_i) . A visual representation of the network model is provided in Figure 1. Data collected by individual SNs is forwarded to their designated Data Fusion Head (DFH), which then routes the aggregated information to the sink node located at a fixed position outside the sensing region.

The network's operational lifetime is organized into rounds. The initial round consists of both a setup phase and a steady-state phase. All subsequent rounds include only the steady-state phase. A key feature of the protocol is that regrouping (DFH reselection) occurs only after a batch of several rounds rather than every round. Therefore, the setup and steady-state phases are executed together in the round that follows a regrouping event. In contrast, the remaining rounds contain only the steady-state phase, as validated in Darabkh et al. [49]. The Round Timelines (RTs) for the proposed protocol—illustrating the efficient handling of data fusion—are depicted in Figure 2 (without routing) and Figure 3 (with multi-hop routing), respectively.

During the setup phase, the sink node first partitions the sensing area into hexagonal Data Fusion Groups (DFGs). Then it applies the Marine Predators Algorithm (MPA) to select the optimal DFHs for each group. This process is explained in greater detail later. The setup phase begins when the sink node broadcasts a Request Information Message to all SNs in the sensing field. In response, each SN_i unicasts an information message back to the sink node containing its unique ID, current Remaining Energy (RE), and geographic coordinates.

Upon collecting responses from all SNs, the sink node divides the sensing region into H equal-sized hexagonal groups, with the group radius dynamically adjusted based on the overall network scale. Each Data Fusion Group is symbolically represented as DFG_h and is uniquely identifiable by all nodes. After executing the MPA, the sink node compiles the resulting network configuration into a Table Message and broadcasts it to all SNs. This message enables each node to determine its role—specifically, whether it has been designated as DFH_i (where $1 < i < H$) for any particular DFG .

Additionally, the sink node calculates the relative Round Time. It assigns each SN_i a dedicated Time Division Multiple Access (TDMA) Time Slot (TS) for data transmission during the steady-state phase. In this phase, each SN_i transmits its data packet directly to its assigned DFH_i , then switches to sleep mode to conserve energy. The DFH_i aggregates the received data from all its Data Fusion Group Members (DFGMs) and forwards the fused result either directly to the sink node (in the protocol variant without routing) or to the next Primary Data Fusion Relay (PDFR) (in the multi-hop routing variant).

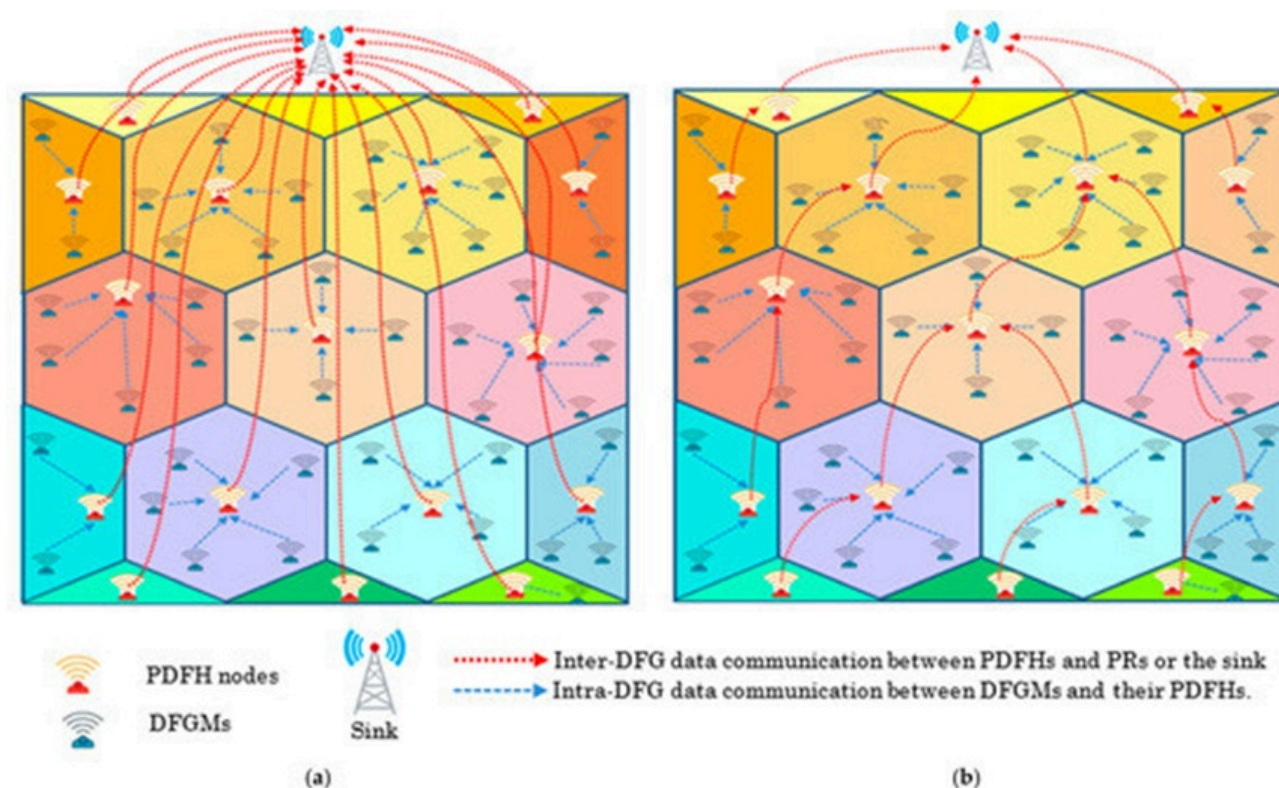


Figure 1. DFG model of an IoT network area for the proposed protocol: (a) DFG model of an IoT network area for the proposed protocol without routing, and (b) DFG model of an IoT network area for the proposed protocol with multi-hop routing.

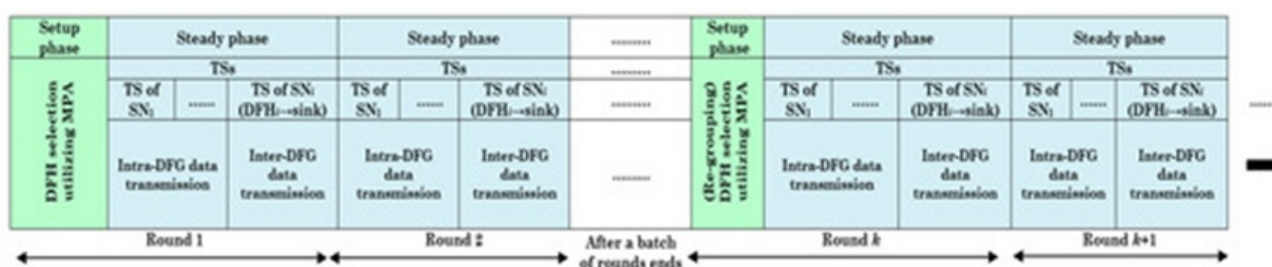


Figure 2. Breaking down the operational lifetime of the IoT network in the proposed protocol without routing.

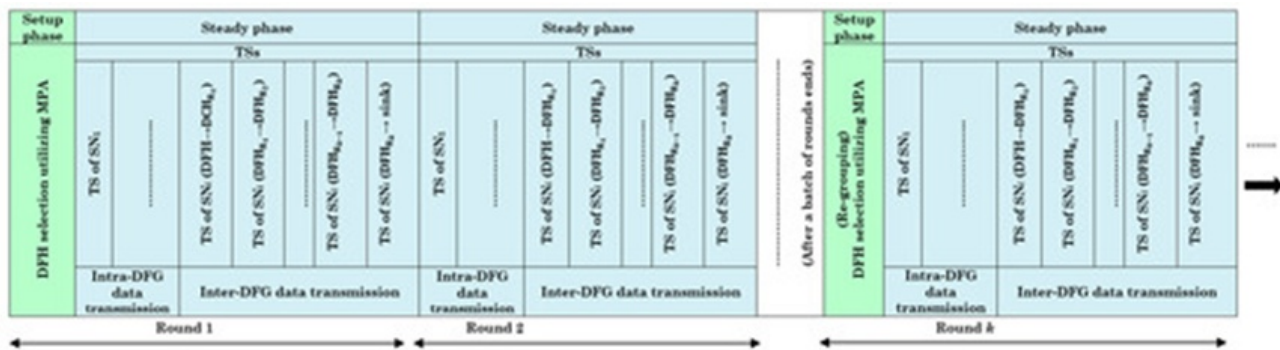


Figure 3. Breaking down the operational lifetime of the IoT network in the proposed protocol with multi-hop routing.

A Sensor Node SN_i is declared dead when its remaining energy becomes insufficient to transmit even a single packet to its DFH or the sink node. In such cases, the node is removed from the sink's network table. When a DFH's energy level falls below a predefined threshold, it broadcasts a Disjoin Message to all SNs in the sensing area, signaling the impending regrouping process. Simultaneously, the DFH sends a separate unicast Disjoin Message directly to the sink node, including details of its DFGMs and their current RE values. Upon receiving this notification, the sink node initiates network-wide synchronization and triggers the regrouping (DFH reselection) procedure.

The proposed protocol employs several distinct message types, which are described below:

- (1) Request Information Message: Broadcast by the sink node to all SNs, soliciting their status information.
- (2) SN's Information Message: Unicast reply sent by each SN_i to the sink node, containing the node's ID, geographic location, and current Remaining Energy (RE).
- (3) Sink Node's Table Message: Broadcast by the sink node after partitioning the area and executing MPA. This message contains comprehensive network configuration data and is used by each SN to determine its role. It exists in two variants:
 - DFGM Table Message: Includes fields specific to the receiving SN_i , such as SN ID, RE, node type, status, DFG number, DFG ID, distance to the sink node, distance to its DFH, and message arrival time (in seconds).
 - DFH Table Message: In the protocol variant without routing, it details the DFH and its associated DFGMs, listing SN ID, RE, node type, status, DFG number, distance to the sink node, and priority of the assigned TDMA Time Slot. In the multi-hop routing variant, the message includes the same fields, plus additional PDFR-related information: the next PDFR ID, the distance between the PDFR and its originating PDFH, RE status, DFG number, distance to the sink node, and the TDMA Time Slot priority.
- (4) Disjoin Message (to sink node): Unicast message sent by a DFH whose energy has dropped below the threshold. It informs the sink node that regrouping is required and includes the IDs and RE levels of the DFH's DFGMs as well as DFH-specific information.
- (5) Disjoin Message (to SNs): Broadcast by a DFH to all SNs in the network when its energy falls below the threshold, instructing them to stop transmitting data to that DFH and to prepare for regrouping, thereby maintaining network-wide synchronization.

The Proposed Protocol

One of the primary determinants of network longevity is the effectiveness with which individual nodes conserve their energy resources. Should the rate of energy depletion accelerate, the network would fall short of meeting its intended operational objectives much sooner than planned. The proposed protocol addresses this concern by designing the Primary Data Fusion Head (PDFH) selection mechanism to favor Sensor Nodes (SNs) that possess higher Remaining Energy (RE). This prioritization strategy enhances network stability and significantly extends its functional lifespan.

In addition, incorporating the distance from each SN to the sink node into the PDFH selection process further reduces energy expenditure. By preferentially assigning the PDFH role to nodes located closer to the sink, the protocol minimizes the occurrence of high-cost, long-range transmissions. Another vital factor incorporated into the selection criteria is the average distance between Data Fusion Group Members (DFGMs) within each group. Collectively, these considerations position the proposed protocol as a promising approach for optimizing energy usage across the entire network. Promoting balanced energy consumption among nodes helps prolong the operational life of IoT deployments and reduces the likelihood of premature PDFH failures.

The PDFH selection is accomplished by applying the Marine Predators Algorithm (MPA) to evaluate and minimize a carefully designed Data Fusion Fitness Function (DFFF). The growing popularity of MPA in the research community stems from several key strengths: its conceptual simplicity, broad applicability, realistic computational efficiency, fast convergence behavior, strong overall performance, and superior ability to tackle continuous, multi-objective, and binary optimization challenges compared with many established metaheuristics. To highlight the practical value of MPA, we previously introduced its first application in an IoT context, closely aligned with the present study, and reported initial experimental results [50, 51]. A notable distinction between those earlier efforts lies in the dimensionality of the fitness functions employed: one used a three-dimensional fitness function [50], while the other used a five-dimensional one [51].

For the proposed protocol variant without multi-hop routing, once the PDFHs have been selected, each SN forwards its collected data directly to its assigned PDFH. The PDFH then aggregates this incoming data along with its own measurements and transmits the fused result straight to the sink node, as illustrated in **Figure 1a** for the non-routing case. In contrast, the multi-hop routing variant operates differently. After aggregation, the PDFH forwards the fused data to an appropriate Primary Relay (PR) selected via the Relay Data Fusion Cost Function (RDFCF), as depicted in **Figure 1b**.

Network area division (DFGs Division)

Hexagonal-based data fusion grouping provides uniform spatial coverage, mitigates boundary-related inefficiencies, and improves overall data management by eliminating orphaned SNs. Circular grouping schemes, by comparison, tend to produce coverage gaps, uneven neighbor distributions, and suboptimal space utilization. Hexagonal tessellations, on the other hand, deliver clearer adjacency relationships and more consistent connectivity, making them particularly well-suited for organizing IoT sensor deployments.

Efficient management of Data Fusion Groups (DFGs) in IoT networks enhances scalability, reduces energy waste, and conserves limited battery resources. By performing in-network aggregation at a designated leader (the DFH) rather than requiring every SN to transmit directly to distant destinations, the protocol substantially lowers both communication volume and power consumption—critical advantages for resource-constrained devices.

The entire sensing field, which measures $W \times W$ meters, is subdivided into equal-sized hexagonal DFGs arranged in a pointy-top orientation, following an approach analogous to that described in Darabkh *et al.* [32]. Density metrics are computed both globally (across the full network) and locally (within each group). The sink node dynamically adjusts the size and placement of these DFGs to achieve a balanced relationship between overall network density and per-group density. Whenever a mismatch is detected, appropriate merging or redistribution techniques are applied to optimize connectivity, distribute communication and computational load more evenly, minimize redundant transmissions, and maintain high operational efficiency.

The side length (S) of each hexagonal DFG is calculated using the formula presented in **Eq. 18**:

$$S = W/f, \quad (18)$$

Where W denotes the side length of the square network area (in meters), and f is a tunable constant determined by simulation parameters.

In the pointy-top hexagonal layout, the vertical spacing (v) between the centers of adjacent hexagons is $\times \left(\frac{3}{4}\right) \times height$ given $\times \left(\frac{3}{2}\right) \times S$ by $\sqrt{3} \times S$, while the horizontal spacing (h) between centers of hexagons $\times \frac{width}{\sqrt{3}}$ in the same row equals $1.5 \times S$, as illustrated in **Figure 4**.

Figure 5a shows the initial random distribution of SNs across an example $200 \times 200 \text{ m}^2$ sensing area before partitioning. **Figures 5b and 5c** then depict the network after division into hexagonal DFGs and the resulting assignment of SNs to their respective groups.

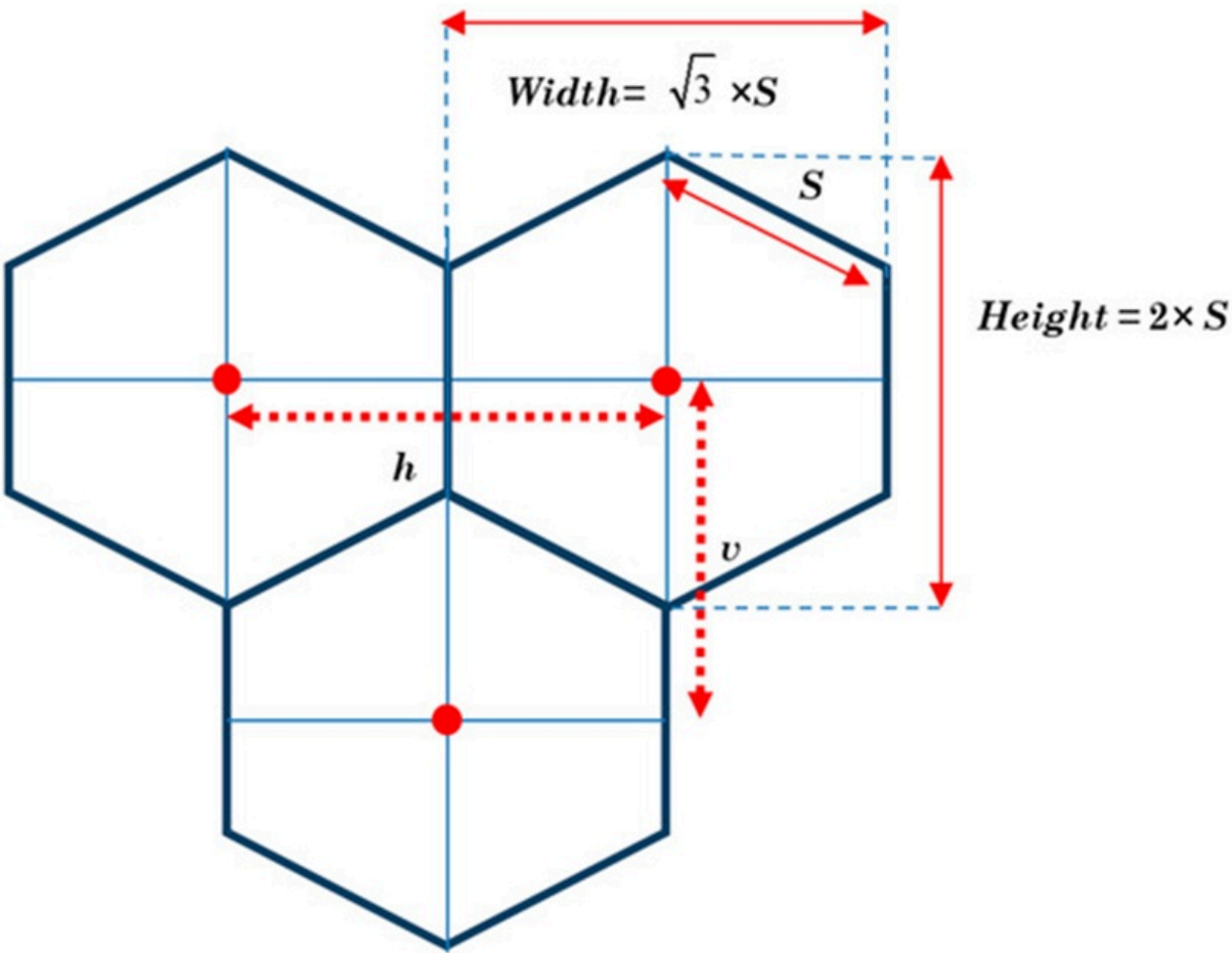


Figure 4. Partitioning the network area into equally sized hexagonal DFGs.

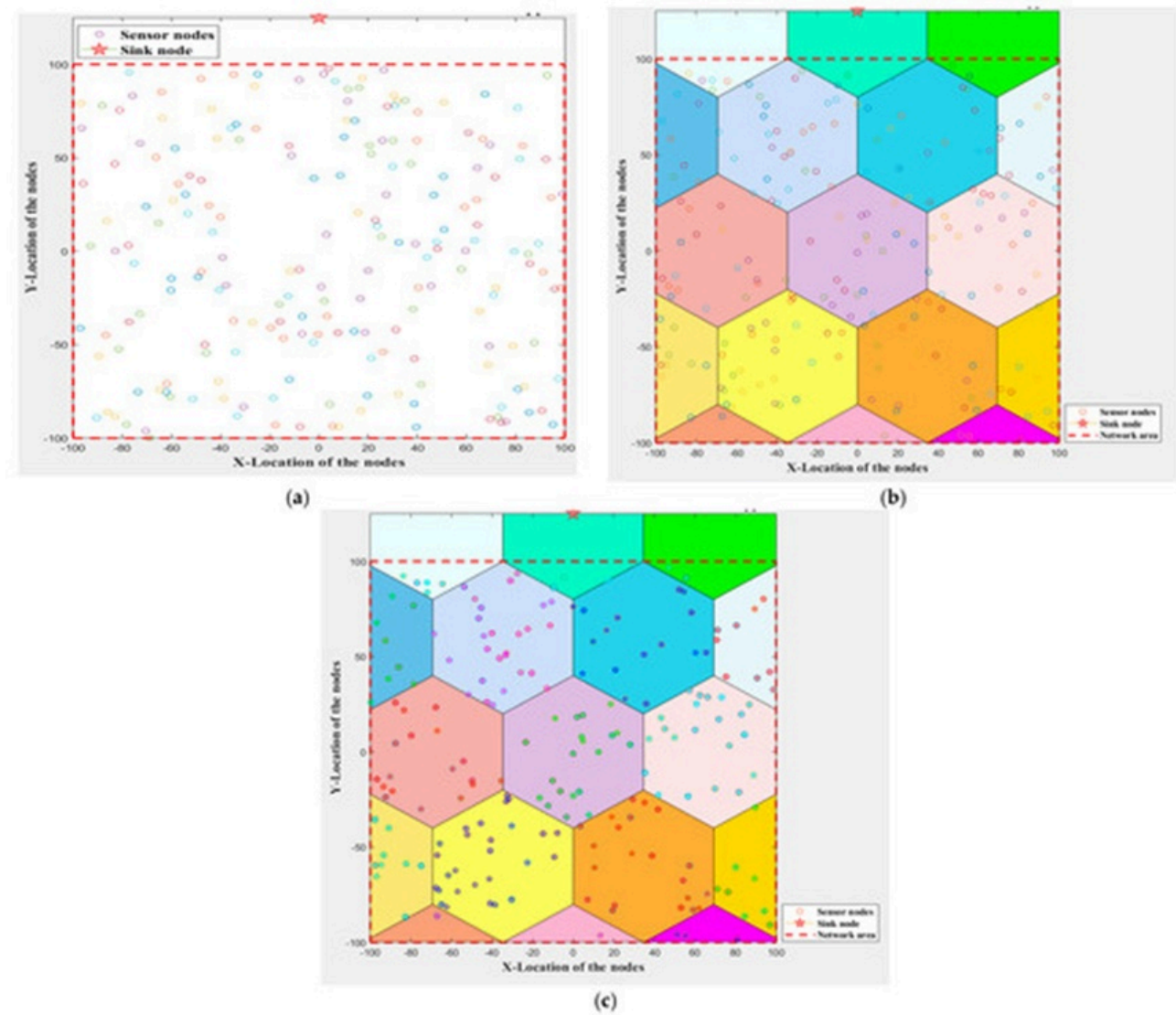


Figure 5. Partitioning the network area into equally sized hexagonal DFGs: (a) A $300 \times 300 \text{ m}^2$ network area before partitioning, (b) Partitioning the network area into equal-sized hexagonal DFGs, and (c) Distributing SNs into the formed DFGs based on their locations.

In the proposed protocol, each network dimension is tailored to represent a specific IoT application scenario, with a uniform assumed network density of 0.0025 nodes/m^2 , which also corresponds to the target density per Data Fusion Group (DFG). **Table 3** summarizes the characteristics of these representative applications. Furthermore, **Figure 6a–6e** illustrates example deployments for a smart logistics park, smart medical center, smart factory zone, smart university campus, and smart city, respectively, each matched to its corresponding network area.

Input:

In: N (Number of SNs in the network)

Output:

Out: C (Optimal number of DFG)

Out: N_c (Number of DFGMs in each DFG)

Out: RMs (Routing metrics for each SN)

Out: $\overrightarrow{Prey}$ matrix (the initialization matrix for the MPA algorithm)

Main:

For each $i = 1$ to N

 Distribute SN_i randomly

 Set SN_i ID, X-location, Y-location, and E_o

 Sink node sends request information message $\rightarrow SN_i$

SN_i replies with SN's information message (SN ID, X-location, Y-location, and E_o) $\rightarrow$ Sink node

End For

Set the network area size (W)

Sink node calculates the length size of hexagonal (S) using Equation (18)

Sink node splits the network area into identical hexagonal DFGs

$C \leftarrow$ Number of hexagonal DFGs

For each $i = 1$ to N

 Calculate the Euclidian distance between SN_i and the sink node using Equation (23) // RM₁

 Calculate the ratio between the E_o and RE using Equation (21) // RM₂

 Calculate ASBO using Equation (27) // RM₅

 Distribute SN_i into the formed DFGs based on their locations

End For

$N_c \leftarrow$ number of DFGMs in each DFG

For each $i = 1$ to N_c

 Calculate Average distance between DFGMs using Equation (24) // RM₃

 Calculate X using Equation (26) // RM₄

End For

For each $i = 1$ to N_c

$Prey(i) \leftarrow$ Save the values of RM₁, RM₂, RM₃, RM₄, and RM₅ using Equation (20)

End For

Return C , N_c , and $\overrightarrow{Prey}$

End Main

Algorithm 1. DFGs formation (setup phase) for the proposed protocol.



Figure 6. A variety of IoT applications spanning different network areas: (a) Model for a smart logistics park for 100×100 m² of network area, (b) Model for a smart medical center for 200×200 m² of network area, (c) Model for a smart factory zone for 300×300 m² of network area, (d) Model for a smart university for 400×400 m² of network area, and (e) Model for a smart city for 500×500 m² of network area.

Table 3. Variations in network area sizes and their corresponding details.

Network deployment area (W × W)	Hexagonal cell side length (S)	Number of sensor nodes (N)	IoT application domain	Sink node coordinates (x, y)
100 × 100 m ²	20 m	25	Smart logistics and warehousing park	(0, 75)
200 × 200 m ²	40 m	100	Smart healthcare / medical facility	(0, 125)

300 × 300 m ²	60 m	225	Smart industrial / manufacturing zone	(0, 175)
400 × 400 m ²	80 m	400	Smart campus/university environment	(0, 225)
500 × 500 m ²	100 m	625	Comprehensive smart city deployment	(0, 275)

PDFH and SDFH selection utilizing the MPA

Since Data Fusion Heads (DFHs) serve as critical intermediaries for both data aggregation and routing in Wireless Sensor Network (WSN)-based IoT systems, selecting appropriate DFHs is a pivotal step. This is particularly true in homogeneous networks where every Sensor Node (SN) starts with identical energy levels. A well-designed DFH election strategy is therefore essential to more evenly distribute energy consumption across the entire network [13].

Over the years, various optimization approaches have been proposed for Cluster Head (CH) or DFH selection, including Genetic Algorithms (GA) [52, 53], Particle Swarm Optimization (PSO) [35, 54], fuzzy logic systems [55, 56], and the Equilibrium Optimizer (EO) [31, 57], among others. The present protocol distinguishes itself by adopting the Marine Predators Algorithm (MPA)—a highly effective, state-of-the-art metaheuristic—to identify the optimal Primary Data Fusion Head (PDFH), First Secondary DFH (FSDFH), and Second Secondary DFH (SSDFH) for each Data Fusion Group (DFG) through the minimization of a carefully formulated Data Fusion Fitness Function (DFFF).

MPA stands out as an innovative nature-inspired metaheuristic that has demonstrated strong adaptability across a wide range of real-world optimization challenges. Its core strength lies in faithfully replicating the foraging strategies and adaptive behaviors of marine predators. By drawing on these biologically plausible mechanisms, MPA efficiently explores complex, high-dimensional search spaces, making it particularly well-suited to solving the PDFH selection problem [58]. The algorithm effectively balances global exploration and local exploitation phases, offers a straightforward implementation accessible to researchers and engineers alike, and includes a memory-retention feature that preserves and reuses promising solutions across iterations, thereby enabling progressive refinement [44].

Within the proposed protocol, MPA has proven highly effective at selecting superior PDFHs, resulting in substantial energy savings and a marked extension of overall network lifetime. To achieve this, the sink node evaluates five key reference metrics (RMs) when computing the DFFF: the communication distance between each DFGM and the sink node, the Remaining Energy (RE) of each DFGM, the average distance among DFGMs within the group, the Average Scale of Building Occlusions (ASBO), and the frequency of PDFH rotation. Figure 8 provides a comprehensive visual depiction of the MPA-driven PDFH selection process.

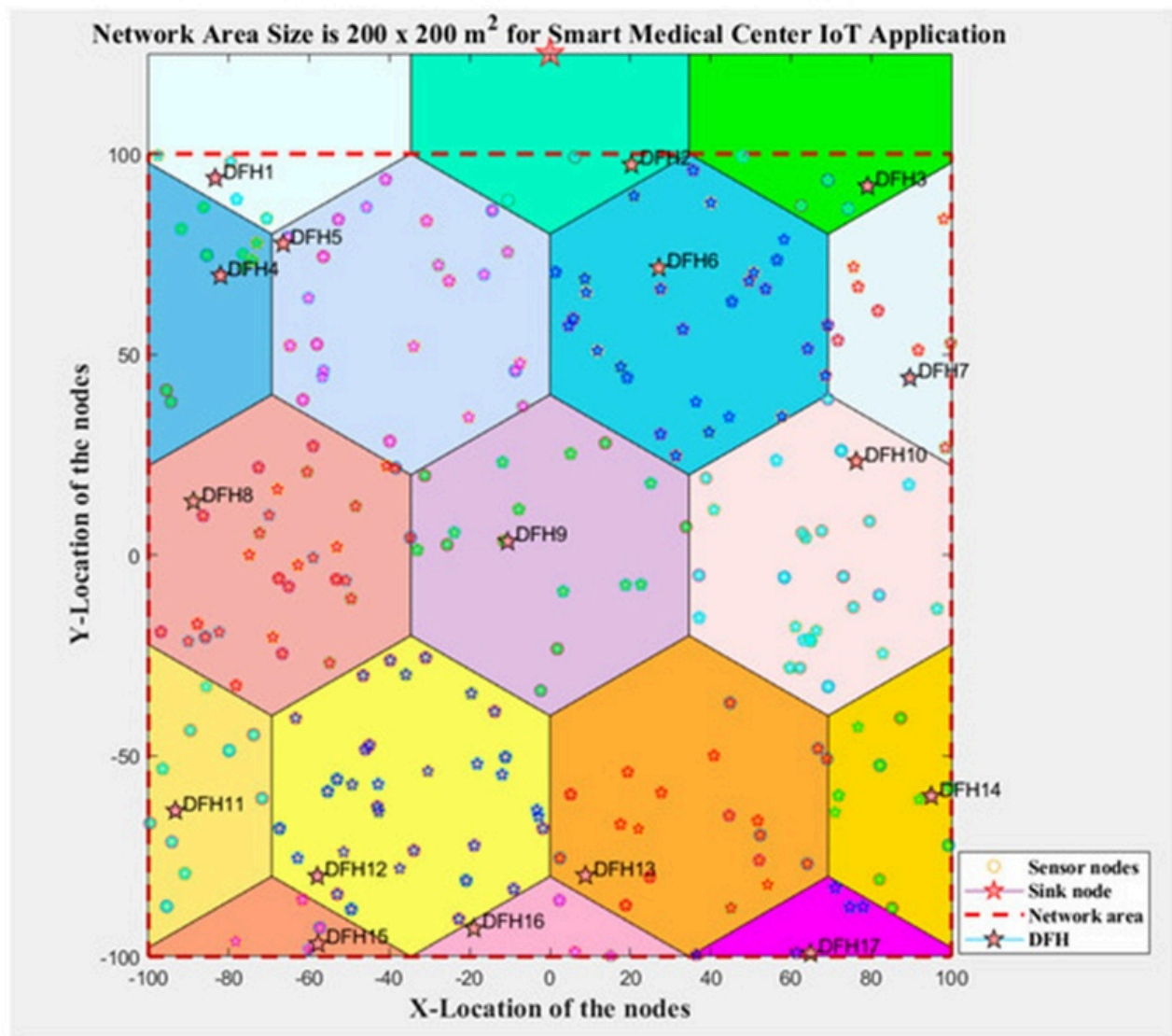


Figure 7. The selected PDFHs are utilizing the MPA for the proposed protocol.

The entire data collection process begins with the sink node broadcasting a Request Information Message to all Sensor Nodes in the network. Each SN responds by unicasting an information message containing its unique ID, X and Y coordinates, and current Remaining Energy (RE) value. Having gathered complete location and energy status for every node, the sink node proceeds with the setup phase: it partitions the sensing field into hexagonal DFGs, assigns SNs to their respective groups, and executes the MPA to determine the best candidate DFGM for the PDFH role in each DFG.

In summary, this work employs the well-established MPA as the core optimization engine to identify the most suitable DFGM for PDFH. As detailed in the following subsections, the MPA operates in two main stages: an initialization phase followed by an iterative refinement phase.

The initialization phase

In the initialization stage of the Marine Predators Algorithm (MPA), several key input parameters must first be defined and configured. These include the following:

- (a) Number of search agents (SAs): This value corresponds to the total number of Data Fusion Group Members (DFGMs) within a given DFG that are eligible to compete for the Data Fusion Head (DFH) role (denoted as N_c).
- (b) Dimension (Dim): This parameter specifies the number of Reference Metrics (RMs) considered in the present study; it is fixed at 5.
- (c) Maximum number of iterations (Max_Iter): This controls how many times the algorithm iterates to identify the most suitable DFH candidate. In the proposed protocol, Max_Iter is set to 500.
- (d) Lower and upper bounds (LB and UB): The lower bound (LB) is set to 0 and the upper bound (UB) to 1. The proposed protocol normalizes all metrics to the $[0, 1]$ interval. This choice is motivated by the absence of precise physical or absolute bounds for three of the RMs used in the study. Normalization ensures fair weighting and consistent scaling across the criteria, placing every metric on an equal footing within the standardized range $[0, 1]$.
- (e) Positions: This parameter consists of an array containing the unique identifiers (IDs) of the DFGMs inside a particular DFG. It is used to identify the DFH candidate ultimately selected as optimal.
- (f) Main parameters of MPA: As originally specified by Faramarzi et al. [43], these include $FADs = 0.2$ and $P = 0.5$. These default values are retained unchanged in the current implementation. The parameters listed above serve as the starting configuration and are progressively updated during the subsequent iterative phases.

$$\begin{aligned}
 Top_predator_pos &= [0 \ 0 \ 0 \ 0 \ 0], Top_predator_fit = \infty \\
 \overrightarrow{StepSize} &= \begin{matrix} SA_1 \\ SA_2 \\ \vdots \\ SA_{N_c} \end{matrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, fitness = \begin{matrix} SA_1 \\ SA_2 \\ \vdots \\ SA_{N_c} \end{matrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\
 \overrightarrow{X}_{min} &= \begin{matrix} SA_1 \\ SA_2 \\ \vdots \\ SA_{N_c} \end{matrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \overrightarrow{X}_{max} = \begin{matrix} SA_1 \\ SA_2 \\ \vdots \\ SA_{N_c} \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}
 \end{aligned} \tag{19}$$

- (g) $\overrightarrow{Prey}$ matrix generation: The $\overrightarrow{Prey}$ matrix constitutes the initial population of candidate solutions in MPA. Each DFG is assumed to contain a collection of DFGMs, every one of which is regarded as a potential Candidate Data Fusion Head (CDFH). These candidates are collectively represented as $CDFH = [CDFH_1, CDFH_2, \dots, CDFH_{N_c}]$, where N_c equals the number of rows in the $\overrightarrow{Prey}$ matrix. Each DFGM is evaluated according to five Reference Metrics (RMs), which form the columns of the matrix: $RM = [RM_1, RM_2, RM_3, RM_4, RM_5]$. Specifically:

- RM_1 represents the distance from a given $CDFH_i$ to the sink node ($DCDFH_i \rightarrow Sink$),
- RM_2 denotes the normalized inverse of the remaining energy ratio for $CDFH_i$ ($I(RECDFH_i)$),
- RM_3 is the average distance from $CDFH_i$ to all other DFGMs within the group ($AvgDisCDFH_i \rightarrow DFGMs$),

- RM_4 corresponds to the ratio of the average number of times all SNs in the DFG have served as DFH (across rounds or batches) to the number of times $CDFH_i$ has performed this role ($XCDFH_i$),
- RM_5 stands for the Average Scale of Building Occlusions affecting $CDFH_i$ ($ASBOCDFH_i$).

The precise mathematical definitions and formulations of these RMs are provided later, as detailed in [Eq. 20](#).

$$\overrightarrow{Prey} = \begin{bmatrix} CDFH_1 \\ CDFH_2 \\ \vdots \\ CDFH_{N_c} \end{bmatrix} \begin{bmatrix} RM_1 & RM_2 & RM_3 & RM_4 & RM_5 \\ D_{CDFH_1 \rightarrow Sink(1,1)} & Inv(RE_{CDFH_1}(1,2)) & AvgDis_{CDFH_1 \rightarrow DFGMs(1,3)} & X_{CDFH_1(1,4)} & ASBO_{CDFH_1(1,5)} \\ D_{CDFH_2 \rightarrow Sink(2,1)} & Inv(RE_{CDFH_2}(2,2)) & AvgDis_{CDFH_2 \rightarrow DFGMs(2,3)} & X_{CDFH_2(2,4)} & ASBO_{CDFH_2(2,5)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D_{CDFH_{N_c} \rightarrow Sink(N_c,1)} & Inv(RE_{CDFH_{N_c}}(N_c,2)) & AvgDis_{CDFH_{N_c} \rightarrow DFGMs(N_c,3)} & X_{CDFH_{N_c}(N_c,4)} & ASBO_{CDFH_{N_c}(N_c,5)} \end{bmatrix}_{N_c \times 5} \quad (20)$$

It is important to highlight that the $\overrightarrow{Prey}$ matrix is fully normalized. In other words, a separate normalization procedure is applied to every RM value. This rescaling step maps each metric to the uniform interval [0, 1], which is crucial for ensuring an unbiased and equitable comparison among diverse criteria during the selection of the best DFH.

The iterative phase

In this phase, the Data Fusion Fitness Function (DFFF) is employed to identify the most suitable Data Fusion Head (DFH). We recognize that the DFFF plays a central role in guiding the Marine Predators Algorithm (MPA) toward locating and capturing the prey (i.e., the target node that best satisfies the optimization criteria). This subsection first defines the individual parameters of the DFFF and then presents its complete, robust formulation.

- DFFF parameters definitions:

The effectiveness of MPA in selecting the optimal DFH hinges critically on the precise formulation of the DFFF. To this end, the proposed DFFF incorporates the following objectives: minimizing the distance from the candidate DFH to the sink node ($DCDFH_i \rightarrow Sink$), minimizing the average distance to all DFGMs ($AvgDis_{CDFH_i \rightarrow DFGMs}$), maximizing the Remaining Energy (RE) level, minimizing the Average Scale of Building Occlusions (ASBO), and minimizing the frequency of PDFH rotations. These factors collectively aim to extend network lifetime while substantially lowering overall energy consumption. Below is a detailed explanation of each parameter:

(1) The ratio of $CDFH_i$'s initial energy (E_0) to its current RE (RE_{CDFH_i}) — denoted as $Inv(RE_{CDFH_i})$: This metric quantifies the residual energy available in each Candidate Data Fusion Head (CDFH) after deployment and operation. It is the single most influential factor in DFH selection. Because the DFH is responsible for collecting, aggregating, and forwarding data from all DFGMs to the sink node, it consumes significantly more energy than ordinary Sensor Nodes (SNs). This parameter, labeled $DFFFP_1$, is to be minimized. Following normalization, its value is confined to the interval [0, 1], as expressed in [Eq. 21](#). Including this term ensures that candidates with higher remaining energy are preferentially selected as DFHs.

$$DFFFP_1 = Inv(RE_{CDFH_i}) = E_{0_{CDFH_i}} / RE_{CDFH_i}, \quad 0 < DFFFP_1 < 1. \quad (21)$$

(2) Distance of $CDFH_i$ to the sink node ($DCDFH_i \rightarrow Sink$): Represented as $DFFFP_2$, this parameter must be minimized to reduce the total transmission distance and associated energy cost. Normalization maps its value into the range [0, 1].

$$DFFFP_2 = D_{CDFH_i \rightarrow Sink}, \quad 0 < DFFFP_2 < 1. \quad (22)$$

The Euclidean distance formula is used to compute $D_{CDFH_i \rightarrow Sink}$, as shown in Eq. 23. This metric favors candidates located closer to the sink, thereby promoting energy-efficient inter-group communication.

$$D_{CCH_i \rightarrow Sink} = \sqrt{(sink_y - CCH_{iy})^2 + (sink_x - CCH_{ix})^2}. \quad (23)$$

(3) Average distance between DFGMs ($AvDis_{CDFH_i \rightarrow DFGMs}$): This parameter measures how centrally positioned a candidate $CDFH_j$ is relative to the other members of its Data Fusion Group (DFGMs). Lower values indicate greater geographic centrality, thereby reducing intra-DFG communication overhead and energy expenditure. $DFFFP_3$ represents this metric and is to be minimized. Euclidean distance is applied to calculate the separation between $CDFH_j$ and each neighboring DFGM, as detailed in Eq. 24.

$$DFFFP_3 = AvDis_{CDFH_i \rightarrow DFGMs} = \frac{1}{N_c - 1} \sum_{j=1, j \neq i}^{N_c} (D_{CDFH_j \rightarrow CDFH_i}), \quad 0 < DFFFP_3 < 1. \quad (24)$$

$$D_{CDFH_j \rightarrow CDFH_i} = \sqrt{(CDFH_{jy} - CDFH_{iy})^2 + (CDFH_{jx} - CDFH_{ix})^2}. \quad (25)$$

(4) $XCDFH_i$: This parameter quantifies the rotation fairness of the DFH role. It is defined as the ratio of the number of times (in rounds or batches) the current $CDFH_j$ has served as DFH to the average number of times all DFGMs in the group have served as DFH. Expressed as $DFFFP_4$, it is formulated as follows:

$$DFFFP_4 = X_{CDFH_i} = DFHR_{CDFH_i} / \left(\frac{\sum_{k=1}^{N_c} DFHR_{CDFH_k}}{N_c} \right), \quad 0 < DFFFP_4 < 1. \quad (26)$$

Where $DFRCDFH_k$ denotes the number of times (rounds or batches) that candidate $CDFH_k$ has been assigned the DFH role.

(5) Average Scale of Building Occlusions ($ASBO_{CDFH_i}$): ASBO estimates the degree to which surrounding buildings obstruct line-of-sight, signal propagation, or other environmental factors (e.g., sunlight exposure or wireless connectivity) at the candidate's location. This parameter compares the height/position of $CDFH_j$ against the heights of nearby buildings within its coverage area, as illustrated in Table 4 and Figure 8. Represented as $DFFFP_5$, ASBO is defined by:

$$DFFFP_5 = ASBO_{CDFH_i} = \frac{\sum_{i=1}^{N_b} SBO_{CDFH_i}}{N_b}, \quad 0 < DFFFP_5 < 1. \quad (27)$$

Where $SBOCDFH_i$ is the individual building occlusion scale value assigned to $CDFH_i$, and N_b is the total number of buildings that potentially obstruct the candidate within its effective coverage radius. Further interpretation of the occlusion scale is provided in [Tables 5, 6, and 7](#), which categorize the impact at low, moderate, and high heights, respectively. Lower scale values are always preferred.

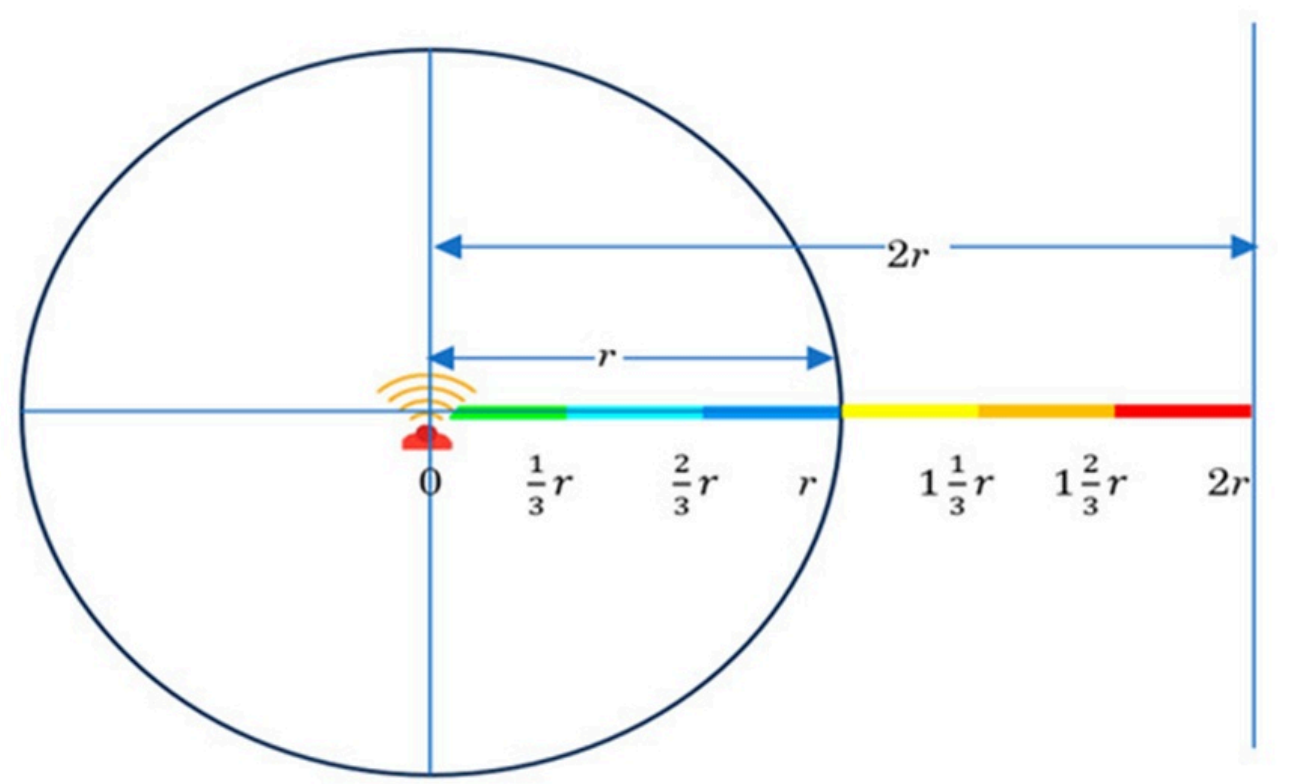


Figure 8. Representation of building occlusions based on a PDFH position and the coverage area.

Table 4. Representation of building occlusions based on a PDFH position.

	Range	Interpretation
	$0-1/3r$	Extremely close
	$1/3r-2/3r$	Close
	$2/3r-r$	Moderately close
	$r-1\ 1/3r$	Moderately far
	$1\ 1/3r-1\ 2/3r$	Far
	$1\ 2/3r-2r$	Extremely far

Table 5. The scale for building occlusions at low height.

Scale	Interpretation

0.1	Extremely close
0.2	Close
0.3	Moderately close
0.4	Moderately far
0.5	Far
0.6	Extremely far

Table 6. The scale for building occlusions at moderate height.

Scale	Interpretation
0.2	Extremely close
0.3	Close
0.4	Moderately close
0.5	Moderately far
0.6	Far
0.7	Extremely far

Table 7. The scale for building occlusions at high height.

Scale	Interpretation
0.3	Extremely close
0.4	Close
0.5	Moderately close
0.6	Moderately far
0.7	Far
0.8	Extremely far

• DFFF estimation:

The DFFF serves as the core objective function that MPA minimizes to identify the best Primary Data Fusion Head (PDFH) and First Secondary DFH (FSDFH). By integrating all five parameters described above, the DFFF enables the algorithm to explore and converge on high-quality solutions. The candidate CDFH_i that yields the lowest DFFF value is designated as the PDFH, and the second-lowest is designated as the FSDFH. The final DFFF is therefore expressed as the following weighted linear combination of DFFFP₁ through DFFFP₅:

$$DFFF = w_1 \times DFFFP_1 + w_2 \times DFFFP_2 + w_3 \times DFFFP_3 + w_4 \times DFFFP_4 + w_5 \times DFFFP_5, \quad (28)$$

Where w_1 , w_2 , w_3 , w_4 , and w_5 represent the weight coefficients assigned to RM_1 , RM_2 , RM_3 , RM_4 , and RM_5 , respectively, with the constraint that $w_1 + w_2 + w_3 + w_4 + w_5 = 1$. The specific weight values used are $w_1 = 0.3375$, $w_2 = 0.2475$, $w_3 = 0.18$, $w_4 = 0.135$, and $w_5 = 0.1$. These weights were determined through extensive trial-and-error experimentation to maximize overall network performance.

Algorithm 2 presents the full procedure for selecting PDFH, FSDFH, and SSDFH using the MPA within the proposed protocol.

Input:
In: N_c (Number of CDFHs)
In: $DFFFP_1, DFFFP_2, DFFFP_3, DFFFP_4, DFFFP_5$ (RMs (Routing Metrics for each CDFH))

Output:
Out: Optimal PDFH, FSDFH, SSDFH IDs

Main:
Set the parameters $P = 0.5$; $FADs = 0.2$; $Dim = 3$; $UB = 1$; $LB = 0$; $Iter = 0$; $MAX_Iter = 500$;
 $w_1 = 0.3375$; $w_2 = 0.2475$; $w_3 = 0.18$; $w_4 = 0.135$; $w_5 = 0.1$.
Set Top_Predator_Position to 0
Set Top_Predator_Fitness to ∞
Set StepSize matrix to 0
Set DFFF to ∞
Construct $\overline{Prey}$ matrix based on the RMs for each $CDFH_i$ using Equation (20)

Construct $\overline{X}_{min}$ and $\overline{X}_{max}$ matrices

While $Iter < Max_Iter$ **do**
For $i = 1$ to N_c **do**
Evaluate DFFF of $\overline{Prey}$ matrix using eq (28)
If $DFFF(CDFH_i) < Top_Predator_Fitness$ **then**
Top_Predator_Fitness $\leftarrow DFFF(CDFH_i)$
Top_Predator_Position $\leftarrow Prey(i, :)$
PDFH $\leftarrow i$ // return the ID of $CDFH_i$
End If
End For
Perform memory saving
Construct $\overline{Elite}$ matrix based on eq (2)
Calculate CP based on eq (10)
Generate $\overline{RL}$, $\overline{RB}$, and $\overline{R}$
If $Iter < Max_Iter/3$ **then** // Phase 1
Update $\overline{Prey}$ matrix using eqs (4-5)
Elseif $Max_Iter/3 < Iter < 2(Max_Iter)/3$ **then** // Phase 2
// The first half of the population
For $k = 1$ to $(N_c/2)$ **do**
Update $\overline{Prey}$ matrix using eqs (6-7)
End For
// The second half of the population
For $k = N_c/2 + 1$ to N_c **do**
Update $\overline{Prey}$ matrix using eqs (8-9)
End For
Elseif $Iter > 2(Max_Iter)/3$ **then** // Phase 3
Update $\overline{Prey}$ matrix using eqs (10-11)
End if
For each $i = 1$ to N_c **do**
Evaluate DFFF of $\overline{Prey}$ matrix using eq (28)
If $DFFF(CDFH_i) < Top_Predator_Fitness$ **then**
Top_Predator_Fitness $\leftarrow DFFF(CDFH_i)$
Top_Predator_Position $\leftarrow Prey(i, :)$
PDFH $\leftarrow i$ // return the ID of $CDFH_i$
End If
End For
Perform memory saving
Apply Eddy formation and the effect of FADs' effect using eq (12)
Iter++
End While
Return PDFH, FSDFH, and SSDFH, which are the Top_Predator_Fitness, the second-least DFFF, and the third-least DFFF, respectively.
End Main

Algorithm 2. PDFH, FSDFH, and SSDFH selection using the MPA in the proposed protocol.

Optimizing data fusion regrouping (PDFH Reselection) procedure

To ensure that the PDFH selection procedure does not occur in every round, this paper proposes that, after the sink node initially selects the PDFH for each DFG using the MPA during the first round, the regrouping decision for subsequent rounds is handled locally by the current PDFH (i.e., in a distributed manner).

The proposed protocol requires each PDFH to compare its remaining energy level (REPDFH) against the energy threshold (E_{th}) in every following round after the setup phase. Eq. 29, which defines E_{th} , expresses it as the average remaining energy across all DFGMs, including the PDFH itself.

If REPDFH is greater than or equal to E_{th} , the PDFH may continue to perform communication within the DFG with its DFGMs. Otherwise, it must cease its role and transmit a disjoint message to the sink node and all sensor nodes (SNs) throughout the network, prompting the selection of a new PDFH according to Eq. 30. This disjoint message includes the current remaining energy levels of all DFGMs.

$$E_{th} = AVG_RE_{(DFGMs)} = \left(\sum_{j=1}^{N_c} RE(j) / N_c \right). \quad (29)$$

$$PDFH \text{ reselection} = \begin{cases} \text{Select a new PDFH,} & \text{if } E_{th} > RE_{PDFH} \\ \text{Keep the current PDFH,} & \text{if } E_{th} < RE_{PDFH} \end{cases}. \quad (30)$$

This approach significantly reduces control message overhead in the setup phase (namely, during PDFH selection) by ensuring the protocol avoids repeating the PDFH election in every round.

Only when the remaining energy (RE) of the currently assigned PDFHs falls below a specific threshold does reselection become necessary. The process is considered complete once another qualified DFGM is elected to assume the PDFH duties. As a consequence, the same PDFH is permitted to serve across multiple successive rounds until its RE eventually falls beneath E_{th} . This explains the protocol's classification as batch-based. Imposing this restriction is vital to stop the PDFH from exhausting its energy too soon and causing network partitions or early node failures.

Moreover, the protocol incorporates a valuable enhancement: a dynamically adjusted RE threshold. This ensures that sensor nodes (SNs) remain eligible for the PDFH position until the network completely fails. Put differently, E_{th} gets recalculated at the start of each round and is dynamically updated in accordance with the latest remaining energy values among the DFGMs, exactly as formulated in Eq. 29. The experimental results clearly show that this mechanism markedly prolongs the overall network lifetime compared with standard techniques.

Whenever a PDFH's REPDFH drops under E_{th} , that node is downgraded to regular DFGM status, allowing it to keep performing its normal sensing responsibilities. Selection of a successor PDFH is then performed using the residual metrics (RMs) and the MPA algorithm. At the same time, the network discards any SNs that have already died. Given that PDFHs must remain active without interruption, their energy expenditure is considerably higher. For this reason, the protocol periodically rotates the PDFH responsibility among different SNs to achieve fair energy distribution. Algorithm 3 presents the complete procedures for both initial PDFH selection and subsequent reselection in the proposed scheme.

Input:
In₁: N_c (Number of DFGMs in each DFG)
In₂: C (Optimal number of DFGs)
In₃: RM_s (Routing metrics for each SN)
In₄: *Prey matrix* (the initialization matrix for the MPA algorithm)
In₅: N (Number of SNs in the network)
In₆: *Round* (Number of rounds)

Output:
Out₁: PDFH (Optimal PDFH for each DFG)

Main:
// PDFH selection at the setup phase of the first round

If (*Round* = 1) **then**
 For $i = 1$ to C **do**
 For $j = 1$ to N_c **do**
 Sink node runs the MPA algorithm
 PDFH ← Best solution from the MPA algorithm
 End For
 End For
 Sink node identifies the type of SN (either PDFH or DFGM)
 Sink node sends sink' table message → SNs
 For $i = 1$ to N
 SN_i receives sink' table message ← Sink node
 End For
End If

If (*Round* > 1) **then**
 If ($E_{th} > RE_{PDFH}$) **then**
 PDFH sends disjoin message contains (DFGM ID, Current RE) → sink node
 PDFH sends disjoin message → all SNs
 Sink node updates the RE of all SNs based on the received disjoin message information
 // PDFH reselection begins
 For $i = 1$ to C **do**
 For $j = 1$ to N_c **do**
 Sink node runs the MPA algorithm
 PDFH ← Best solution from the MPA algorithm
 End For
 End For
 Sink node identifies the type of SN (either PDFH or DFGM)
 Sink node sends sink' table message → SNs
 For $i = 1$ to N **do**
 SN_i receives sink' table message ← Sink node
 End For
 Elseif
 $Round++$
 End If
End If

For each Round do // at the end of each round
 Update the values of the RE of all SNs
 Calculate the E_{th} based on eq (29)
End For
Return PDFH for the next round
End Main

Algorithm 3. PDFH selection and reselection procedures for the proposed protocol.

Mastering data fusion communications: efficient scheduling and transmission over intra- and inter-DFG

As explained earlier, sensor nodes (SNs) transmit their information messages in response to a query broadcast by the sink node. The sink carefully logs the exact arrival time of every incoming message for later use.

Following this, the sink divides the entire deployment region into congruent hexagonal DFGs and elects the PDFH, FSDFH, and SSDFH for each DFG by applying the MPA across the DFGMs within it. Once these roles have been assigned, the sink disseminates a comprehensive table message throughout the network. The message carries, for each node, its identifier, arrival timestamp, designated role (PDFH or DFGM), corresponding DFG index, PDFH identifier, assigned time slots (TSs), and distances to both the sink and the relevant PDFH. As soon as all SNs have received and processed this table message, the network transitions into the steady-state phase.

As stated, each SN delivers its collected data to its own PDFH; the PDFH, in turn, forwards the information to the sink. The sink therefore maintains detailed awareness of every PDFH identifier, its associated DFGMs, DFG identifiers, round durations (RT), and the arrival times of messages from all SNs. Consequently, it becomes the sink's responsibility to distribute TDMA time slots (TSs) to the sensor nodes. A thorough explanation of this scheduling approach appears shortly below.

In reality, the TDMA framework originally described in Wang and Henning [59] has been substantially refined here. The improved TDMA mechanism exploits local neighborhood knowledge to organize time slots more effectively and reduce the likelihood of packet collisions. Within its local zone (the DFG), the elected PDFH serves as the central authority overseeing both data aggregation and forwarding operations. Additionally, the sink constructs a dedicated TDMA schedule tailored to each DFG's DFGMs, allocating each node a precise, exclusive time slot for sending its data during steady-state operation.

Thanks to this scheduling discipline, interference among nodes belonging to the same DFG is completely avoided, and intra-DFG data collisions are prevented because transmissions occur strictly within each node's designated TS. This design also greatly simplifies the adoption of a sleep–wake duty cycle by the DFGMs. During the sleep portion of the cycle, nodes remain inactive except when their assigned TDMA TS arrives, thereby conserving energy. In sleep mode, SNs can further reduce power consumption by operating at a lower activity level.

Upon completion of each round (during the steady-state phase), the PDFH collects the sensed readings submitted by all DFGMs according to their scheduled TSs, fuses the received information, and transmits the consolidated result to the sink node—either directly (single-hop) or via intermediate relays (multi-hop). Irrespective of how many individual packets arrive, the PDFH always compresses them into a single unified packet, substantially reducing the transmitted data volume before sending the fused outcome directly to the sink.

Crucially, communication between the sink and each PDFH relies on Direct Sequence Spread Spectrum (DSSS) technology. DSSS is deliberately chosen to suppress interference between different DFGs. By assigning every DFG its own distinct spreading code—unique with respect to all neighboring DFGs—inter-DFG interference is effectively eliminated, allowing coexisting DFGs to operate without mutual disruption.

To better illustrate how TDMA scheduling works in this routing-free protocol, once the data fusion grouping has been finalized, the sink assigns each DFGM within a given DFG a priority rank based on the arrival timestamp of its information message. For nodes categorized as Normal DFGM ("NDFGM"), ranking adheres to a simple "first to arrive, first to be assigned" rule: the earliest arriving message at the PDFH obtains rank 1, the next obtains rank 2, and the process continues sequentially until every NDFGM in that DFG receives its rank.

By contrast, the PDFH is automatically granted the top rank within its DFG—regardless of when its own message arrived—since it is responsible for collecting data from all DFGMs, merging it with its local readings, and delivering the aggregated packet to the sink. Leveraging these rank values, the sink then applies its TDMA scheduling logic to determine each node’s transmission time slot. The rank is recorded in the “Turn of TDMA TS” field and serves as the basis for allocation. The exact computation of the assigned time slot is provided in Eq. 31, while an example of the intra-DFG and inter-DFG time-slot arrangement for a specific group (e.g., DFG 6) is depicted in Table 8.

$$TS = \begin{cases} \frac{RT}{N_c + 1}, & \text{if single-hop communication is employed,} \\ \frac{RT}{N_c + N_{PDFRs} + 1}, & \text{if multi-hop communication is employed,} \end{cases} \quad (31)$$

Where N_{PDFRs} denotes the number of PDFRs required to forward data all the way to the sink node.

Table 8. TDMA schedule for intra- and inter-DFG communication in DFG 6.

SN ID	Message Arrival Time (ms)	Node Type	DFG ID	PDFH ID	Assigned Rank	Allocated TDMA Time Slot (TS Number)
20	10	NDFGM	6	10	1	1
16	15	NDFGM	6	10	2	2
4	16	NDFGM	6	10	3	3
32	20	NDFGM	6	10	4	4
10	7	PDFH	6	10	5	5

Transmission through multiple hops using an innovative relay data fusion cost function for improving routing technique

In Internet of Things (IoT) networks, data fusion is a key strategy for overcoming energy constraints, especially since sensor nodes (SNs) operate under strict energy budgets. By aggregating and performing preliminary processing of raw data at local nodes, data fusion substantially reduces the volume of information that must travel across the network. This reduction in transmitted data directly lowers the energy spent on communication—a factor that depends heavily on both packet length and the distance over which signals propagate. The process removes duplicate or unnecessary information, transmits only meaningful content, and supports localized computation, thereby decreasing dependence on long-range transmissions that demand far greater power. These advantages collectively prolong the operational life of individual SNs and the entire network, rendering IoT systems more durable, energy-efficient, and practical for large-scale or remote deployments where battery replacement is difficult or impossible. Thus, data fusion emerges as an indispensable mechanism for energy optimization and overall performance improvement in IoT environments.

In the routing-free protocol proposed here, each PDFH aggregates data from its member SNs and transmits the fused result directly to the sink node. Consequently, PDFHs farther from the sink deplete their energy reserves more rapidly and fail earlier than those closer to it. To promote balanced energy usage, ensure reliable performance, and minimize communication costs (particularly transmission distances), the protocol introduces multi-hop forwarding. Notably, this multi-hop mechanism integrates seamlessly with the DFG structure: each PDFH independently collects sensed data from

all SNs within its own DFG and arranges for delivery to the sink node. To accomplish efficient long-range forwarding, PDFHs rely on selected intermediate nodes—acting as PDFRs, FSDFRs, and SSDFRs—to relay the aggregated data.

The multi-hop strategy adopted in this protocol delivers dependable, stable, and well-balanced load distribution by effectively shortening individual transmission ranges. Building on the concepts outlined earlier, the protocol distinguishes between two primary communication modes: intra-DFG and inter-DFG. When the network is segmented into multiple DFGs to enable multi-hop inter-DFG forwarding, each DFG is assigned both a PDFH and an SDFH. The PDFH oversees intra-DFG interactions, meaning it handles all communication among DFGMs within the same DFG. Among its responsibilities, the PDFH collects data from every DFGM in its group and decides how to forward it: if the distance to the sink is less than do , it sends the data directly; otherwise, it forwards it via one or more intermediate PDFHs (which function as PDFRs and SDFRs). In the proposed design, the sink node performs all necessary calculations to identify the most suitable PDFRs and SDFRs for each PDFH.

This method markedly improves inter-DFG data delivery, especially when substantial distances separate PDFHs from the sink node. As a result, overall energy expenditure decreases, and the network's functional lifespan is extended. Importantly, the selection of an appropriate PDFR and SDFR must satisfy the following key conditions:

(1) The Euclidean distance separating the Candidate PDFR (CPDFR) from the originating PDFH should be minimized while remaining oriented toward the sink node (i.e., in the upstream direction), and the distance must be less than do . This preference for proximity is ensured by the numerator of [Eq. 32](#), which is formulated to yield smaller values for better candidates.

(2) Among all potential upstream candidates, the CPDFR should possess the highest remaining energy (RE); this priority is enforced by dividing the average RE of all candidate PDFRs by the RE of the specific CPDFR j .

(3) The value $ASBO_{CPDF}[j]$ is additionally evaluated with reference to [Tables 4-7](#), [Figure 8](#), and [Eq. 22](#).

These considerations give rise to a novel Relay Data Fusion Cost Function (RDFCF). Using this function, each PDFH selects the relay node (SR) with the second-lowest RDFCF value and the primary forwarder (PDFR) with the absolute lowest RDFCF value. The forwarding chain continues iteratively until the aggregated data finally reaches the sink node. Accordingly, the RDFCF for any candidate PDFR j (denoted $RDFCF[j]$) is computed as follows:

$$RDFCF[j] = \left(\frac{do}{D_{PDFH_i \rightarrow CPDFR_s[j]}} \right) + \left(\frac{Avg(RE_{CPDFR_s[K]})}{RE_{CPDFR_s[j]}} \right) + \left(\frac{Avg(ASBO_{CPDFR_s[K]})}{ASBO_{CPDFR_s[j]}} \right), j = 0, 1, 2, \dots, K-1, \quad (32)$$

Where $D_{PDFH_i \rightarrow CPDFR_s[j]}$ denotes the Euclidean distance separating PDFH $_i$ from CPDFRs $[j]$, with each element representing a unique entry in the array of candidate PDFRs; on the other hand, $Avg(RE_{CPDFR_s[K]})$ stands for the mean remaining energy across the entire set of candidate PDFRs. At the same time, $RE_{CPDFR_s[j]}$ refers to the remaining energy specifically of candidate CPDFRs $[j]$.

Through this formulation, the multi-hop routing strategy implemented in the present study substantially enhances communication performance in scenarios characterized by large separations between PDFHs and the sink node. It achieves this without incurring wasteful energy expenditure during data forwarding, thereby meaningfully prolonging the operational lifetime of the entire network.

Moreover, data fusion plays a central and decisive role in this framework. By enabling intermediate PDFRs to combine and process incoming data streams from various sources, data fusion dramatically reduces redundant packet transmissions. This aggregation step greatly reduces the amount of unprocessed data delivered to the sink node, thereby

saving energy at each relay hop and alleviating overall network congestion. Consequently, the communication subsystem operates with markedly higher efficiency, further extending the network lifespan.

Algorithm 4 offers a detailed walkthrough of the proposed protocol's operation, covering scheduling procedures, data transmission flows, and the strategic application of data fusion to achieve optimized performance.

Input:

Ins: N_c (Number of DFGMs in each DFG)
Ins: C (Optimal number of DFGs)
Ins: DFGM [] (list of DFGMs for each DFG)
Ins: PDFH [] (list of PDFHs)
Ins: Max_Rounds (Maximum number of rounds)
Ins: N (Number of SNs in the network)

Output:

Outs: TSs (Scheduling TDMA TS for each SN)
Outs: AI_1, AI_2 (Array index of the minimum RDFCF and the second minimum one, respectively)
Outs: data_packets (Total Number of transmitted data packets to the sink node)

Main:

```

For Round=1 to Max_Rounds do
  For k=1 to N do
    Sink node identifies the type of SNs (either PDFH or NDFGM)
    Sink node assigns TDMA TSs based on the message arrival time and eq (29)
    Sink node advertises the schedule (TDMA TSs) → SNs
  End for
  For j=1 to C do
    For i=1 to  $N_c$  do
      If (DFGM[j]. type == "NDFGM" and it is in wakeup mode then
        DFGM[j] sends its data packet → PDFH [j]
      Elseif (DFGM[j]. type == "PDFH" and it is turn for transmission then
        Switch (type of the communication)
          // The proposed protocol without routing (single hop communication)
          Case 1:
            PDFH [j] aggregates the received data from DFGMS
            PDFH [j] sends the aggregated data directly → sink node; data_packets++;
          End Case1
          // The proposed protocol with multi-hop routing (multi-hop communication)
          Case 2:
            If (Distance  $p_{CH}[j] \rightarrow \text{Sink} < d_{\omega}$ ) then
              Array CPDFRs[0] // array of CPDFRs of size 0 (empty array)
              PDFH [j] aggregates the received data from DFGMS
              PDFH [j] sends the aggregated data directly → sink node; data_packets++;
            Elseif
              Count the number of CPDFRs → Q
              Array CPDFRs[Q] // array of CPDFRs of size Q
               $Temp_1 \leftarrow \infty$  // the minimum RDFCF
               $Temp_2 \leftarrow \infty$  // the second minimum RDFCF
              For j=0 to Q-1 do
                Calculate RDFCF[j] based on eq 27
                If (RDFCF[j] <  $Temp_1$ )
                   $Temp_2 = Temp_1$ 
                   $Temp_1 = RDFCF[j]$ 
                   $AI_1 \leftarrow j$  // Array Index of the minimum RDFCF
                Elseif (RDFCF[j] <  $Temp_2$  && RDFCF[j] !=  $Temp_1$ )
                   $Temp_2 = RDFCF[j]$ 
                   $AI_2 \leftarrow j$  // Array Index of the second minimum RDFCF
                End If
              End For
              PDFR=CPDFRs[ $AI_1$ ]
              SDFR=CPDFRs[ $AI_2$ ]
            End If
          End Case2
        End Switch
        PDFH [j] aggregates the received data from DFGMS
        PDFH [j] sends the aggregated data → PDFR1; PDFR1 sends the aggregated data → PDFR2;
        PDFR2-1 sends the aggregated data → PDFR3; PDFR3 sends the aggregated data → sink node; data_packets++;
      End If
    End For
  End For
End For
Return TSs, PDFR, SDFR, data_packets,
End Main

```

Algorithm 4. Scheduling and data transmission for the proposed protocol.

Communication overhead of the proposed protocol

The objective of this subsection is to offer readers a thorough and clear explanation of how the proposed network architecture controls and manages energy consumption. Fundamentally, the energy dissipation model adopted here relies on established radio and channel propagation models introduced in Darabkh et al. [47] and Shahzad et al. [48].

Within the proposed protocol, total energy expenditure is divided into two distinct operational phases: the setup phase and the steady-state phase.

- The first round, encompassing both the setup and steady-state phases, proceeds according to the following steps:

✓ Setup phase:

1. The sink node broadcasts a request information message to every sensor node (SN). Accordingly, the energy consumed by each SN for receiving this request information message is given by:

$$E_{RSNs}(NL_C, d_{SN \rightarrow sink}) = E_{elec} \times NL_C, \quad (33)$$

where NLC represents the length of the control message.

2. The energy expended by each SN to transmit its own information message back to the sink node.

$$E_{TSNs}(NL_C, d_{SN \rightarrow sink}) = \begin{cases} NL_C \times E_{elec} + NL_C \times \epsilon_{fs} \times d_{SN \rightarrow sink}^2, & \text{if } d_{SN \rightarrow sink} < d_0, \\ NL_C \times E_{elec} + NL_C \times \epsilon_{mp} \times d_{SN \rightarrow sink}^4, & \text{if } d_{SN \rightarrow sink} \geq d_0. \end{cases} \quad (34)$$

3. The energy required by each SN to receive the sink's table message.

$$E_{RSNs}(NL_C, d_{SN \rightarrow sink}) = E_{elec} \times NL_C. \quad (35)$$

Therefore, the total energy consumed during the setup phase (E_{Setup}) can be expressed as follows:

$$E_{Setup} = 2 \times NE_{RSNs} + NE_{TSNs}. \quad (36)$$

✓ Steady-state phase:

1. The energy consumed by sensor nodes (SNs) within a given DFG when transmitting their sensed data to the corresponding PDFH.

$$E_{TSNs}(NL_D, d_{SN \rightarrow PDFH}) = \begin{cases} NL_D \times E_{elec} + NL_D \times \epsilon_{fs} \times d_{SN \rightarrow PDFH}^2, & \text{if } d_{SN \rightarrow PDFH} < d_0, \\ NL_D \times E_{elec} + NL_D \times \epsilon_{mp} \times d_{SN \rightarrow PDFH}^4, & \text{if } d_{SN \rightarrow PDFH} \geq d_0, \end{cases} \quad (37)$$

where NLD denotes the length of the data message.

2. The energy expended by a PDFH for receiving data from all its DFGMs, performing aggregation on the collected data (including its own readings), and subsequently forwarding the fused result to the sink node.

$$E_{RPDFH}(NL_D, d_{SN \rightarrow PDFH}) = (N_c - 1) \times NL_D \times E_{elec}. \quad (38)$$

$$E_{DAPDFH}(NL_D, d) = N_c \times NL_D \times E_{fusion}. \quad (39)$$

$$E_{TPDFH}(NL_D, d_{PDFH \rightarrow sink}) = \begin{cases} NL_D \times E_{elec} + NL_D \times \varepsilon_{fs} \times d_{PDFH \rightarrow sink}^2, & \text{if } d_{PDFH \rightarrow sink} < d_0, \\ NL_D \times E_{elec} + NL_D \times \varepsilon_{mp} \times d_{PDFH \rightarrow sink}^4, & \text{if } d_{PDFH \rightarrow sink} \geq d_0. \end{cases} \quad (40)$$

Accordingly, the aggregate energy consumption across all PDFHs is expressed as follows:

$$E_{PDFHs} = C \times (E_{RPDFH} + E_{DAPDFH} + E_{TPDFH}). \quad (41)$$

3. In later rounds, if a PDFH's remaining energy exceeds the average energy level of its DFG, only the steady-state phase is executed. However, should a PDFH's energy fall below the average remaining energy of its DFG, the regrouping process is triggered. During regrouping, the sink node continues to hold accurate positional data for all SNs but lacks up-to-date information on their current energy levels. The primary sources of energy consumption in this regrouping scenario are as follows:

i. The energy used by all PDFHs to transmit disjoint messages to the sink node.

$$E_{TCH}(NL_C, d_{PDFH \rightarrow sink}) = \begin{cases} NL_C \times E_{elec} + NL_C \times \varepsilon_{fs} \times d_{PDFH \rightarrow sink}^2, & \text{if } d_{PDFH \rightarrow sink} < d_0, \\ NL_C \times E_{elec} + NL_C \times \varepsilon_{mp} \times d_{PDFH \rightarrow sink}^4, & \text{if } d_{PDFH \rightarrow sink} \geq d_0. \end{cases} \quad (42)$$

ii. The energy required by a PDFH to broadcast disjoint messages to the SNs within its DFG.

$$E_{TCH}(NL_C, d_{SN \rightarrow PDFH}) = \begin{cases} NL_C \times E_{elec} + NL_C \times \varepsilon_{fs} \times d_{SN \rightarrow PDFH}^2, & \text{if } d_{SN \rightarrow PDFH} < d_0, \\ NL_C \times E_{elec} + NL_C \times \varepsilon_{mp} \times d_{SN \rightarrow PDFH}^4, & \text{if } d_{SN \rightarrow PDFH} \geq d_0. \end{cases} \quad (43)$$

iii. The energy consumed by SNs when receiving disjoint messages from their PDFH.

$$E_{RSNs}(NL_C, d) = E_{elec} \times NL_C. \quad (44)$$

As a result, the total energy dissipation associated with the disjoint messaging mechanism is computed as follows:

$$E_{disjoin} = C \times E_{TCHs} + N \times E_{TCHs} + N \times E_{RSNs}. \quad (45)$$

Therefore, the overall energy consumed during the steady-state phase (E_{Steady}) is given by:

$$E_{Steady} = (N - C) \times E_{TSNs} + E_{CHs} + E_{disjoin}. \quad (46)$$

Ultimately, the total energy dissipation per round—encompassing both the setup and steady-state phases—can be calculated by combining [Eqs. 34 and 44](#):

$$E_{Round} = E_{Setup} + E_{Steady}. \quad (47)$$

Dealing with data fusion impairments: more optimized data fusion techniques, including building occlusion management, devices/links failures and recovery, device deployment redundancy, redundant coverage, and data redundancy

Beyond their role in minimizing energy consumption and improving communication performance, data fusion grouping strategies in IoT environments can also effectively address challenges related to building occlusion, device or link failures, and various forms of data redundancy. These considerations are vital to ensuring stable and dependable network operation over extended periods.

A particularly important aspect of the proposed protocol is its incorporation of self-healing capabilities. These capabilities refer to the network's ability to independently detect, analyze, and resolve faults or disruptions without requiring external human intervention. In large-scale or geographically remote IoT deployments—where manual maintenance is either impractical or prohibitively costly—self-healing mechanisms prove especially valuable.

Whenever the protocol automatically identifies potential anomalies or failures and assesses the incoming data to pinpoint their underlying causes, it activates repair procedures and, if necessary, restructures parts of the network topology to adapt to the changed conditions. Further elaboration on these mechanisms is provided in the subsequent subsections.

Data fusion management in building occlusion impairment

In densely built urban settings, building occlusion represents a major obstacle, especially for applications that depend on precise spatial information. Tall buildings obstruct direct line-of-sight paths, making it difficult to gather comprehensive, trustworthy data from sensors. Overcoming this challenge requires sophisticated data fusion methods that combine inputs from diverse sources, including ground-level sensor nodes (SNs).

In urban Internet of Things (IoT) deployments, occlusion caused by structures at different elevations significantly degrades wireless communication quality and reliability. Buildings of varying heights disrupt line-of-sight links, leading to multipath effects, signal fading, increased packet loss rates, and reduced throughput. The severity of occlusion grows with building height: taller structures create stronger signal barriers, larger coverage dead zones, and greater disruption for sensor nodes positioned behind them or at lower elevations. The intricate mix of low-rise and high-rise buildings in metropolitan areas further complicates the reliable propagation of signals.

To effectively counteract building occlusion within the proposed protocol, a comprehensive strategy is required—one that merges cutting-edge technological solutions with careful network design. A primary tactic involves strategically positioning PDFHs at higher elevations to minimize exposure to signal obstruction. In this study, the integration of advanced data fusion algorithms with a newly introduced residual metric, ASBO, incorporated into both the DFFF and RDFCF formulations, enables effective compensation for coverage gaps caused by occlusions. These techniques facilitate the intelligent merging of multi-sensor observations, allowing reconstruction of shadowed regions and strengthening the reliability of urban-generated datasets.

Additionally, the protocol leverages data fusion grouping reinforced by a solid redundancy framework, thereby improving the network's capacity to withstand interruptions. Multipath routing is another essential element, ensuring continuous data delivery by providing backup routes when primary paths become blocked. The protocol further incorporates adaptive algorithms that dynamically respond to current network conditions, automatically reorganizing groupings and rerouting paths to substantially reduce the adverse impact of occlusion. Signal strength is also improved by applying the two-ray ground-reflection propagation model, which enhances penetration and reduces attenuation caused by physical obstacles. Collectively, these measures not only improve communication reliability and data integrity but also enable real-time responsiveness in complex, obstacle-rich urban landscapes.

Optimized data fusion in devices/links failures and recovery

Failures of devices or communication links in Internet of Things (IoT)-enabled wireless sensor networks (WSNs) are inevitable and arise from environmental hazards, limited battery endurance, or hardware malfunctions. Data fusion serves as a cornerstone for preserving operational continuity and network robustness in the event of such failures. Optimized data fusion grouping algorithms must therefore be capable not only of detecting and recovering from these disruptions but also of effectively aggregating and reconciling information from multiple sensor nodes, even when segments of the network become unreliable.

By embedding redundancy-aware mechanisms, the proposed protocol minimizes data loss and maintains high decision accuracy. Recovery procedures are further strengthened by adaptive confidence intervals (CIs), which dynamically recalibrate the network configuration in response to observed failure patterns in real time. This approach simultaneously improves fault tolerance and accelerates the restoration of full functionality, establishing a dependable foundation for mission-critical IoT systems. The specific techniques employed to achieve this resilience are outlined below.

In the proposed protocol, the inclusion of a First-Secondary Data Fusion Head (FSDFH) and a Second-Secondary Data Fusion Head (SSDFH) is vital to maintaining network resilience and operational continuity. The primary responsibility of the FSDFH and SSDFH is to seamlessly assume the duties of the main PDFH in the event of its failure, thereby ensuring uninterrupted data aggregation and intra-DFG communication with minimal disruption.

Since the PDFH acts as the central coordinator for data collection and fusion within its DFG, its failure could lead to substantial information loss and communication breakdowns; thus, redundancy is indispensable. The presence of an FSDFH and an SSDFH enables rapid recovery from PDFH failures, greatly enhancing the protocol's overall stability and longevity.

Selection of the PDFH, FSDFH, and SSDFH occurs simultaneously through application of the MPA algorithm: the PDFH is assigned the lowest DFFF value among the DFGMs, the FSDFH receives the second-lowest DFFF, and the SSDFH receives the third-lowest DFFF. Consequently, all DFGMs transmit their data concurrently to the PDFH, FSDFH, and SSDFH. Both the FSDFH and SSDFH are also tasked with forwarding their collected data to the PDFH, ensuring efficient utilization of all designated nodes. In this arrangement, the FSDFH and SSDFH maintain identical copies of the data held by the PDFH. The PDFH then performs final aggregation and forwards the consolidated result—either directly or via designated PDFRs—to the sink node.

Should the PDFH fail, the FSDFH immediately assumes control: it aggregates the incoming data and transmits the fused packet in the same manner as the original PDFH. The FSDFH operates under the same TDMA scheduling discipline as the PDFH and remains continuously active to guarantee receipt of data from every DFGM, mirroring the PDFH's behavior. Likewise, the SSDFH is prepared to perform identical functions if both the PDFH and FSDFH become unavailable.

When developing the proposed protocol, one of the essential features incorporated to ensure network reliability and enable prompt identification of device or link failures is the use of periodic health check messages.

Under this mechanism, sensor nodes (SNs) periodically transmit their regular data messages along with embedded status updates. These updates are directed to their assigned PDFH, FSDFH, and SSDFH, or, in certain cases, sent directly to the sink node. In general, the status information encompasses battery charge levels, transmit/receive circuitry functionality, received signal strength, current data transmission state, and overall operational condition. Continuous monitoring of these parameters enables the network to detect device or link failures early, preventing them from escalating into major disruptions.

This proactive monitoring strategy extends the operational lifetime of the protocol and supports immediate corrective measures, including both localized and network-wide repairs, thereby sustaining peak performance. In essence, periodic health check messages promote dependable and uninterrupted data delivery—even amid device or link failures—by optimizing resource usage and bolstering the network's overall robustness and fault tolerance.

To handle failures effectively across different scopes, the proposed protocol implements both local repair and global repair strategies, facilitating rapid recovery and consistent high performance.

More precisely, local repair addresses issues confined to a limited segment of the network, typically affecting a single DFG or a small group of SNs. The key advantages of local repair include its speed—repairs can be executed swiftly, minimizing service disruption—and its low overhead, since only a limited portion of the network is involved, resulting in reduced communication and processing costs. The specific scenarios requiring local repair, together with the corresponding recovery procedures, are as follows:

- For a Normal DFGM (NDFGM): In the proposed scheme, DFGMs regularly share health check messages with their PDFH, FSDFH, and SSDFH. If a DFGM fails to send its data during its allocated time slot (TS), or if the PDFH detects irregular parameters through the health status details appended to the DFGM's data message, the PDFH immediately issues an updated TDMA schedule that excludes the malfunctioning DFGM.
- For a PDFH: The protocol mandates that each PDFH include its own health status information within the data message it forwards to the sink node, with copies also sent to its FSDFH and SSDFH. As previously noted, both the FSDFH and SSDFH remain continuously active and monitor the PDFH's outgoing data messages. If no transmission is received from the PDFH after a period equivalent to $1.15 \times RT$ (where RT is the round time), or if the FSDFH identifies abnormal conditions—particularly in the transmit/receive circuitry—from the PDFH's health status data, the FSDFH immediately assumes the PDFH role: it aggregates the incoming data. It transmits the fused result to the sink node. Similarly, the SSDFH performs the same takeover if the FSDFH fails. More specifically, if $1.3 \times RT$ elapses without any transmission from either the PDFH or the FSDFH, the SSDFH steps in as the new PDFH. This graduated timeout approach helps eliminate the need for global repair in most cases.
- For an FSDFH or SSDFH: According to the protocol, both the FSDFH and SSDFH independently append their health status information to their data messages and forward these to the sink node (via their PDFH). Should the sink node detect irregular parameters in these status updates, the affected FSDFH or SSDFH is removed from the sink's table message and excluded from future network operations. However, since this situation is not time-critical, full regrouping will be deferred until the next batch cycle.

In contrast to local repair, global repair involves detecting and correcting failures that affect a substantial portion—or even the entire—network. Global repair offers the benefit of comprehensive restoration, addressing widespread issues to ensure long-term network health, efficient resource allocation, and prevention of recurring problems. Within the proposed protocol, global repair is integrated into the dynamic regrouping process—a fundamental recovery mechanism that enhances network resilience and operational efficiency. This process involves reassigning PDFH, FSDFH, and SSDFH roles within DFGs, as well as reconfiguring inter-DFG forwarding paths (PDFRs, FSDFRs, and SSDFRs). It is indispensable for adapting to changes such as node failures, energy exhaustion, or shifts in network topology.

In summary, the primary goal of dynamic regrouping is to achieve equitable energy distribution among SNs, thereby maximizing the network's overall lifespan. Additionally, it reduces data loss and minimizes communication latency by delegating critical responsibilities to the most suitable nodes. The specific conditions that trigger global repair, along with the associated recovery methods, are as follows:

- If the sink node fails to receive any data from a PDFH, its FSDFH, its SSDFH, or the associated PDFRs after $1.45 \times RT$ (i.e., 45% beyond the expected reception time), accounting for a 15% additional delay tolerance per role.
- If the sink node receives health status messages from the PDFH, FSDFH, and SSDFH and identifies abnormal parameters within those messages.

It should be emphasized that multipath routing is a supplementary feature of the proposed protocol, designed not only to enhance the resilience, efficiency, and dependability of data fusion and transmission throughout the network but also to facilitate both equal and unequal load distribution.

In practice, rather than relying on a solitary route, this approach establishes multiple independent paths connecting sensor nodes (SNs) to the sink node. By doing so, it mitigates risks arising from traffic congestion, fluctuating network conditions, and link or node failures. In addition, multipath routing spreads the communication load across diverse routes, thereby achieving balanced energy consumption and prolonging the operational lifetime of SNs by preventing overuse of individual paths. In environments that are both highly dynamic and severely resource-limited, this built-in redundancy significantly boosts fault tolerance. It elevates overall service quality, making it a fundamental technique for sustaining uninterrupted, high-performance communication and strengthening the network's ability to withstand disruptions.

Ultimately, the protocol implements multipath communication via a newly developed Relay Data Fusion Cost Function (RDFCF) that selects the PDFR (primary forwarder), FSDFR (first secondary forwarder), and SSDFR (second secondary forwarder), corresponding to primary and backup routes, respectively. Detecting routing failures—specifically failures of PDFR, FSDFR, or SSDFR—is of utmost importance. Since these entities are themselves sensor nodes within the network, any such failures are managed using precisely the same local repair and global repair procedures described earlier.

Device deployment redundancy and redundant coverage optimizations

When designing an effective data fusion grouping protocol tailored for IoT networks, careful attention must be given to the implications of redundant device placement and the challenges posed by overlapping (redundant) coverage within the monitored area. Excessive overlap in sensing ranges or unnecessary duplication of devices can lead to wasteful energy expenditure and suboptimal use of network resources. The proposed protocol explicitly accounts for these factors, resulting in substantially improved network longevity and operational effectiveness.

Redundant device deployment involves strategically placing multiple SNs in close physical proximity to ensure continuity and fault tolerance. While such redundancy strengthens network robustness against individual node failures, it can also lead to unnecessary repeated transmissions and inefficient energy use. To counteract these drawbacks, the proposed protocol applies sophisticated data fusion methods that consolidate duplicate readings from neighboring SNs, thereby significantly reducing the total volume of data forwarded across the network and conserving precious energy reserves. By performing aggregation at designated fusion nodes (PDFHs), the protocol discards redundant or low-value information while preserving the accuracy and completeness of the collected dataset. Consequently, the protocol guarantees that the density of each DFG remains exactly equal to, but never exceeds, the overall network density, as illustrated in **Table 3**. This constraint effectively prevents the unnecessary deployment of duplicate devices.

Redundant coverage, on the other hand, occurs when multiple SNs' sensing fields overlap, covering the same physical region. This overlap can generate excessive communication traffic and redundant data copies. The protocol counters this issue through intelligent fusion mechanisms that intelligently merge overlapping sensor readings into a single coherent

dataset, thereby eliminating redundancy at its source. Within the proposed framework, SNs activate exclusively during their assigned TDMA time slot (TS), transmitting their data only to the corresponding PDFH during that window before returning to sleep mode according to the TDMA schedule. Should an SN mistakenly transmit to a different PDFH (one to which it does not belong), the receiving PDFH discards the data, as the SN is not a member of its DFG. This mechanism eliminates intra-DFG interference. To address inter-DFG interference, the protocol employs Direct Sequence Spread Spectrum (DSSS), assigning each DFG a unique spreading code distinct from those of adjacent DFGs. This code differentiation ensures that neighboring DFGs do not interfere with one another, enabling efficient data fusion, optimal resource utilization, robust fault tolerance, and stable long-term operation.

Data redundancy and data fusion optimization

Data redundancy refers to the intentional replication of information across multiple sensor nodes to safeguard data integrity and availability. While redundancy can enhance fault tolerance, it must be carefully managed to avoid excessive energy and bandwidth consumption. Data fusion complements this by intelligently combining inputs from diverse sources to generate more accurate, concise, and actionable information. By drastically reducing the amount of raw data that needs to be transmitted and processed, data fusion effectively controls redundancy and drives greater efficiency, particularly in resource-constrained sensor networks.

The proposed protocol seamlessly integrates data redundancy management with advanced fusion techniques to ensure continuous communication, maximize resource efficiency, and elevate overall network performance. Redundancy levels are dynamically tuned to prevailing network conditions, while fusion strategies refine the data-acquisition and processing pipeline. Together, these enhancements achieve balanced energy distribution, lower end-to-end latency, and consistently high reliability. The following specialized techniques are employed to optimize the protocol's handling of data fusion and redundancy:

- **Redundant data aggregation:** The proposed protocol implements redundant data aggregation, collecting, processing, and consolidating multiple copies of the same information from different SNs within a DFG at the PDFH, FSDFH, or SSDFH. This process removes duplicates, extracts essential features, and forwards only the refined result to the sink node. Through this approach, data integrity and completeness are preserved even if individual packets are lost or corrupted during transmission. By substantially reducing the amount of data sent, the technique optimizes bandwidth usage, enhances fault tolerance, and improves overall data reliability. Furthermore, it reduces total energy expenditure since fewer packets need to traverse long distances. Notably, this aggregation strategy is uniformly applied across all IoT application scenarios outlined in **Table 3**, equipping the protocol to effectively address the challenges inherent to dynamic, resource-limited IoT environments.

- **Energy-adaptive and selective redundancy methods:** Further details on these techniques are provided in the following sections:

- i. **Energy-adaptive redundancy levels:** Within the proposed protocol, energy-adaptive redundancy levels involve dynamically adjusting both the amount and placement of redundant data and operations according to the current remaining energy of sensor nodes (SNs) and prevailing real-time network conditions. This adaptive mechanism continuously evaluates network health and dynamically determines the optimal degree of redundancy. For instance, when network stability is at risk, the protocol increases redundancy to ensure dependable data delivery and enhanced fault tolerance—provided that SN energy levels remain above the critical energy threshold. Conversely, under stable conditions, redundancy is reduced to conserve bandwidth and energy resources. This inherent flexibility allows the network to eliminate unnecessary overhead while preserving superior performance and robustness. By modulating redundancy in response to key factors such as SN energy status, current traffic load, and data priority, the approach achieves an effective balance between reliability guarantees and efficient resource consumption.

- ii. **Selective redundancy:** Selective redundancy, as implemented in the proposed protocol, is a targeted strategy aimed at strengthening network resilience and reliability through deliberate replication of only the most essential data and functions

at carefully chosen points or segments of the network. Redundancy is applied selectively based on the information's criticality and the application's specific requirements. For example, in IoT healthcare scenarios—such as continuous patient monitoring—critical data streams demand higher levels of redundancy. As outlined in **Table 3**, the proposed protocol supports a diverse range of IoT applications, which are classified into medical and non-medical categories. In medical applications, priority is assigned to redundant handling of vital data: key nodes such as PDFHs, FSDFHs, and SSDFHs are identified as critical and instructed to maintain duplicate copies of high-priority data and operations, given their responsibility for time-sensitive and mission-critical information. This focused approach keeps overall network resource consumption relatively low while enabling rapid recovery from node failures or link disruptions. Selective redundancy is also integrated into the multipath routing framework to establish backup transmission paths in case the primary route becomes unavailable. By concentrating redundancy where it matters most, this tailored strategy minimizes the trade-off between fault tolerance and resource efficiency, thereby ensuring uninterrupted operation of essential services and improving the network's overall stability and performance.

In summary, building on the techniques described above, the protocol introduces a novel parameter called the Redundancy Energy Threshold (RET). Notably, SNs whose remaining energy falls below RET will automatically restrict redundancy measures and prioritize aggressive data fusion to conserve power. In contrast, SNs handling high-importance (urgent and critical) data and possessing energy levels above RET are permitted—and encouraged—to store and transmit redundant copies as needed.

Results and Discussion

This section presents a thorough evaluation and analysis of the performance achieved by the newly proposed protocol through an extensive series of simulations. First, the simulation environment and the specific parameters employed are described. Next, the key performance metrics used to quantify the protocol's improvements are defined. Finally, the main simulation outcomes and key insights are presented and discussed.

Simulation environment and setting parameters

Growing academic interest in enhancing IoT network performance has prompted the development and comparison of various simulation platforms and modeling tools. This trend stems largely from the practical difficulties of real-world experimentation, including high complexity, long durations, and substantial costs. The present study selects an appropriate simulation framework capable of representing a wide range of smart service scenarios and smart city use cases. All simulations were performed using MATLAB version 9.8.0.1323502 (R2020a) on a laptop equipped with a 12th Gen Intel Core i7-12700H processor, 16 GB of RAM, and running Windows 11 Pro.

The network model consists of a rectangular sensing field populated with randomly distributed sensor nodes (SNs) and a single external sink node, as illustrated in **Figure 9**. To ensure fair comparisons, the simulation parameters—summarized in **Table 9**—were aligned with those used by competing protocols. In the intra-DFG communication phase, SNs transmit R-bit data packets to their respective PDFHs. In the inter-DFG phase, PDFHs forward aggregated data either directly to the sink node or via designated PDFRs (PRs). To mitigate the impact of randomness in SN placement, each scenario was executed 50 times, and the reported results represent the average across all runs.

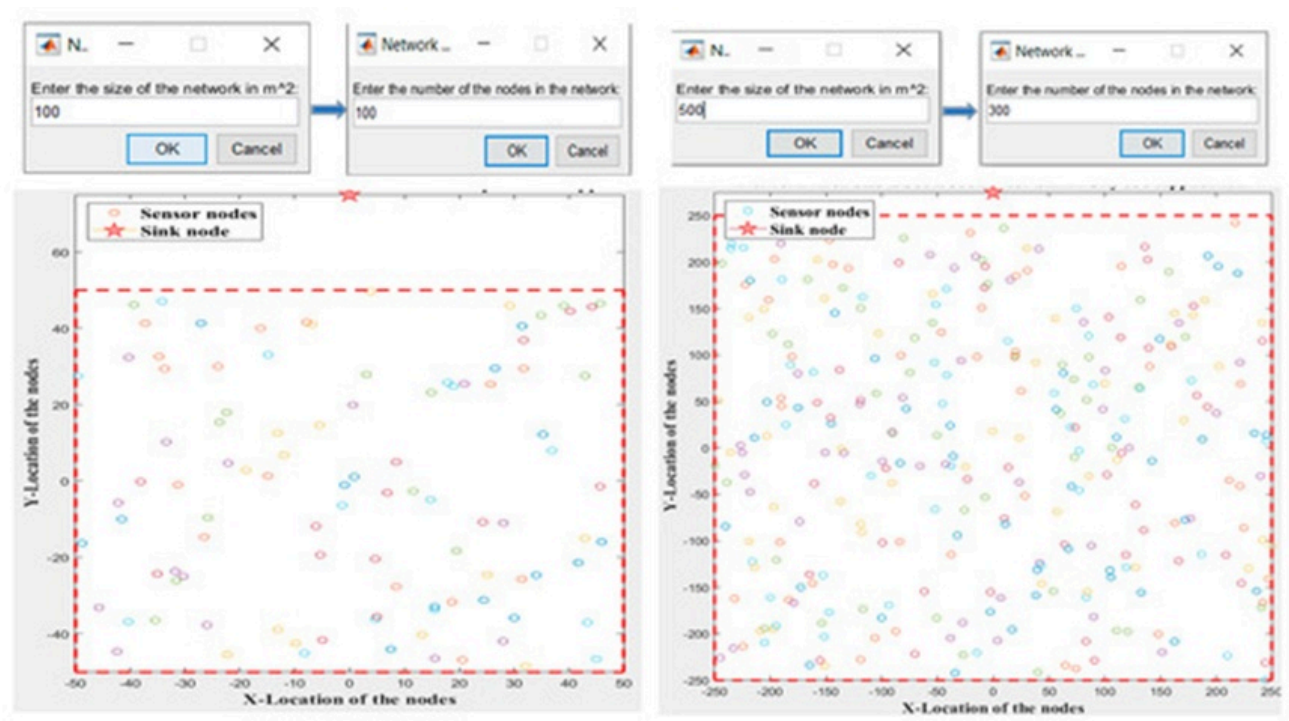


Figure 9. Distribution of SNs in the first and second simulation scenarios.

Table 9. Simulation parameters of the proposed protocol.

Category	Parameter name	Value in scenario 1	Value in scenario 2
Network parameters	Network sensing field dimensions (W × W)	100 × 100 m ²	500 × 500 m ²
	Total number of sensor nodes (N)	100 nodes	300 nodes
	Sink node coordinates (x, y)	(0, 75)	(0, 275)
	Number of Data Fusion Groups (C)	Dynamic	Dynamic
	Sensor node deployment pattern	Random	Random
Packet parameters	Length of data packet (NLD)	4000 bits	4000 bits
	Length of control packet (NLC)	200 bits	200 bits
	Number of packets transmitted	Variable	Variable
Simulation environment	Simulation platform	MATLAB R2020	MATLAB R2020
Energy model parameters	Initial energy per sensor node (E ₀)	1 J	1 J
	Electronic circuit energy consumption (E _{elec})	50 nJ/bit	50 nJ/bit
	Energy required for data fusion/aggregation (E _{fusion})	5 nJ/bit	5 nJ/bit

	Free-space path loss coefficient (ϵ_{fs})	10 pJ/bit/m ²	10 pJ/bit/m ²
	Multi-path fading coefficient (ϵ_{mp})	0.0013 pJ/bit/m ⁴	0.0013 pJ/bit/m ⁴
	Threshold transmission distance (d_0)	87.7 m	87.7 m
Execution parameters	Total number of simulation rounds	Variable	Variable
	Maximum number of rounds per experiment	60,000 rounds	60,000 rounds
	Number of independent simulation runs	50 runs	50 runs

Performance metrics

A comprehensive set of simulations was conducted to rigorously evaluate the robustness and overall effectiveness of the proposed protocol, employing a diverse array of performance indicators. These simulations provide a detailed assessment of network durability and offer clear insight into the rate at which sensor nodes (SNs) deplete their available energy reserves.

i. Network Lifetime:

This metric is defined as the total number of rounds until the final SN exhausts its energy and ceases functioning. Network lifetime is quantified using two primary benchmarks:

- Half Node to Die (HND): The number of rounds elapsed from the initial deployment of the IoT network until 50% of the SNs have depleted their energy and become inoperative.
- Last Node to Die (LND): The total number of rounds required for every SN in the network to exhaust its energy supply and cease operation. This is determined by tracking the count of active (alive) and inactive (dead) SNs over time.

ii. Total Energy Consumption:

This metric quantifies the cumulative energy expended by all SNs at the end of each operational round. Energy usage is measured in Joules.

iii. Throughput:

Throughput evaluates the network's data-handling efficiency by recording the total volume of data packets—measured in bits—that are successfully delivered to the sink node by the time the Last Node to Die (LND) is reached. Expressed in bits per second (bps), higher throughput indicates greater capacity for effective, sustained data transmission across the network.

iv. Average Delay:

This metric represents the average end-to-end travel time experienced by each data packet from its source SN to the sink node, expressed in milliseconds (ms). It reflects the typical latency encountered by packets as they traverse the network, serving as an indicator of data-flow efficiency. The average is computed by aggregating transmission delays across multiple rounds and then taking the mean, thereby providing a reliable estimate of the expected time required for successful packet delivery.

Simulation scenarios and results

The current section reviews and interprets the simulation outcomes in detail. Performance comparisons are drawn among several existing schemes—LDIWPSOC [40], PSO-EEC [39], OFCA [41], NPSOP [42]—as well as the newly introduced protocol in its non-routed form and its multi-hop variant. Simulations rely on the configuration parameters listed in Table 9 and explore two distinct cases. The first case involves randomly placing 100 sensor nodes across a $100 \times 100 \text{ m}^2$ region; the second deploys 300 nodes over a $500 \times 500 \text{ m}^2$ field. To align with the baseline methods, both scenarios initially set the number of DFGs to 10% and 5% of the total node count, respectively. Unlike those approaches, however, our method dynamically determines the appropriate number of DFGs. Evaluation focuses on three core indicators: overall network survival time, cumulative energy usage, and achieved throughput.

Simulation scenario 1: alive nodes versus diversity of network sizes (network scalability)

evaluationSimulation Scenario 1: Alive Nodes versus Diversity of Network Sizes (Network Scalability) Evaluation

The ability to scale effectively is a critical requirement for IoT systems, which are routinely deployed for broad-area sensing and continuous data gathering across many domains. Protocols that demonstrate robust scalability in both data fusion grouping and data forwarding can gracefully handle increases in node population or coverage area while sustaining acceptable energy efficiency, throughput, and service quality.

This experiment specifically studies the influence of varying deployment scale on the longevity of the proposed protocol. We examine two different coverage regions— $100 \times 100 \text{ m}^2$ and $500 \times 500 \text{ m}^2$ —together with node populations of 100, 300, and 600. Figure 10 plots how long the network remains operational (measured via Half Node to Die and Last Node to Die metrics) as both area and node count increase. Results are compared against LDIWPSOC, PSO-EEC, OFCA, NPSOP, the proposed protocol operating without routing, and the proposed protocol using multi-hop forwarding.

As shown in Figure 10, extending the deployment area results in a noticeable reduction in network lifetime. This outcome aligns with expectations: wider areas result in longer average intra-DFG communication distances (between SNs and their PDFHs) and inter-DFG forwarding (from PDFHs to the sink), accelerating energy drain across the board. Even so, every competing scheme underperforms our protocol across all tested configurations.

Several distinctive design choices account for this superiority. Most importantly, the network is subdivided into an adaptively chosen number of hexagonal DFGs. This hexagonal tessellation delivers multiple advantages, particularly in terms of coverage quality and uniformity. Compared with conventional rectangular or square partitioning, hexagons naturally produce more even spatial distribution, minimize uncovered zones or redundant overlap, and place neighboring nodes at roughly equal distances from one another. These geometric benefits reduce mutual interference, optimize each node's effective transmission radius, and limit signal degradation.

The arrangement also supports better energy management: nodes coordinate more efficiently, waste fewer resources on unnecessary retransmissions, and extend their collective operational duration. These properties prove especially valuable in expansive IoT deployments—such as those required for smart urban infrastructure—where dependable connectivity, low energy footprint, and consistent data delivery are paramount.

Another decisive factor is the superior performance of the MPA optimization technique employed for PDFH election. As summarized in Table 2, MPA strikes an excellent compromise between global search (exploration) and local refinement (exploitation), enabling it to solve the challenging PDFH assignment problem reliably. Unlike many conventional heuristics, it consistently avoids entrapment in suboptimal solutions. Its flexibility proves particularly valuable during PDFH selection, where it intelligently weights critical attributes and consistently favors high-residual-energy candidates while avoiding nodes nearing depletion. This strategy helps delay individual node failures and distributes workload more evenly throughout the network.

Moreover, by tightly coupling the data-fusion grouping phase with routing decisions, the protocol drastically reduces control-packet overhead during topology setup and maintenance. Energy expenditure is carefully optimized in both phases, yielding substantially better scalability than the reference methods.

In addition to lowering energy use and improving communication performance, the data fusion grouping framework also effectively mitigates several common network impairments—such as building-induced occlusion, node or link failures, and excessive data duplication. Successfully addressing these challenges is essential for achieving long-term operational stability and dependability.

Tables 10 and 11 quantify the relative improvement in network lifetime delivered by the proposed protocol (without routing and with multi-hop routing, respectively) when compared against each baseline method across the different network scales evaluated.

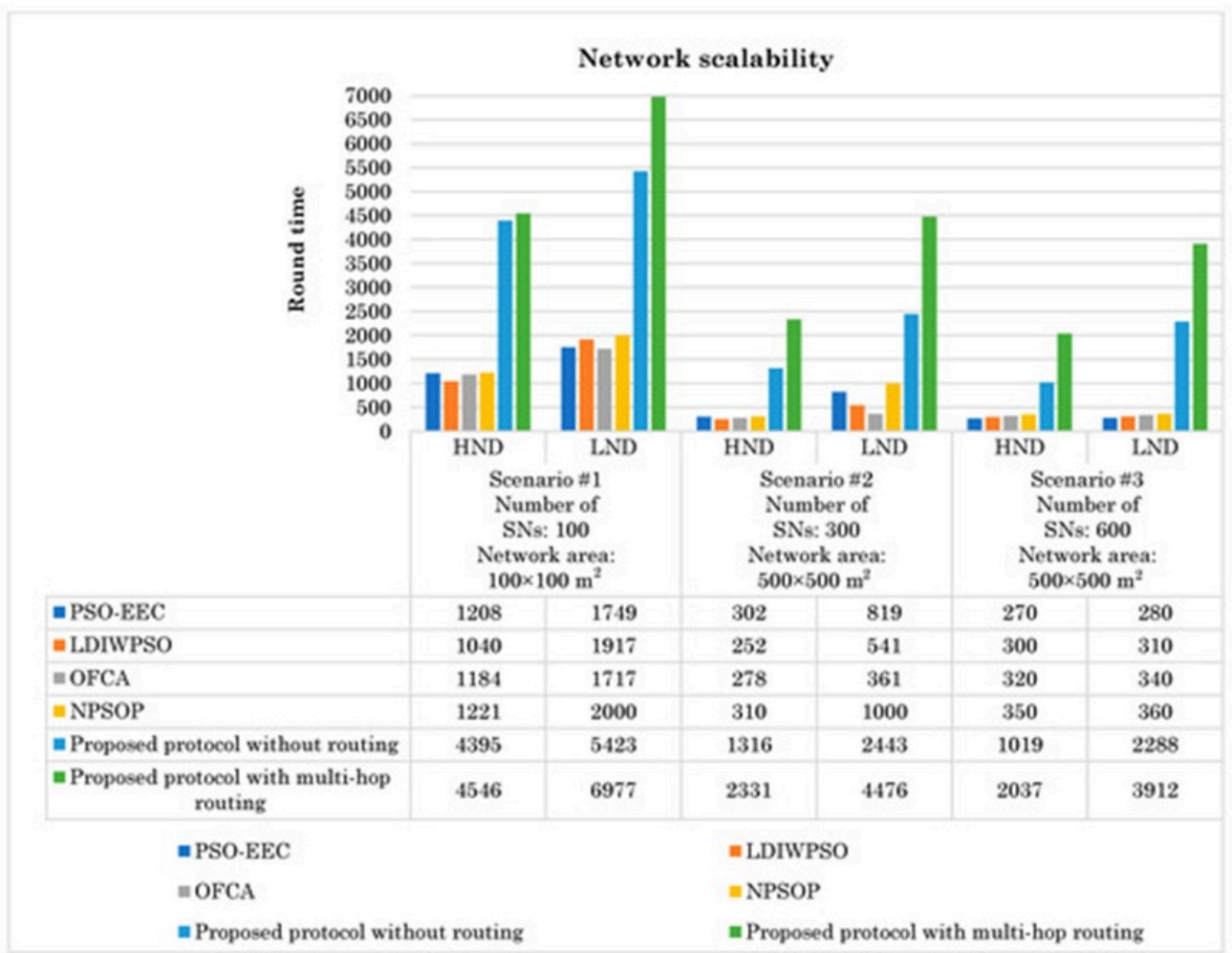


Figure 10. Evolution of the number of alive nodes over simulation rounds: side-by-side comparison with reference protocols (PSO-EEC [39], LDIWPSOC [40], OFCA [41], NPSOP [42]) for various IoT network scales and densities.

Table 10. Percentage gains in network lifetime achieved by the proposed protocol (routing-free version) relative to the benchmark protocols across the tested network sizes.

Scenario description	Metric	PSO-EEC	LDIWPSOC	OFCA	NPSOP
----------------------	--------	---------	----------	------	-------

Scenario #1 Number of sensor nodes (SNs): 100 Network deployment area: 100 × 100 m ²	HND	264%	323%	271%	260%
	LND	210%	183%	216%	171%
Scenario #2 Number of sensor nodes (SNs): 300 Network deployment area: 500 × 500 m ²	HND	336%	422%	373%	325%
	LND	198%	352%	577%	144%
Scenario #3 Number of sensor nodes (SNs): 600 Network deployment area: 500 × 500 m ²	HND	277%	240%	218%	191%
	LND	717%	638%	573%	536%

Table 11. Percentage gains in network lifetime achieved by the proposed protocol (multi-hop version) relative to the benchmark protocols across the tested network sizes.

Scenario description	Metric	PSO-EEC	LDIWPSOC	OFCA	NPSOP	Proposed protocol (without routing)
Scenario #1 Number of sensor nodes (SNs): 100 Network deployment area: 100 × 100 m ²	HND	276%	337%	284%	272%	3%
	LND	299%	264%	306%	249%	29%
Scenario #2 Number of sensor nodes (SNs): 300 Network deployment area: 500 × 500 m ²	HND	672%	825%	738%	652%	77%
	LND	447%	727%	1140%	348%	83%
Scenario #3 Number of sensor nodes (SNs): 600 Network deployment area: 500 × 500 m ²	HND	654%	579%	537%	482%	100%
	LND	1297%	1162%	1051%	987%	71%

Simulation scenario 2: alive nodes evaluation

In this scenario, we assess network durability by tracking the number of remaining active sensor nodes (SNs) using the Half Node to Die (HND) and Last Node to Die (LND) indicators. **Figure 11** visually summarizes the findings by comparing the survival curves of our proposed protocol—implemented both without routing and with multi-hop forwarding—against those of the reference schemes: LDIWPSOC, PSO-EEC, OFCA, and NPSOP.

The curves in **Figure 11** reveal a consistent, predictable behavior shared by all evaluated protocols: the count of alive SNs steadily declines as the simulation progresses through additional rounds. This gradual drop is entirely anticipated, since prolonged operation leads to cumulative energy expenditure driven by frequent message exchanges and increasingly long communication paths, gradually pushing the network toward complete depletion.

The markedly superior performance delivered by our protocol stems from several carefully engineered improvements. One of the most impactful is adopting a dynamic PDFH count that adapts to the network’s actual requirements at any given moment, rather than enforcing a fixed number of cluster heads. Much of this gain arises because the majority of intra-group communication occurs over short distances governed by the Friis free-space propagation model, which incurs far lower energy cost than long-range transmission.

A second major enhancement lies in the highly effective mechanism for electing PDFHs, FSDFHs, and SSDFHs, powered by the MPA optimization algorithm. Beyond simply selecting an advanced optimizer, the protocol integrates five carefully chosen residual metrics (RMs) into the Data Fusion Fitness Function (DFFF). This combination produces dramatically better outcomes. Among these metrics, the newly introduced ASBO stands out as a pioneering contribution:

it explicitly accounts for real-world signal degradation caused by physical obstacles, bringing the model significantly closer to practical deployment conditions where buildings, walls, and other barriers impair wireless links.

Another decisive factor is the protocol's use of batch-based data fusion grouping, which sharply contrasts with the round-by-round regrouping employed by many competing approaches. In our design, regrouping is triggered only when a PDFH's remaining energy falls below the average energy level of its DFG members (i.e., below the energy threshold E_{th}). The sink node centrally oversees this process, promoting uniform energy drain across all SNs. By avoiding unnecessary regrouping events, the protocol substantially reduces control message overhead—an effect that directly translates into prolonged network lifetime.

Additionally, once PDFHs have been selected, the sink node distributes a single TDMA scheduling table message containing all necessary timing and role information. Because this critical configuration message is broadcast only once per batch, the number of subsequent control exchanges among the sink, PDFHs, and DFGMs is minimized, further lowering energy expenditure and extending operational endurance.

The multi-hop version of the proposed protocol achieves particularly impressive lifetime gains, clearly outperforming all baseline methods. This improvement results from replacing direct (single-hop) transmissions to the sink with carefully orchestrated multi-hop forwarding via selected PDFRs (PRs). The choice of each relay node is guided by a novel Relay Data Fusion Cost Function (RDFCF) that incorporates multiple decisive criteria: proximity to the sink, residual energy of candidate relays, and the ASBO metric to account for environmental obstructions. These combined innovations produce demonstrably superior results.

Across all tested configurations and network scales, both the HND and LND values recorded for our protocol significantly exceed those of the competing schemes. To quantify these gains precisely, **Table 12** reports the percentage improvement in HND and LND achieved by the proposed protocol relative to each benchmark method.

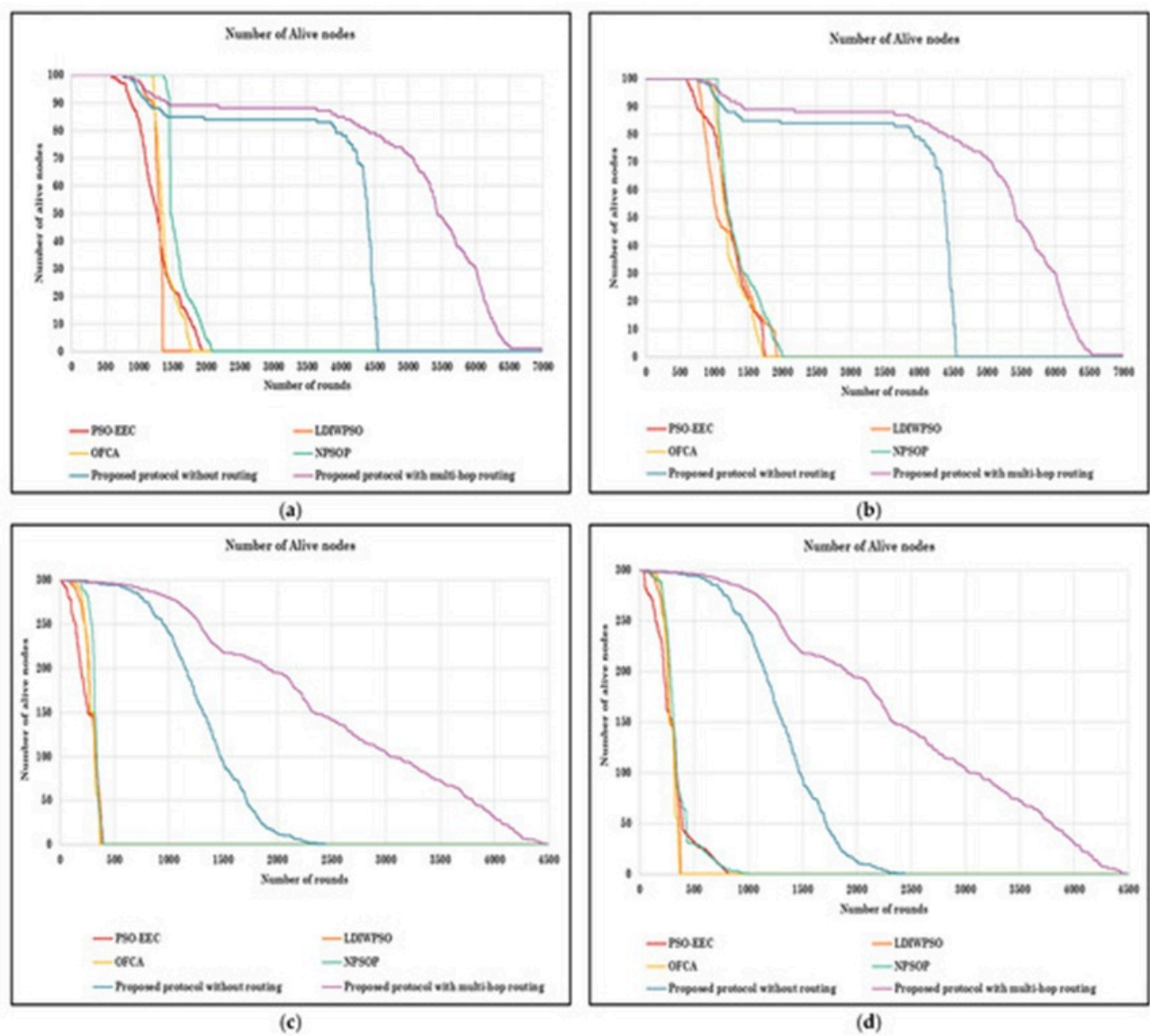


Figure 11. Number of alive nodes versus number of rounds: side-by-side comparison with reference protocols for Scenarios 1 and 2 (PSO-EEC [39], LDIWPSOC [40], OFCA [41], NPSOP [42]): (a) Scenario #1: 100 × 100 m² area, 100 nodes, 10% DFHs, (b) Scenario #1: 100 × 100 m² area, 100 nodes, 5% DFHs, (c) Scenario #2: 500 × 500 m² area, 300 nodes, 10% DFHs, (d) Scenario #2: 500 × 500 m² area, 300 nodes, 5% DFHs.

Table 12. Comparative improvement ratios for HND and LND: proposed protocol versus counterpart protocols.

Scenario description	Protocol compared	HND (rounds)	HND improvement ratio – proposed protocol without routing	HND improvement ratio – proposed protocol with multi-hop routing	LND (rounds)	LND improvement ratio – proposed protocol without routing	LND improvement ratio – proposed protocol with multi-hop routing

Scenario #1 Number of SNs: 100 Number of DFHs: 10% Network area: 100 × 100 m ²	PSO-EEC	1310	235%	314%	1933	135%	261%
	LDIWPSOC	1288	241%	321%	1352	236%	416%
	OFCA	1415	211%	283%	1776	156%	293%
	NPSOP	1465	200%	270%	2078	119%	236%
	Proposed protocol without routing	4395	NA	23%	4546	NA	53%
	Proposed protocol with multi-hop routing	5423	NA	NA	6977	NA	NA
Scenario #1 Number of SNs: 100 Number of DFHs: 5% Network area: 100 × 100 m ²	PSO-EEC	1208	264%	349%	1749	160%	299%
	LDIWPSOC	1040	323%	421%	1917	137%	264%
	OFCA	1184	271%	358%	1717	165%	306%
	NPSOP	1221	260%	344%	2000	127%	249%
	Proposed protocol without routing	4395	NA	23%	4546	NA	53%
	Proposed protocol with multi-hop routing	5423	NA	NA	6977	NA	NA
Scenario #2 Number of SNs: 300 Number of DFHs: 10% Network area: 500 × 500 m ²	PSO-EEC	256	414%	811%	396	517%	1030%

	LDIWPSOC	277	375%	742%	380	543%	1078%
	OFCA	307	329%	659%	367	566%	1120%
	NPSOP	318	314%	633%	397	515%	1027%
	Proposed protocol without routing	1316	NA	77%	2443	NA	83%
	Proposed protocol with multi-hop routing	2331	NA	NA	4476	NA	NA
Scenario #2 Number of SNs: 300 Number of DFHs: 5% Network area: 500 × 500 m ²	PSO-EEC	302	336%	672%	819	198%	447%
	LDIWPSOC	252	422%	825%	541	352%	727%
	OFCA	278	373%	738%	361	577%	1140%
	NPSOP	310	325%	652%	1000	144%	348%
	Proposed protocol without routing	1316	NA	77%	2443	NA	83%
	Proposed protocol with multi-hop routing	2331	NA	NA	4476	NA	NA

NA: not applicable (self-comparison yields no meaningful value).

Simulation scenario 3: energy consumption evaluation

This subsection examines simulation outcomes for energy expenditure patterns across the evaluated protocols. We specifically contrast the cumulative energy consumption of our advanced protocol against four established schemes: LDIWPSOC, PSO-EEC, OFCA, and NPSOP. **Figure 12** clearly demonstrates that energy usage increases proportionally with greater data transfer volumes and longer communication ranges. Although the benchmark protocols achieve reasonable efficiency, they are consistently outperformed by our proposed protocol—both in its routing-free configuration and, to an even greater extent, when employing multi-hop forwarding.

Several carefully designed innovations enable our protocol to achieve markedly lower energy consumption. These include an optimized scheduling mechanism, a highly effective DFG-formation process that determines the optimal number of

PDFHs, and other efficiency-focused enhancements. A particularly important feature is the delayed regrouping strategy: reorganization occurs only after multiple rounds. It is activated solely when a PDFH's remaining energy drops below the average energy level of its DFG members. This selective triggering prevents low-energy nodes from being assigned as PDFHs too early, thereby preserving their limited reserves and substantially prolonging overall network survival.

By proactively addressing various network impairments, the protocol delivers a significant leap forward in energy-efficient operation, establishing a new benchmark for sustainable IoT networking.

Furthermore, the incorporation of backup cluster heads—namely FSDFH, SSDFH, FSDFR, and SSDFR—plays an indispensable role in preserving data transmission stability, reliability, and continuity. These secondary nodes stand ready to collect data from sensor nodes and forward it either to the sink or to subsequent relays. Given that IoT deployments frequently experience node failures due to battery exhaustion, wireless link disruptions, or harsh environmental conditions, the presence of backup heads ensures uninterrupted operation even when a primary PDFH becomes unavailable. The FSDFH and SSDFH can immediately assume responsibility for aggregation and forwarding, thereby minimizing data loss, reducing service interruptions, and sustaining acceptable performance levels. Such redundancy is especially valuable in mission-critical domains such as remote health monitoring and industrial process control, where any interruption in data flow could compromise safety or operational effectiveness.

In the multi-hop variant of the protocol, the routing mechanism—built around a novel Relay Data Fusion Cost Function (RDFCF)—further reduces energy consumption by shortening individual transmissions to the sink node. This approach also achieves better load distribution among PDFHs farther from the sink. **Table 13** summarizes the percentage reduction in energy consumption achieved by our protocol relative to each comparison method across a wide range of simulation rounds.

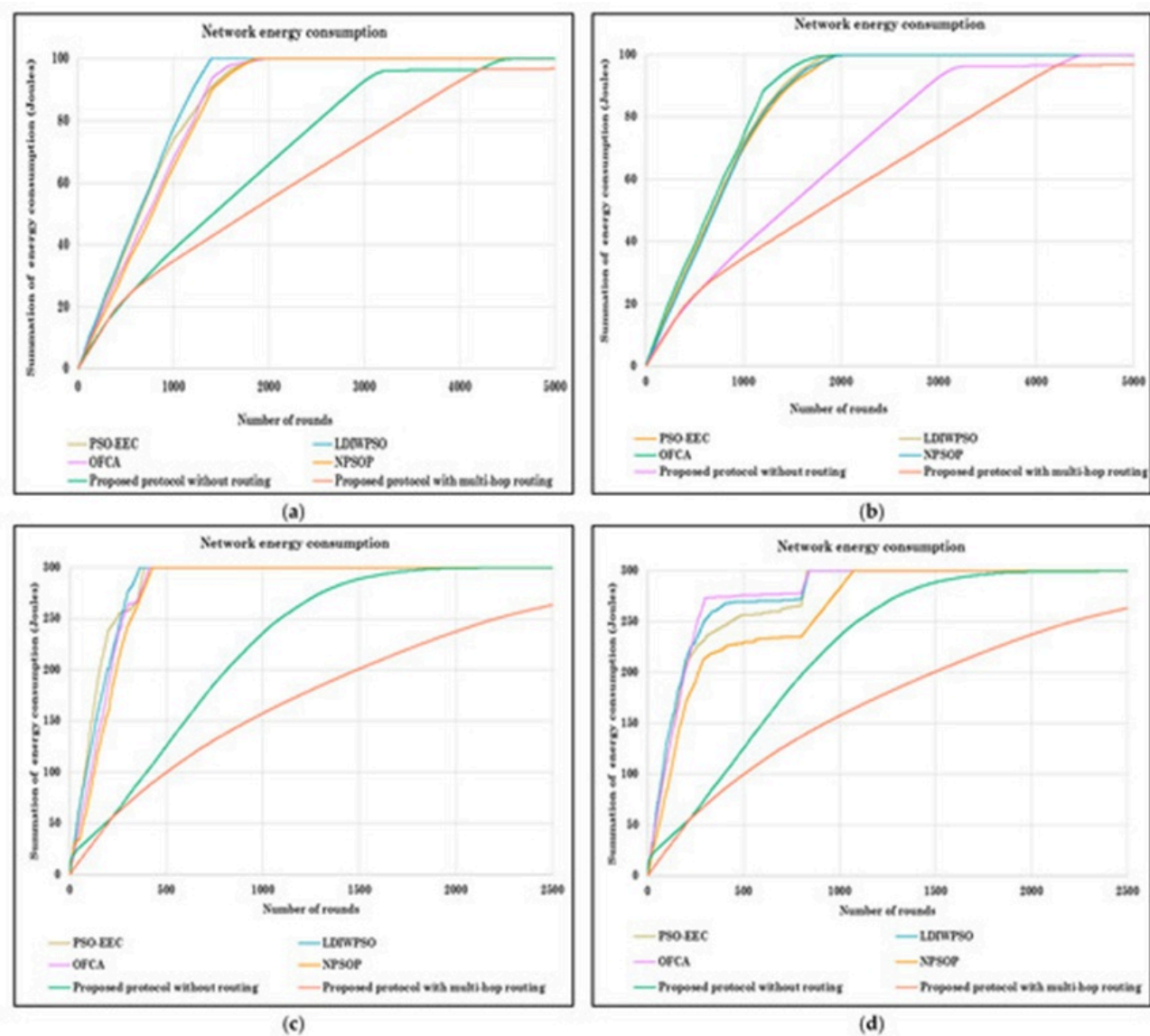


Figure 12. Total network energy consumption versus number of rounds: comparison with reference protocols for Scenarios 1 and 2 (PSO-EEC [39], LDIWPSOC [40], OFCA [41], NPSOP [42]): (a) Scenario #1: 100 × 100 m² area, 100 nodes, 10% DFHs, (b) Scenario #1: 100 × 100 m² area, 100 nodes, 5% DFHs, (c) Scenario #2: 500 × 500 m² area, 300 nodes, 10% DFHs, and (d) Scenario #2: 500 × 500 m² area, 300 nodes, 5% DFHs.

Table 13. Percentage improvement in energy consumption for the proposed protocol compared to the counterpart protocols.

Scenario description	Protocol compared	Improvement ratio – proposed protocol without routing	Improvement ratio – proposed protocol with multi-hop routing
Scenario #1 Number of SNs: 100 Number of CHs: 10% Network area: 100 × 100 m ²	PSO-EEC	135%	260%

	LDIWPSOC	225%	399%
	OFCA	133%	258%
	NPSOP	143%	273%
	Proposed protocol without routing	NA	53%
	Proposed protocol with multi-hop routing	NA	NA
Scenario #1 Number of SNs: 100 Number of CHs: 5% Network area: 100 × 100 m ²	PSO-EEC	131%	254%
	LDIWPSOC	137%	264%
	OFCA	126%	247%
	NPSOP	130%	253%
	Proposed protocol without routing	NA	53%
	Proposed protocol with multi-hop routing	NA	NA
Scenario #2 Number of SNs: 300 Number of CHs: 5% Network area: 500 × 500 m ²	PSO-EEC	194%	438%
	LDIWPSOC	190%	432%
	OFCA	190%	432%
	NPSOP	128%	317%
	Proposed protocol without routing	NA	83%
	Proposed protocol with multi-hop routing	NA	NA
Scenario #2 Number of SNs: 300 Number of CHs: 10% Network area: 500 × 500 m ²	PSO-EEC	534%	1063%
	LDIWPSOC	573%	1133%
	OFCA	490%	981%
	NPSOP	468%	941%
	Proposed protocol without routing	NA	83%

	Proposed protocol with multi-hop routing	NA	NA
--	--	----	----

NA: not applicable (self-comparison provides no meaningful value).

Simulation scenario 4: throughput evaluation

Here we validate the effectiveness of our proposed protocol—both in its routing-free and multi-hop configurations—by measuring achievable throughput up to the point of Last Node to Die (LND). **Figure 13** presents the cumulative throughput (in bits) delivered to the sink node as a function of simulation rounds for our protocol and the four baseline methods.

The results consistently show that throughput decreases as network area expands, a trend observed across all protocols due to increased average communication distances. Nevertheless, our protocol exhibits substantially higher throughput than any competing scheme across all tested configurations. This superiority arises directly from the suite of innovations previously outlined.

Not only does the proposed protocol roughly double the network's operational lifetime compared with earlier approaches, but it also maintains higher data delivery rates over that extended duration while consuming less energy per transmitted bit. The prolonged survival of active nodes naturally results in a greater total volume of data reaching the sink, directly translating into higher overall throughput.

Table 14 provides a precise numerical quantification of these throughput gains, detailing the percentage improvement achieved by our protocol relative to each benchmark method.

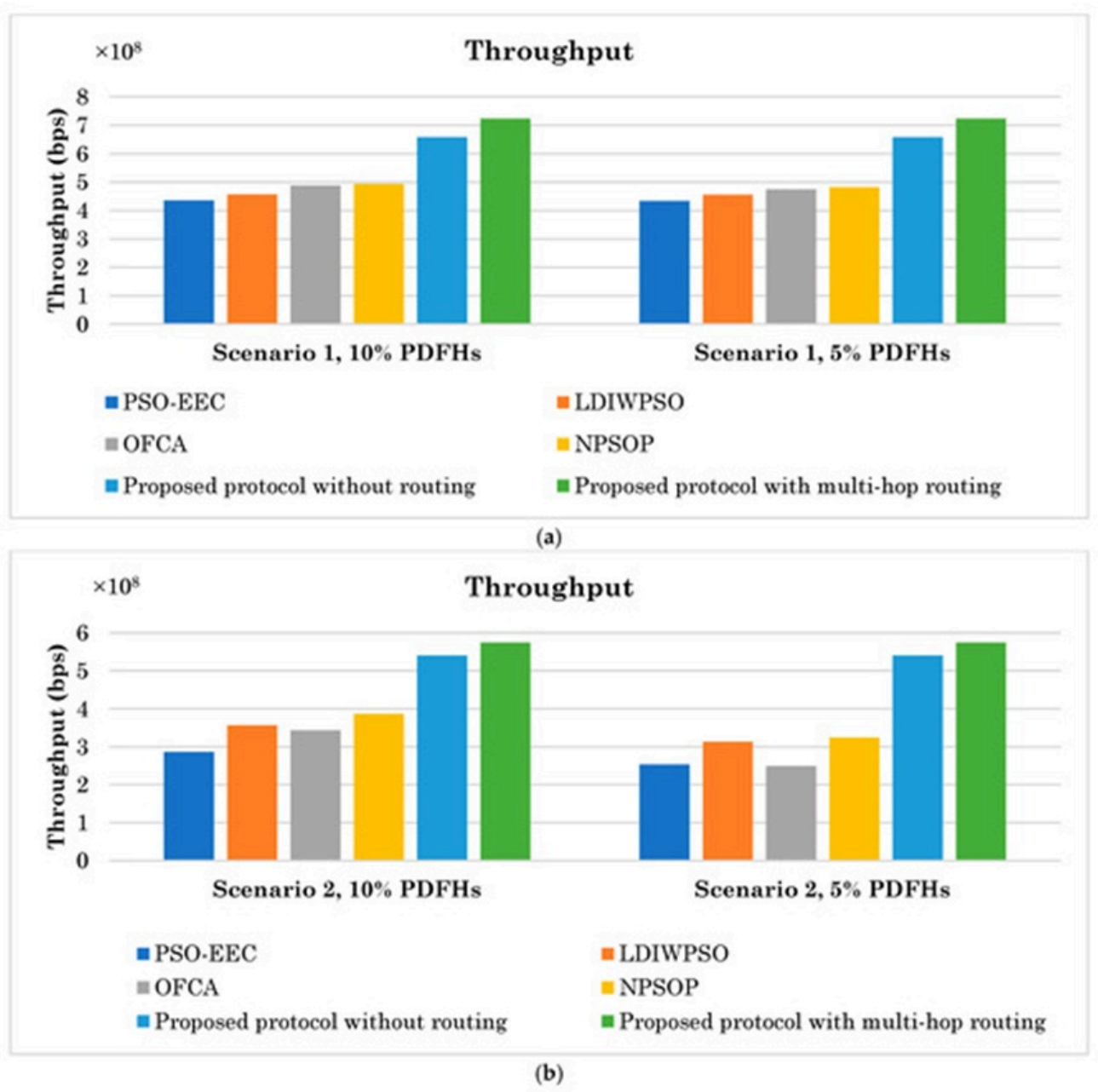


Figure 13. Cumulative throughput versus number of rounds: comparison with reference protocols for Scenarios 1 and 2 (PSO-EEC [39], LDIWPSOC [40], OFCA [41], NPSOP [42]): (a) Scenario #1: 100 $\times$ 100 m² area, 100 nodes, 10% and 5% PDFHs, and (b) Scenario #2: 500 $\times$ 500 m² area, 300 nodes, 10% and 5% PDFHs.

Table 14. Percentage improvement in throughput for the proposed protocol compared to the counterpart protocols.

Scenario description	Protocol compared	Throughput ($\times 10^8$ bits)	Improvement ratio–proposed protocol without routing	Improvement ratio–proposed protocol with multi-hop routing

Scenario #1 Number of SNs: 100 Number of DFHs: 10% Network area: 100 × 100 m ²	PSO-EEC	4.3543	51%	66%
	LDIWPSOC	4.5695	44%	58%
	OFCA	4.8841	35%	48%
	NPSOP	4.9338	33%	47%
	Proposed protocol without routing	6.5752	NA	10%
	Proposed protocol with multi-hop routing	7.2344	NA	NA
Scenario #1 Number of SNs: 100 Number of DFHs: 5% Network area: 100 × 100 m ²	PSO-EEC	4.3295	52%	67%
	LDIWPSOC	4.5530	44%	59%
	OFCA	4.7434	39%	53%
	NPSOP	4.8179	36%	50%
	Proposed protocol without routing	6.5752	NA	10%
	Proposed protocol with multi-hop routing	7.2344	NA	NA
Scenario #2 Number of SNs: 300 Number of DFHs: 10% Network area: 300 × 300 m ²	PSO-EEC	2.8610	89%	101%
	LDIWPSOC	3.5661	51%	61%
	OFCA	3.4373	57%	67%
	NPSOP	3.8712	40%	48%
	Proposed protocol without routing	5.4008	NA	6%
	Proposed protocol with multi-hop routing	5.7432	NA	NA
Scenario #2 Number of SNs: 300 Number of DFHs: 5% Network area: 300 × 300 m ²	PSO-EEC	2.5288	114%	127%
	LDIWPSOC	3.1322	72%	83%
	OFCA	2.4949	116%	130%

	NPSOP	3.2475	66%	77%
	Proposed protocol without routing	5.4008	NA	6%
	Proposed protocol with multi-hop routing	5.7432	NA	NA

NA: not applicable (self-comparison provides no meaningful value).

Simulation scenario 5: average delay evaluation

Figure 14 provides a detailed side-by-side comparison of the average end-to-end delay experienced by packets in the proposed protocol relative to the benchmark protocols. The findings clearly indicate that our protocol—whether configured with direct (single-hop) forwarding or multi-hop routing—consistently achieves the lowest average delay across all tested configurations. This substantial improvement stems directly from the protocol's advanced, context-aware algorithms and adaptive mechanisms, as described in earlier sections, enabling it to outperform existing approaches by a significant margin. Taken together, these results highlight the protocol's capability to establish new standards for low-latency performance in IoT environments, making it especially valuable for delay-sensitive applications.

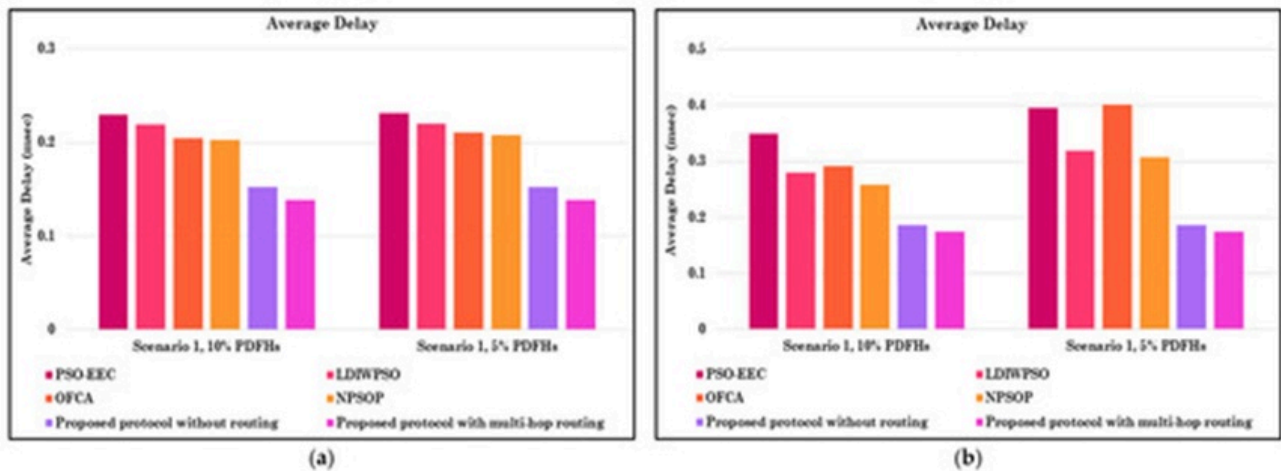


Figure 14. Average end-to-end delay versus number of rounds: comparison with reference protocols for Scenarios 1 and 2 (PSO-EEC [39], LDIWPSOC [40], OFCA [41], NPSOP [42]): (a) Scenario #1: 100 × 100 m² area, 100 nodes, 10% and 5% PDFHs, and (b) Scenario #2: 500 × 500 m² area, 300 nodes, 10% and 5% PDFHs.

Analysis of underlying causes and correlations of results

The superior performance of the proposed protocol across network lifetime, energy expenditure, throughput, and packet delay can be traced to a series of purposeful architectural and algorithmic choices. The adoption of hexagonal-shaped DFGs promotes highly uniform spatial coverage while minimizing signal attenuation, thereby reducing energy waste and extending operational endurance. The MPA optimization technique, combined with a carefully selected set of dynamic residual metrics, produces highly effective PDFH assignments that evenly distribute the workload and help prevent early node exhaustion. In the multi-hop configuration, communication ranges are shortened, further conserving energy, and the

novel RDFCF selects relay nodes based on multiple efficiency criteria, thereby minimizing overall power consumption. The batch-oriented data fusion grouping strategy significantly reduces control traffic overhead. At the same time, the inclusion of backup heads and relays guarantees uninterrupted data flow even in the presence of node failures. Collectively, these elements drive higher throughput, lower latency, and more efficient energy utilization—qualities that are particularly advantageous in time-critical IoT use cases.

Several meaningful relationships emerge when examining the interplay among the four primary performance indicators:

- **Network lifespan and energy consumption:** An inverse relationship exists between these two metrics. Schemes that achieve lower energy expenditure naturally enable sensor nodes to remain active for more rounds, thereby extending the network's overall functional duration. Our protocol leverages this advantage through refined routing decisions and intelligent PDFH selection, thereby eliminating wasteful energy consumption.
- **Network lifespan and throughput:** Extended network survival time directly translates into greater cumulative throughput. A network that stays operational longer can forward a larger total volume of data packets to the sink before any node fails. This linkage is clearly demonstrated in the proposed protocol, where careful energy stewardship enables prolonged, reliable data delivery.
- **Energy consumption and average delay:** Reduced energy usage tends to correspond with shorter average packet delays. When nodes and PDFHs deplete their batteries more slowly, they remain responsive and can forward packets promptly, reducing the time packets spend in transit. The protocol's optimized TDMA scheduling and efficient routing paths reinforce this beneficial correlation by simultaneously improving energy economy and transmission responsiveness.
- **Throughput and average delay:** Although elevated throughput typically signals strong data-handling capacity, it can occasionally lead to congestion and elevated latency in less optimized designs. The proposed protocol successfully mitigates this potential drawback by combining multi-hop forwarding with adaptive scheduling mechanisms, thereby sustaining high data rates while keeping delay increases minimal.

Conclusion

Most sensor nodes (SNs) deployed in smart city Internet of Things (IoT) applications suffer from non-rechargeable batteries with limited capacity. This inherent constraint renders such networks particularly susceptible to rapid energy depletion during periods of heightened activity, underscoring the urgent need for effective energy-saving strategies to achieve sustained network functionality. To tackle this issue, the present study introduces an advanced IoT routing protocol that integrates novel energy-efficient mechanisms for both data fusion grouping and multi-hop data forwarding.

Central to the protocol is its data fusion grouping algorithm, which significantly boosts energy efficiency by partitioning the monitored sensing field into uniform hexagonal Data Fusion Groups (DFGs). This structured division substantially reduces redundant transmissions while promoting balanced energy dissipation across all nodes, thereby considerably prolonging the network's operational lifetime.

Whenever the sink node broadcasts a request for information, each SN replies with a message containing its identifier, geographical coordinates, and current remaining energy (RE). The protocol then applies the DFG formation algorithm to systematically organize the sensing region, followed by the Marine Predators Algorithm (MPA) to select the optimal Primary Data Fusion Head (PDFH), First-Secondary Data Fusion Head (FSDFH), and Second-Secondary Data Fusion Head (SSDFH) within each group.

The selection process is guided by five carefully chosen residual metrics (RMs): the individual SN's remaining energy, the mean distance among DFG members, the distance from each candidate to the sink node, the Area-Specific Building

Occlusion (ASBO) factor, and the ratio of the average number of rounds all DFGMs have served as PDFHs to the rounds served by the current candidate PDFHi. These metrics ensure highly effective routing decisions and resource distribution.

In addition, the protocol introduces an enhanced TDMA scheduling approach, a newly designed Relay Data Fusion Cost Function (RDFCF), and an innovative Data Fusion Fitness Function (DFFF). Extensive simulation experiments clearly demonstrate that the proposed protocol outperforms existing solutions in key performance areas—namely, network lifetime, total residual energy, achieved throughput, and average packet delay—across multiple rounds, both in single-hop and multi-hop configurations.

Rigorous comparisons with established protocols, including LDIWPSOC [40], PSO-EEC [39], OFCA [41], and NPSOP [42], confirm the substantial gains delivered by the proposed enhancements. Beyond energy and communication improvements, the protocol effectively addresses real-world deployment challenges, including building occlusions of varying heights, device or link failures, and issues arising from data redundancy. These capabilities give the protocol a clear advantage over prior methods, particularly in complex and obstacle-rich environments.

The robust data fusion strategy not only streamlines transmission efficiency but also actively mitigates redundant data flows, serving as a foundational element for extending both network longevity and operational dependability. By comprehensively addressing these critical concerns, the proposed protocol lays a solid groundwork for building energy-efficient, trustworthy IoT infrastructures tailored to smart city requirements.

Consequently, we strongly recommend that the Greater Amman Municipality (GAM) and stakeholders in the Information and Communication Technology (ICT) sector seriously consider adopting and deploying the models presented here in practical settings, given the exceptional performance demonstrated by the proposed smart frameworks. This call to action is particularly urgent, as neither GAM nor the local ICT industry has yet initiated or implemented any comprehensive smart city projects in Amman, Jordan's capital.

Looking ahead, future investigations could concentrate on further refining PDFH election procedures and routing strategies for smart city IoT deployments. A particularly promising direction involves integrating machine learning algorithms to enable predictive energy management—anticipating consumption surges and dynamically adjusting routing paths in advance. Another valuable line of inquiry would involve large-scale field trials in realistic smart city testbeds featuring diverse physical obstacles and real-time application demands, thereby providing a deeper understanding of the protocol's scalability limits and resilience under actual operating conditions.

Acknowledgements

The corresponding author would like to acknowledge the Deanship of Scientific Research at the University of Jordan for the financial support granted for this work.

Conflict of interest

None

Financial support

This work was supported by the Deanship of Scientific Research at the University of Jordan [2404].

Ethics statement

None

Received: 09 Oct 2023

Revised: 27 Nov 2023

Accepted: 09 Jan 2024

Published online: 20 March 2024

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Darabkh KA, Al-Akhras M, Khalifeh AF, Jafar IF, Jubair F. An innovative RPL objective function for broad range of IoT domains utilizing fuzzy logic and multiple metrics. *Expert Syst Appl*. 2022;205:117593.
- Vishwakarma AK, Chaurasia S, Kumar K, Singh YN, Chaurasia R. Internet of things technology, research, and challenges: A survey. *Multimed Tools Appl*. 2024;1–36.
- Zeng F, Pang C, Tang H. Sensors on Internet of Things systems for the sustainable development of smart cities: A systematic literature review. *Sensors*. 2024;24:2074.
- Zhang Z, Zhang Y, Tian H, Martin A, Liu Z, Ding W. A survey of evidential clustering: definitions, methods, and applications. *Inf Fusion*. 2025;115:102736.
- Liu Z, Li J, Zhang X, Wang X. Multi-level information fusion for missing multi-label learning based on stochastic concept clustering. *Inf Fusion*. 2024;115:102775.
- Yu B, Xu R, Cai M, Ding W. A clustering method based on multi-positive–negative granularity and attenuation-diffusion pattern. *Inf Fusion*. 2023;103:102137.
- Yu B, Xu R, Cai M, Ding W. An adaptive density clustering approach with multi-granularity fusion. *Inf Fusion*. 2024;106:102273.
- Darabkh KA, Al-Akhras M, Zomot JN, Atiquzzaman M. RPL routing protocol over IoT: A comprehensive survey, recent advances, insights, bibliometric analysis, recommendations, and future directions. *J Netw Comput Appl*. 2022;207:103476.
- Zaman M, Puryear N, Abdelwahed S, Zohrabi N. A review of IoT-based smart city development and management. *Smart Cities*. 2024;7:1462–501.
- Samaiyya D, Radhika S, Chandrasekar A. An optimal model for enhancing network lifetime and cluster head selection using hybrid snake whale optimization. *Peer Peer Netw Appl*. 2023;16:1959–74.
- Hosseinzadeh M, Hemmati A, Rahmani AM. Clustering for smart cities in the internet of things: a review. *Cluster Comput*. 2022;25:4097–127.
- Wang Z, Duan J, Xing P. Multi-hop clustering and routing protocol based on enhanced snake optimizer and golden jackal optimization in WSNs. *Sensors*. 2024;24:1348.
- Darabkh KA, AlAdwan HH, Al-Akhras M, Jubair F, Rahamneh S. A revolutionary RPL-based IoT routing protocol for monitoring building structural health in smart city domain utilizing equilibrium optimizer algorithm. *Soft Comput*. 2024;28:10099–138.

Chen Y, Liu X, Rao M, Qin Y, Wang Z, Ji Y. Explicit speed-integrated LSTM network for non-stationary gearbox vibration representation and fault detection under varying speed conditions. *Reliab Eng Syst Saf*. 2024;254:110596.

Darabkh KA, Al-Akhras M. The potential of computational intelligence to extend the lifespan of Internet of Things power-limited sensor networks. In: *Proceedings of the 2024 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. Split, Croatia; 2024. p. 1–6.

Ramalingam S, Dhanasekaran S, Sinnasamy SS, Salau AO, Alagarsamy M. Performance enhancement of efficient clustering and routing protocol for wireless sensor networks using improved elephant herd optimization algorithm. *Wirel Netw*. 2024;30:1773–89.

Chai X, Lee BG, Hu C, Pike M, Chieng D, Wu R, et al. IoT-FAR: A multi-sensor fusion approach for IoT-based firefighting activity recognition. *Inf Fusion*. 2025;113:102650.

Huang C, Coskun S, Karimi HR, Ding W. A distributed state and fault estimation scheme for state-saturated systems with quantized measurements over sensor networks. *Inf Fusion*. 2024;110:102452.

Ji Y, Huang Y, Zeng J, Ren L, Chen Y. A physical–data-driven combined strategy for load identification of tire type rail transit vehicle. *Reliab Eng Syst Saf*. 2024;253:110493.

Kaleybar HJ, Davoodi M, Brenna M, Zaninelli D. Applications of genetic algorithm and its variants in rail vehicle systems: A bibliometric analysis and comprehensive review. *IEEE Access*. 2023;11:68972–93.

Kumaravel V, Panneerselvam A. A multi-objective tabu particle swarm optimization for effective cluster head selection in WSN. *Cluster Comput*. 2019;22:12275–82.

Bharathi M, Srinivas TA S. Exploring ant colony optimization for enhanced routing in IoT networks: A survey. *Adv Image Process Pattern Recognit*. 2024;7:68–83.

Placzek B. Prediction-based data reduction with dynamic target node selection in IoT sensor networks. *Future Gener Comput Syst*. 2023;152:225–38.

Darabkh KA, AlAdwan HH, Al-Akhras M, Jubair F, Rahamneh S. A new routing protocol for low-power and lossy networks utilizing computational intelligence over IoT networks. In: *Proceedings of the 2023 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*. Dubai, UAE; 2023. p. 97–102.

Darabkh KA, Asma'a BA, Al-Akhras M, Wafa'a KK. Improving network lifetime in IoT sensor network based on particle swarm optimization, clustering, and mobile sink. In: *Proceedings of the 2022 4th IEEE Middle East and North Africa COMMUNICATIONS Conference (MENACOMM)*. Amman, Jordan; 2022.

Amutha J, Sharma S, Sharma SK. An energy efficient cluster-based hybrid optimization algorithm with static sink and mobile sink node for wireless sensor networks. *Expert Syst Appl*. 2022;203:117334.

Suresh SS, Prabhu V, Parthasarathy V, Senthilkumar G, Gundu V. Intelligent data routing strategy based on federated deep reinforcement learning for IoT-enabled wireless sensor networks. *Meas Sens*. 2024;31:101012.

Priyadarshi R. Energy-efficient routing in wireless sensor networks: a meta-heuristic and artificial intelligence-based approach: A comprehensive review. *Arch Comput Methods Eng*. 2024;31:2109–37.

Gao J, Wang Z, Lei Z, Wang R-L, Wu Z, Gao S. Feature selection with clustering probabilistic particle swarm optimization. *Int J Mach Learn Cybern*. 2024;15:3599–617.

Ravi G, Das MS, Karmakonda K. Reliable cluster-based data aggregation scheme for IoT network using hybrid deep learning techniques. *Meas Sens*. 2023;27:100744.

Darabkh KA, Al-Akhras M. An improved routing protocol for IoT sensors utilizing clustering techniques and optimization methods. In: *Proceedings of the 6th IEEE International Conference on Advanced Communication Technologies and Networking (CommNet 2023)*. Rabat, Morocco; 2023.

Darabkh KA, Zomot JN, Al-qudah Z. EDB-CHS-BOF: energy and distance based cluster head selection with balanced objective function protocol. *IET Commun*. 2019;13:3168–80.

Heidari E. A novel energy-aware method for clustering and routing in IoT based on whale optimization algorithm and Harris Hawks optimization. *Computing*. 2024;106:1013–45.

He S, Li Q, Khishe M, Mohammed AS, Mohammadi H, Mohammadi M. The optimization of nodes clustering and multi-hop routing protocol using hierarchical chimp optimization for sustainable energy efficient underwater wireless sensor networks. *Wirel Netw*. 2023;30:233–52.

Darabkh KA, Amareen AB, Al-Akhras M, Kassab WK. An innovative cluster-based power-aware protocol for Internet of Things sensors utilizing mobile sink and particle swarm optimization. *Neural Comput Appl*. 2023;35:19365–408.

Somula R, Cho Y, Mohanta BK. SWARAM: Osprey optimization algorithm-based energy-efficient cluster head selection for wireless sensor network-based Internet of Things. *Sensors*. 2024;24:521.

Bian Z, Qu J, Zhou J, Jiang Z, Wang S. Weighted adaptively ensemble clustering method based on fuzzy co-association matrix. *Inf Fusion*. 2023;103:102099.

Zhou X, Yang Q, Liu Q, Liang W, Wang K, Liu Z, et al. Spatial-temporal federated transfer learning with multi-sensor data fusion for cooperative positioning. *Inf Fusion*. 2023;105:102182.

Rawat P, Chauhan S. Particle swarm optimization-based energy efficient clustering protocol in wireless sensor network. *Neural Comput Appl*. 2021;33:14147–65.

Choudhary S, Sugumaran S, Belazi A, El-Latif AAA. Linearly decreasing inertia weight PSO and improved weight factor-based clustering algorithm for wireless sensor networks. *J Ambient Intell Humaniz Comput*. 2023;14:6661–79.

Giri A, Dutta S, Neogy S. An optimized fuzzy clustering algorithm for wireless sensor networks. *Wirel Pers Commun*. 2022;126:2731–51.

Huangshui H, Xinji F, Chuhan W, Ke L, Yuxin G. A novel particle swarm optimization-based clustering and routing protocol for wireless sensor networks. *Wirel Pers Commun*. 2023;133:2175–202.

Faramarzi A, Heidarinejad M, Mirjalili S, Gandomi AH. Marine predators algorithm: a nature-inspired metaheuristic. *Expert Syst Appl*. 2020;152:113377.

Al-Betar MA, Awadallah MA, Makhadmeh SN, Alyasseri ZAA, Al-Naymat G, Mirjalili S. Marine predators algorithm: a review. *Arch Comput Methods Eng*. 2023;30:3405–35.

Rai R, Dhal KG, Das A, Ray S. An inclusive survey on marine predators algorithm: variants and applications. *Arch Comput Methods Eng*. 2023;30:3133–72.

Mugemanyi S, Qu Z, Rugema FX, Dong Y, Wang L, Bananeza C, et al. Marine predators algorithm: A comprehensive review. *Mach Learn Appl*. 2023;12:100471.

Darabkh KA, Zomot JN, Al-Qudah Z, Khalifeh AF. Impairments-aware time slot allocation model for energy-constrained multi-hop clustered IoT nodes considering TDMA and DSSS MAC protocols. *J Ind Inf Integr*. 2021;25:100243.

Shahzad MK, Islam SMR, Kwak K-S, Nkenyereye L. AEF: Adaptive en-route filtering to extend network lifetime in wireless sensor networks. *Sensors*. 2019;19:4036.

Darabkh KA, El-Yabroudi MZ, El-Mousa AH. BPA-CRP: a balanced power-aware clustering and routing protocol for wireless sensor networks. *Ad Hoc Netw*. 2018;82:155–71.

Darabkh KA, Al-Akhras M. Prolonging IoT sensor networks lifetime for different smart city applications utilizing a three-dimension MPA based fitness function. In: *Proceedings of the 2024 IEEE 15th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON 2024)*, Berkeley, CA, USA, 2024.

Darabkh KA, Al-Akhras M. A five-dimension MPA based fitness function for optimizing energy in IoT sensor networks considering various smart city applications. In: *Proceedings of the 2024 IEEE 15th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON 2024)*, Berkeley, CA, USA;2024.

Dehkordi AB. EDBLSD-IIoT: A comprehensive hybrid architecture for enhanced data security, reduced latency, and optimized energy in industrial IoT networks. *J Supercomput*. 2025;81:359.

Rathee M, Kumar S, Dilip K, Dohare U. Towards energy balancing optimization in wireless sensor networks: A novel quantum inspired genetic algorithm based sinks deployment approach. *Ad Hoc Netw*. 2023;153:103350.

Gad AG. Particle swarm optimization algorithm and its applications: A systematic review. *Arch Comput Methods Eng*. 2022;29:2531–61.

Abdulzahra AMK, Al-Qurabat AKM, Abdulzahra SA. Optimizing energy consumption in WSN-based IoT using unequal clustering and sleep scheduling methods. *Internet Things*. 2023;22:100765.

Rui K. Improving energy efficiency in wireless sensor networks (WSNs) using two-level fuzzy clustering and Artificial Bee Colony (ABC) optimization. *Int J Electron*. 2025:1–26.

Al-Betar MA, Abu Doush I, Makhadmeh SN, Al-Naymat G, Alomari OA, Awadallah MA. Equilibrium optimizer: a comprehensive survey. *Multimed Tools Appl*. 2023;83:29617–66.

Shen B, Khishe M, Mirjalili S. Evolving marine predators algorithm by dynamic foraging strategy for real-world engineering optimization problems. *Eng Appl Artif Intell*. 2023;123:106207.

Wang Y, Henning I. A deterministic distributed TDMA scheduling algorithm for wireless sensor networks. In: *Proceedings of the 2007 International Conference on Wireless Communications, Networking and Mobile Computing*, Shanghai, China; 2007. p. 2759–62.